AIDS Epidemiology
Methodological Issues

Nicholas P. Jewell
Klaus Dietz
Vernon T. Farewell
Editors

Birkhäuser
Boston · Basel · Berlin
1992

RECOVERY OF INFORMATION AND ADJUSTMENT FOR DEPENDENT CENSORING USING SURROGATE MARKERS

James M. Robins and Andrea Rotnitzky

Abstract: A class of tests and estimators for the parameters of the Cox proportional hazards model, the accelerated failure time model, and a model for the effect of treatment on the mean of a response variable of interest are proposed that use surrogate marker data to recover information lost due to independent censoring and to adjust for bias due to dependent censoring in randomized clinical trials. We construct an adaptive test that (i) is asymptotically distribution free under the null hypothesis of no treatment effect on survival, (ii) incorporates surrogate marker data, and (iii) is guaranteed to be locally more powerful than the ordinary log-rank test against proportional hazards alternatives when the baseline failure time distribution is Weibull. The proposed test is shown to outperform the log-rank test in a series of simulation experiments. We also prove the optimal estimator within our class is semiparametric efficient by first showing that our estimation problem is a special case of the general problem of parameter estimation in an arbitrary semiparametric model with data missing at random, and then deriving a representation for the efficient score in this more general problem.

1. Introduction

Randomized clinical trials of the effect of a new treatment on mortality from a chronic fatal disease such as AIDS must be conducted over prolonged periods of time. It is important to be able to stop such trials the moment that it be determined that the new treatment prolongs survival. To this end, interim analyses comparing the survival experience in the two treatment arms are typically conducted at 6 monthly or yearly intervals. Although subject-specific data on the evolution of time-dependent covariates that predict subsequent survival (such as CD4-count and serum HIV-antigen levels) will often be available at the time of an interim analysis, typically treatment arm specific survival curves are compared using a log-rank or weighted log-rank test that ignores the surrogate marker data. In this paper, we refer to any post treatment variable that, conditional on treatment arm, predicts subsequent survival as a surrogate marker. One major goal of this paper is to develop statistical methods that increase the power to detect a treatment effect by incorporating information on surrogate markers and yet do not compromise the validity of the usual intention to treat analysis of the null hypothesis of no-treatment effect. Specifically, we shall propose a class of tests that are guaranteed to reject, in large samples, at their nominal rate under the null hypothesis of no effect of treatment on survival, but may be much more powerful than any weighted log-rank test when a treatment effect exists. In this paper, we restrict attention to tests conducted at a single point in time. Group sequential procedures, applicable to repeated interim analyses, will be the subject of a separate report.

The second major goal of this paper is to develop statistical methods that can be used to adjust for non-random non-compliance and dependent censoring in randomized clinical trials. In an AIDS randomized trial comparing a treatment A, say high-dose AZT, with a

ERRATA for

RECOVERY OF INFORMATION AND ADJUSTMENT FOR DEPENDENT CENSORING USING SURROGATE MARKERS

James M. Robins and Andrea Rotnitzky

Page	Line	Uncorrected	Corrected
302	Eq. (3.3b)	$I[S_i^* = S_j^*]\left[1 - \hat{\lambda}_{S_j^*}(X_j) e^{\hat{\alpha}' \cdot W_i(X_j)}\right]$	$\left[1 - \hat{\lambda}_{S_j^*}(X_j) e^{\hat{\alpha}' \cdot W_i(X_j)}\right]^{I[S_i^* = S_j^*]}$
311	3rd to last	$\lambda_0(u) I\{X^* \geq u\}$	$\lambda_0(u) e^{\beta_0 Z} I\{X^* \geq u\}$
311	2nd to last	$\lambda_0(u) I(X^* \geq u)$	$\lambda_0(u) e^{\beta_0 Z} I(X^* \geq u)$
313	Eq. (3.10a)	$\lambda_Q(x)$	$\lambda_Q(x \mid \bar{L}(x), Z)$
313	Eq. (3.10b)	$\lambda_Q(x)$	$\lambda_Q(x \mid \bar{L}(x), Z)$
317	2nd para/line 12	$\| A \|$	$\| A \|^2$
318	following Eq. (4.4)	Π	π
320	2nd	$\mathbf{m}(D) = \overline{\pi}_m$	$\mathbf{m}(D) = \overline{\pi}_M$
327	Eq. (A.5)	$\Pi[S_\beta^{(F)} \mid \Lambda_0^{(F),\perp}]$	$\Pi[S_\beta^{(F)} \mid \Lambda_0^{(F)}]$

treatment B, say low-dose AZT, it is common for subjects to fail to comply with the assigned treatment protocol and to initiate treatment with a new but unproven therapy C, say, aerosolized pentamidine prophylaxis for pneumocystis carinii pneumonia (PCP). In this setting, in order to obtain some useful information regarding the benefits of high- vs low-dose AZT, suppose it is agreed to regard subjects as censored at the time they initiate treatment with aerosolized pentamidine. Unfortunately, as discussed by Lagakos et al. (1991), the actual level of the associated censored data log-rank intention to treat test may differ from its nominal level if (a) subjects in the low-dose AZT arm are more likely than subjects in the high-dose arm to initiate therapy with aerosolized pentamidine, and (b), within each arm, censoring is not independent of failure (i.e., death) because subjects at high risk of failure (that is, subjects with recurrent episodes of PCP) are more likely to be censored (i.e., to initiate therapy with aerosolized pentamidine). Together (a) and (b) imply that censoring and failure are dependent. Suppose, in this setting, based on substantive considerations, the investigators conducting the trial are willing to assume that, among a subset of subjects in a given treatment arm with identical PCP histories up to time t, the decision to initiate prophylaxis therapy with aerosolized pentamidine at t is unrelated to prognosis. Then, the methods proposed in this paper can be used to construct an asymptotically α-level test of the null hypothesis that the distribution of failure times in the high- and low-dose AZT arms would have been the same had no subject abandoned protocol and initiated therapy with aerosolized pentamidine.

The paper is organized as follows. In Section (2), we (a) specify 3 models whose parameters we shall estimate - the proportional hazards (PH) model, the accelerated failure (AF) time model, and a model for the mean of a random variable measured near the end of follow-up, and (b) formally define dependent and independent censoring mechanisms. The combination of one of these models and a censoring mechanism constitute a semiparametric model for the observed data (Begun et al., 1983).

In Section (3) we show that, under an independent censoring mechanism, (a) the non-centrality parameter of the locally most powerful asymptotically distribution free (ADF) test of the null hypothesis of no treatment effect on survival against PH (or AF) alternatives equals the semiparametric information bound (SIB) for estimating the PH (or AF) model parameter β_0 at β_0=0 and (b) the SIB is greater when surrogate marker data is available than when it is not.

In Sections (3a)-(3b), we propose a class of tests that (a) incorporate surrogate marker data, (b) are ADF under the null hypothesis of no treatment effect on survival, and (c) contain tests whose non-centrality parameters converge to the SIB. Further, in Sections (3b)-(3c), we explicitly construct an ADF adaptive test that (a) incorporates surrogate marker data, (b) is guaranteed to be locally more powerful than the ordinary log-rank test against PH or AF alternatives when the baseline failure time distribution is Weibull, and (c), as predicted by theory, outperforms the log-rank test in a series of simulation experiments. In Section (3d) we show that our methods allow one to improve upon the Kaplan-

Meier estimate of the baseline survival curve by incorporating surrogate marker data.

In Section (3e), we propose a class of unbiased estimating equations for our 3 models whose solutions are consistent, asymptotically normal even in the presence of dependent censoring. Our class of estimators are based on recent work by Robins (1991) and Robins, Rotnitzky, and Zhao (1992), and can be viewed as a generalization of the Koul et al. (1981) censored linear regression estimator that uses surrogate marker data to recover information and adjust for dependent censoring.

In Section (4), we calculate the semiparametric variance bound for our semiparametric models and show that the asymptotic variance of the optimal estimator in our proposed class attains the bound. We calculate the semiparametric variance bound by showing that the semiparametric problem we are considering is actually a special case of the general problem of estimating the parameters of an arbitrary semiparametric model in the presence of data missing at random [Rubin (1976)]. In Theorems (4.1)-(4.3) we provide a representation for both (a) the efficient score and (b) the influence function of any regular, asymptotically linear estimator in this general estimation problem [provided that, for each subject, the probability of observing complete (i.e., full) data is bounded away from zero]. These representation theorems were derived by Robins, Zhao, Rotnitzky, and Lipsitz (1991). We show that the efficient score is a solution to an operator (i.e., integral) equation which does not in general have a closed form solution. However, we show that, in the special case of our mean model, a closed form solution exists. Although, for the PH and AF model, the efficient score does not exist in closed form, we describe how one can obtain a semiparametric efficient estimator by using linear combinations of estimating equations in our class (Newey, 1992b).

In Section (4e), we show that, in the <u>absence</u> of surrogate marker data, a slightly modified version of the original Koul et al., (1981) estimator is semiparametric efficient in the AF model under an independent censoring mechanism. In addition, as a by product of our investigation of dependent censoring, we construct, in Section (3h) and Appendix 4, a new estimator for the parameters of an AF model when censoring is due solely to a known end-of-follow-up date.

The paper is organized such that a reader who is more interested in the abstract theory of estimation in semiparametric models with data missing at random (rather than in particular applications of this theory) can proceed directly to the general representation theorems (4.1)- (4.3) of Section 4. With the notable exception of the proof of the general representation Theorem (4.1) provided in Appendix 2, the proofs of most theorems are only sketched and regularity conditions largely ignored.

2. A Formalization of the Problem

To be concrete, consider a two-arm double-blind trial comparing the effect of high versus low dose AZT treatment on the survival of AIDS patients in which dependent censoring due to initiation of prophylaxis therapy with aerosolized pentamidine may occur.

Suppose patient enrollment began on August 1, 1989 and continued until follow-up was artificially terminated on July 31, 1990 for the purposes of conducting an interim analysis. The maximum potential follow-up time c_{max} is one year. Many subjects will have been under follow-up for less than a year.

With time measured as time since randomization, define for subject i, $i=1,\cdots,n$,

Q_{1i} - censoring time due to end of follow-up. Thus, $Q_{1i} \le c_{max} = 1$.

Q_{2i} - time to initiation of prophlyaxis therapy for PCP

$Q_i = \min(Q_{1i}, Q_{2i})$-censoring time

T_i - failure time in the absence of prophylaxis for PCP

$Z_i = (Z_{1i},\cdots,Z_{pi})'$. Z_{1i} is a dichotomous treatment arm indicator. $(Z_{2i},\cdots,Z_{pi})'$ is a vector of pretreatment variables such as age and gender.

$L_i(t)$ and $\bar{L}_i(t) = \{L_i(u) ; 0 \le u \le t\}$ are respectively the recorded value at t and the recorded history up to t of a vector of time-dependent covariates such as CD4-count and HIV-antigen level.

For reasons that will become clear below, we may need to consider artificially ending follow-up at a time c^*, $c^* \le c_{max}$. Let $\Delta_i = I(T_i < c^*)$. If we disregard all events occurring subsequent to $\min(c^*, Q_i, T_i)$, the observable random vectors are

$$\{c^*, Z_i, X_i = \min(Q_i, T_i, c^*), \bar{L}_i(X_i), \tau_i = I[X_i \ne Q_i], \Delta_i \tau_i\}, i=1,\cdots,n \tag{2.1}$$

which we assume are independent and identically distributed. We call (2.1) the c^*-observed data. Note Δ_i is observed only if $\tau_i=1$. If we do not artificially end follow-up prior to c_{max}, we would have $c^* = c_{max}$, $X_i = \min(Q_i, T_i)$, and $\Delta_i \tau_i = \tau_i$. We shall consider three models for the treatment effect. The PH model

$$pr[T>t \mid Z] = \{\bar{F}_0(t)\}^{\exp(\beta_0' Z)} \tag{2.2a}$$

The AF model

$$pr[T>t \mid Z] = \bar{F}_0(e^{\beta_0' Z} t) \tag{2.2b}$$

and a model for the conditional mean of Δ

$$E[\Delta \mid Z] = g(Z, \beta_0) \tag{2.2c}$$

where $g(\cdot,\cdot)$ is a known function, $\bar{F}_0(t)$ is an unspecified survival curve, and we have suppressed the i subscript. Gray and Tsiatis (1989) discuss the potential usefulness of model (2.2c) in studies of a potentially curative therapy. An investigator may be interested in the mean at time c^* of random variables other than Δ - e.g. the random variable recording HIV-antigen level. The results we obtain for model (2.2c) extend straightforwardly to models for the mean at c^* of any random variable.

If Z_i is the dichotomous treatment arm indicator then the null hypothesis $\beta_0=0$ is equivalent to the usual "intention to treat null hypothesis"

$$\text{pr}[T>t \mid Z] = \bar{F}_0(t) \tag{2.3}$$

of equality of treatment arm specific survival curves under both models (2.2a) and (2.2b). We shall assume the censoring mechanism satisfies

$$\lambda_Q(u \mid \bar{L}(u),Z,T) = \lambda_Q(u \mid \bar{L}(u),Z) \tag{2.4a}$$

where $\lambda_Q(u \mid \cdot) = \lim_{h\to 0} h^{-1}\text{pr}[u\leq X<u+h, \tau=0 \mid X\geq u,\cdot]$. (2.4a) would generally hold under the assumptions concerning initiation of prophylaxis therapy given in the introduction, if data on PCP history was recorded in $\bar{L}(u)$, and if there were no secular trends among patients entering the trial over the year of enrollment. Indeed (2.4a) would hold even if Q_1 and T are dependent given Z due to secular changes in the prognosis of patients at entry over the calendar year of enrollment (e.g., due to improvements in the standard care of AIDS patients), provided these changes could be explained by secular changes in the covariates in $\bar{L}(u)$. [In certain settings, an investigator might believe that there had been important changes in the prognosis of patients at entry over the year of enrollment that could not be explained by secular changes in the variables recorded in $\bar{L}(u)$. In Appendix 4, we show how the methods developed below can be extended to estimate the treatment effect in such settings.] We shall at times consider the implications of assuming the following in addition to (2.4a):

$$\lambda_Q[u \mid Z,T] = \lambda_Q[u \mid Z] \tag{2.4b}$$

$$\lambda_Q[u \mid Z,\bar{L}(T),T] = \lambda_Q[u \mid Z] \tag{2.4c}$$

$$\lambda_Q[u \mid \bar{L}(u),Z,T,\bar{L}(T)] = \lambda_Q[u \mid \bar{L}(u),Z] \tag{2.4d}$$

When (2.4a) holds, but (2.4b) is false, we say we have independent censoring given $\bar{L}(u)$ and Z, but dependent censoring given Z. (2.4c) would hold if no one took PCP prophylaxis (i.e., $Q_i = Q_{1i}$), and (conditional on Z) there exists no secular trends among patients entering the trial over the year of enrollment. (2.4c) implies (2.4a), (2.4b), and (2.4d). Finally, (2.4d) implies (2.4a) and is equivalent to the definition of missing at random (MAR) given by Rubin (1976).

3. Estimation and Testing:

3a. Introduction

Under each of the three models (2.2a)-(2.2c), our estimate $\hat{\beta}$ will be a solution to $S\{\hat{D}(\beta)\} = 0$ with

$$S\{\hat{D}(\beta)\} = \Sigma_i \tau_i \hat{D}_i(\beta)/\hat{K}_i \ ; \tag{3.1}$$

and our tests of the hypothesis that a given value of β equals the true β_0 will be based on $S\{\hat{D}(\beta)\}$ divided by the estimate of its asymptotic standard error given in Section (3j.2), where (i) $\hat{D}_i(\beta)$ depends on the model (2.2a) - (2.2c) and will be defined below, and (ii) $\hat{K}_i \equiv \hat{K}_i(X_i)$ is an estimate of $K_i(X_i)$ with $K(t) \equiv \exp\{-\int_0^t \lambda_Q[u \mid \bar{L}(u),Z]du\}$ and is computed as follows. We suppose we have a correctly specified stratified time-dependent Cox model

$$\lambda_Q[u \mid \bar{L}_i(u),Z_i] = \lambda_{S_i^*}(u)\exp[\alpha_0' W_i(u)] \tag{3.2}$$

where α_0 is a parameter vector, $W_i(u)$ is a vector of functions $w(\bar{L}_i(u),Z_i)$ of $\bar{L}_i(u)$ and Z_i, S_i^* is a discrete stratification variable that is a function of Z_i, and $\lambda_{S_i^*}(u)$ are unspecified stratum specific baseline hazard functions. As an example, $W_i(u)$ might include a subject's most recent CD4-count prior to u, the number of PCP bouts up to u, and their interactions with treatment arm. Let $\hat{\alpha}$ be the Cox maximum partial likelihood estimator of α_0. To obtain $\hat{\alpha}$ one can use standard time-dependent Cox proportional hazards model software by regarding the subjects with $X_i = Q_i$ as the "failures." Let

$$\hat{\lambda}_{S_j^*}(X_j) = (1-\tau_j)\left[\sum_{i=1}^{n} e^{\hat{\alpha}' \cdot W_i(X_j)} Y_i(X_j) I[S_i^* = S_j^*]\right]^{-1} \tag{3.3a}$$

where $Y_i(u) = I[X_i \geq u]$ records "at-risk" status at u. $\hat{\lambda}_{S_j^*}(X_j)$ is the Cox baseline hazard estimator for censoring at X_j in stratum S_j^*. Finally,

$$\hat{K}_i(u) \equiv \prod_{\{j;X_j \leq u, \tau_j = 0\}} {}^{I[S_i^* = S_j^*]} \left[1 - \hat{\lambda}_{S_j^*}(X_j) e^{\hat{\alpha}' \cdot W_i(X_j)}\right] . \tag{3.3b}$$

We will occasionally write $\hat{K}_i \equiv \hat{K}_i(X_i)$ as $\hat{K}_i^W$ to stress its dependence on $W_i(u)$. $\hat{K}_i$ is an $n^{1/2}$-consistent estimator of $K_i(X_i)$ (Anderson and Gill, 1982).

$\hat{D}_i(\beta)$ is either exactly equal to or is an estimate of a random variable $\tilde{D}_i(\beta)$, where $0=S^{(F)}\{\tilde{D}(\beta)\} \equiv \Sigma_i \tilde{D}_i(\beta)$ would be an unbiased estimating equation for β with solution $\tilde{\beta}$ if, contrary to fact, there were no Q-censoring and we had thus observed

$$\{c^*, X_i^* \equiv \min(T_i,c^*), \bar{L}_i(X_i^*), Z_i, \Delta_i\},\ i=1,\cdots,n \tag{3.4}$$

which we shall call the c*-full data. That is, $E[\tilde{D}_i(\beta_0)] = 0$, and further, $\tilde{D}_i(\beta)$ and $\tilde{\beta}$ are functions of the c*-full data. In contrast, $\tau_i \hat{D}_i(\beta)$ and, thus, $S\{\hat{D}(\beta)\}$ and $\hat{\beta}$ will depend only on the c*-observed data. Before providing definitions for $\hat{D}_i(\beta)$ and $\tilde{D}_i(\beta)$ in the general case, we shall consider in Sections (3b)-(3d) tests of the null hypotheses (2.3) (i.e., tests of β_0=0 in models (2.2a) and (2.2b)) and estimation of the mean model (2.2c) under the assumptions that Z_i and S_i^* are both the (0,1) treatment arm indicator.

3b. Tests of the Null Hypothesis (2.3)

We define $\hat{D}_i(0) \equiv \tilde{D}_i(0) = R_i(Z_i - n^{-1}\Sigma_i Z_i)$ where $R \equiv r(\Delta, \ell n(X^*)) \equiv \Delta r_1\{\ell n(T)\} + (1-\Delta) r_2\{\ell n(c^*)\}$ with $r_1(\cdot)$ and $r_2(\cdot)$ scoring functions chosen by the analyst. Clearly

$E[\tilde{D}_i(0)] = 0$ under (2.3), since (2.3) implies T is independent of Z, and R is random only through its dependence on T. We can construct an ADF test of (2.3) based on $S\{\hat{D}(0)\}$ since the following fundamental lemma, proved in Section (3h), implies that $n^{-1/2}S\{\hat{D}(0)\}$ has asymptotic mean 0 under (2.3) and (2.4a).

Fundamental Lemma: If $K_i(X_i)>\sigma>0$ with probability one (w.p.1) for some $\sigma>0$ and (2.4a) and (2.3) hold, then $E[\tau_i\hat{D}_i(0)/K_i(X_i)] = 0$.

To help understand the lemma, note that when the stronger condition (2.4d) is true, $K_i(X_i)$ is the probability that subject i survived to X_i without being censored. Hence, if subject *i* is observed to fail at X_i and $K_i(X_i) = .1$, then subject i would need to count for $1/K_i(X_i) = 10$ subjects (himself and 9 others who were censored prior to X_i). Although, under the weaker assumption (2.4a), $K_i(X_i)$ is not the probability of surviving to X_i without being censored, nevertheless the lemma remains true.

To see why the Lemma may not hold if $K_i(X_i)$ is not bounded away from zero, suppose (1) data on $\bar{L}_i(u)$ was not collected or needed; (2) T_i was uniformly distributed on (0,.8) independent of Z_i so (2.3) was true; (3) $Q_i \coprod T_i \mid Z_i$ so (2.4d) was true with (3a) $Q_{1i} = c_{max}=1$ w.p.1, and (3b) Q_{2i} uniformly distributed on (0,.4) given $Z_i = 0$, and $Q_{2i}>1$ w.p.1 given $Z_i = 1$, so that all subjects with $Z_i = 0$ will initiate prophylaxis by time .4 if alive; (4) $c^* \equiv 1.0$; and (5) 50% of subjects were assigned to each treatment arm.

Then $K_i(X_i)>0$ w.p.1, but $K_i(X_i)\rightarrow 0$ as $X_i\rightarrow .4$ for subjects with $Z_i=0$ and $E[\tau_i\hat{D}_i(0)/K_i(X_i)] = (.25)(.5)$ if $r_1(\ell nT)=0$ if $T\leq .4$, and $r_1(\ell nT)=1$ for $T>.4$ [The 25% of subjects observed to fail between .4 and .8 without having initiated prophylaxis are all in the Z=1 arm with $K_i(X_i)=1$. Hence, for this group $\tau_i\hat{D}_i(0)/K_i(X_i) = \hat{D}_i(0)$ has mean $(1-1/2)E[R_i \mid T_i>.4] = .5]$. Artificially ending follow-up at a time c^* such that $K_i(c^*)>\sigma>0$ w.p.1 will insure that $K_i(X_i)>\sigma>0$ w.p.1, and thus, by the fundamental lemma, that $E[\tau_i\hat{D}_i(0)/K_i(X_i)] = 0$. For example, we can choose $c^* = .4-\epsilon, \epsilon>0$ to ensure that $K_i(c^*)>\sigma>0$ w.p.1 for some σ.

In the remainder of this subsection, we shall suppose that no subject initiates PCP prophylaxis so that $Q_i=Q_{1i}$, and that there are no secular trends so that censoring mechanism (2.4c) holds. Eq. (2.4c) implies the Cox model (3.2) is guaranteed to be correctly specified with $\alpha_0=0$. Nonetheless, as we shall see, the estimation of the coefficients α_0 that are known to be zero is the key to recovery of information from the surrogate markers $L_i(u)$.

Write $S\{\hat{D}(0)\}$ divided by the standard error estimate of Section 3j.2 as $\psi(r,w,c^*)$ to emphasize its dependence on r, w, and c^*. $\psi(r,w,c^*)$ is asymptotically distribution free under (2.3) and (2.4c). Specifically, a corollary of Theorem 3.4 of Section (3j.2) is

Theorem 3.1: If $K(X)>\sigma>0$ with probability one, then $\psi(r,w,c^*)$ is asymptotically N(0,1) under (2.3) and (2.4c). Further, under local AF or PH alternatives with $\beta_0=n^{-1/2}$, $\psi(r,w,c^*)$ is asymptotically normal with variance 1 and mean $[NC(r,w,c^*)]^{1/2}$, say.

Notational Convention: The value of the non-centrality parameter NC(r,w,c*) in Theorem (3.1) depends on whether we are considering AF or PH model alternatives. When a quantity such as NC(r,w,c*) depends on the model, we shall write it as $NC^{PH}(r,w,c^*)$ or $NC^{AF}(r,w,c^*)$, as appropriate, whenever it is important to indicate the model. However, when, as in Theorem (3.1), we leave off the model-identifying superscript, a quantity such as NC(r,w,c*) will be used to generically refer to either model.

We briefly discuss how to choose c* to ensure that $K(X)>\sigma>0$ with probability one. If, as we shall allow, the distribution of Q has a point mass at c_{max} (as would be the case if an initial group of subjects were enrolled simultaneously), then K(X)=0 with positive probability if $c^* = c_{max}$, but $K(X)>\sigma>0$ if $c^*=c_{max}^-$. If Q is absolutely continuous with respect to Lebesgue measure with support on (0,1) then (a) $K(X)\geq K(c^*)>\sigma>0$ for all $c^*<c_{max}$ where σ may depend on c*, and (b) K(X)>0 if $c^* = c_{max}$ but will not be bounded away from zero which can create problems (even asymptotically) due to division by $\hat{K}_i(X_i)$. As discussed in Section 3c, the problems associated with setting $c^* = c_{max}$ might be solved practically (and presumably theoretically) by treating the last observation in each stratum S* as a failure when computing $\hat{K}_i(X_i)$ and/or by choosing R such that R/K(X) is bounded as $X\rightarrow c_{max}$.

Let $\hat{K}_i^0$ be $\hat{K}_i$ with $\hat{\alpha} \equiv 0$ in (3.3a) and (3.3b) and let $\psi(r,0,c^*)$ be the associated test statistic. $\hat{K}_i^0$ is the treatment-arm-specific Kaplan-Meier estimator of K(u) evaluated at $u = X_i$ and does not depend on surrogate marker data L(u). Let α_j and $W_j(u)$, $j=1,\cdots,J$, represent the parameter and covariate vector in the j^{th} of J nested Cox models (3.2) ordered by increasing dimension of $W_j(u)$. For a given choice of $R = r[\Delta,\ell n(X^*)]$, the optimal covariate is $W_{op}^r(u) \equiv \{Z-E(Z)\}\{K(u)\}^{-1}E[R \mid \bar{L}(u),Z,Y(u)=1] = \{Z-E(Z)\}E[\tau R/K(X) \mid \bar{L}(u),Z,Y(u)=1]$ where the equality is by (3.10f) below. In Section (4c), we prove

Lemma 3.1: (a) Under censoring mechanism (2.4c), if $j^*>j$ $NC(r,0,c^*)\leq NC(r,w_j,c^*)\leq NC(r,w_{j^*},c^*)\leq NC(r,w_{op}^r,c^*)$ where the last inequality is strict unless there exists a constant matrix b such that $W_{op}^r(u) = b\, W_{j^*}(u)$ with probability one. (b) For

$$W(u) = (W^*(u)',ZW^*(u)')' \tag{3.5a}$$

with $W^*(u) = w^*\{\bar{L}(u)\}$, NC(r,0,c*) is strictly less than NC(r,w,c*) if, for a component $W_m^*(u)$ of $W^*(u)$,

$$E[\{Z-E(Z)\}^2 \int_0^\infty dN_Q(u)\mathrm{Cov}\{R,W_m^*(u) \mid Z,T\geq u\}] \neq 0 \tag{3.5b}$$

with $N_Q(u) = I[X\leq u, \tau=0]$.

Inequality (3.5b) will generally hold if $W_m^*(u)$ and R are dependent given Z. [To obtain a test that is strictly more powerful than $\psi(r,0,c^*)$, it is often important to use covariates of the form (3.5a)]. Lemma (3.1a) implies that increasing the number of covariates in the Cox model (3.2) usually leads to improvements in power. We can obtain an asymptotically distribution-free test of (2.3) with non-centrality parameter approximating that of $NC(r,w^r_{op},c^*)$ either by (i) specifying a richly parameterized Cox model (3.2); or, possibly, by (ii) extending (3.2) by adding, as a covariate, an estimate of $W^r_{op}(u)$ given by $(Z-n^{-1}\Sigma_i Z_i)$ times $\hat{p}r[\tau=1 \mid \bar{L}(u),Z,Y(u)=1]$ multiplied by the predicted value from the least squares fit of a linear regression of $R/\hat{K}(X)$ on functions of $\bar{L}(u)$ and Z among those subjects with $Y(u)=1$ and $\tau=1$. Here $\hat{p}r[\tau=1 \mid \bar{L}(u),Z,Y(u)=1]$ is the predicted value from the fit of a logistic model for the probability that $\tau=1$ on functions of $\bar{L}(u)$ and Z among subjects with $Y(u)=1$. Separate linear and logistic models may be fit at each u in the set of observed Q-censoring times. The resulting test, $\psi(r,\hat{w}^r_{op}, c^*)$, (a) will have the same limiting distribution and thus non-centrality parameter as $\psi(r,w^r_{op},c^*)$ if each of the linear and logistic regression models are correctly specified, and (b) will still be asymptotically N(0,1) under (2.3) and (2.4c) if the models are misspecified or even incompatible with any joint distribution for the data (see Appendix 3). This robustness to model misspecification is convenient since $\bar{L}(u)$ is the entire history of a complex process up to u, and, thus, estimation of $E[\tau R/K(X) \mid \bar{L}(u),Z,Y(u)=1]$ by non-parametric regression is not practical due to the curse of dimensionality.

Since (2.4c) implies independent censoring given Z, ADF competitors to $\psi(r,w,c^*)$ will be the standard intention-to-treat weighted log rank tests $\psi_{LR} \equiv \psi_{LR}(c^*)$ with $c^* = c_{max}$, where $\psi_{LR}(c^*) \equiv \Sigma_i\Delta_i\tau_i\omega_i(Z_i-E_i)/[\Sigma_i\Delta_i\tau_i\omega_i^2E_i(1-E_i)]^{1/2}$, ω_i is a weight function, $E_i = \Sigma_j Y_j(T_i)Z_j/\Sigma_j Y_j(T_i)$. $\psi_{LR}(c^*)$ and $\psi(r,0,c^*)$, in contrast to $\psi(r,w,c^*)$, do not depend on the surrogates $\bar{L}(u)$, but only on

$$\{c^*, \tau_i, \Delta_i\tau_i, X_i, Z_i\},\ i=1,\cdots,n \quad ,$$

which we call the <u>c^*-$\overline{\text{sur}}$-data</u>.

<u>Remark</u>: Another test that incorporates surrogate marker data could be based on modifying the generalized "Buckley-James" estimating function considered by Ritov (1990) to $\sum_{i=1}^{n}(Z_i-n^{-1}\Sigma_j Z_j)\{\tau_i R_i+(1-\tau_i)\hat{E}[R_i \mid \bar{L}_i(u),Y_i(u)=1,Z_i]\}$ evaluated at $u=X_i$ divided by its estimated asymptotic standard error. However, this test is guaranteed to be asymptotically N(0,1) under (2.3) and (2.4c) only if the estimate of the conditional expectation $E[R \mid \bar{L}(u),Z,Y(u)=1] = E[\tau K(u)R/K(X) \mid \bar{L}(u),Z,Y(u)=1]$ is based on a non-parametric regression of $\tau K(u)R/K(X)$ on $\bar{L}(u)$ and Z given $Y(u)=1$, which, as discussed above, is usually not practical due to the curse of dimensionality. Further, it follows from Theorem (3.2f) below that, if $\bar{L}(u)$ was a sufficiently simple process so that the non-parametric regression was practical and, thus, the modified "Buckley-James" test was asymptotically N(0,1), then the test would be asymptotically equivalent to and would have the same non-centrality parameter as some test $\psi(r^{(1)},w,c^*)$, where, in general $R_i \neq R_i^{(1)} \equiv r^{(1)}(\Delta_i,\ell nX_i^*)$.

Write ψ_{LR}^{PH} and ψ_{LR}^{AF} respectively for the tests that have $\omega_i=1$ and $\omega_i = \partial \ell n\lambda_{\ell nT}(u)/\partial u \mid_{u=\ell nT_i}$ where $\lambda_{\ell nT}(u)$ is the hazard of ℓnT. $\psi_{LR}^{PH}(c^*)$ and $\psi_{LR}^{AF}(c^*)$ are locally optimal in the class of weighted log-rank tests against their respective alternatives and are equal when $\bar{F}_0(t)$ is Weibull (Gill, 1980). In fact, results due to Begun et al. (1983) and Ritov and Wellner (1988) imply that, under (2.4c), they are locally optimal in the much larger class of regular, asymptotically linear (RAL) tests of (2.3) that use only the c^*-$\overline{\text{sur}}$-data. Informally, RAL tests are ones which are asymptotically N(0,1) under (2.3) and (2.4c), are asymptotically equivalent to a sample average, and whose convergence is uniform under Pitman sequences of distributions all satisfying the null hypothesis (2.3). See Section (4a) for a formal definition.

In Appendix 1, we show that the non-centrality parameter of the locally most powerful RAL test of (2.3) against PH (or AF) alternatives with $\beta_0=n^{-1/2}$ is the semiparametric information bound (SIB) for the estimation of $\beta_0=0$ in the PH (AF) model under censoring mechanism (2.4c). The SIB is defined in Section (4a) below and in Begun et al. (1983). The SIB depends on the available data. Write $NC_s^{c^*}$ and $NC_{\bar{s}}^{c^*}$ for the SIB at $\beta_0=0$ based on the c^*-observed data and the c^*-$\overline{\text{sur}}$-data respectively [where, in accord with the convention described previously, $NC_s^{c^*}$ refers to $NC_s^{c^*,PH}$ if we are considering the PH model and to $NC_s^{c^*,AF}$ if we are considering the AF model.] $NC_s^{c^*} \to NC_s^{c_{max}} \equiv NC_s$ and $NC_{\bar{s}}^{c^*} \to NC_{\bar{s}}^{c_{max}} \equiv NC_{\bar{s}}$ as $c^* \to c_{max}$. In Theorem (3.2), we shall show $NC_{\bar{s}} \leq NC_s \leq NC_{full}$ where NC_{full} is the SIB at $\beta_0=0$ based on the c^*-full-data (3.4) with $c^*=c_{max}$; and, further, in the absence of staggered entry (i.e., $Q_i = c_{max}$ for all i), $NC_{\bar{s}} = NC_s = NC_{full}$ and no information is gained from surrogate marker data. Thus we can view $(NC_{full} - NC_{\bar{s}})$ as the amount of information lost due to staggered entry in the absence of surrogate data and $(NC_s - NC_{\bar{s}})/(NC_{full} - NC_{\bar{s}})$ as the fraction of the lost information that can be recovered by optimally utilizing surrogate data. Begun et al. (1983) and Ritov and Wellner (1988) prove that the non-centrality parameter of ψ_{LR}^{PH} and ψ_{LR}^{AF} equal $NC_{\bar{s}}^{PH}$ and $NC_{\bar{s}}^{AF}$ respectively. We prove in Section (4e)

Theorem 3.2: Under (2.4c) and with Z dichotomous, (a) $NC_{\bar{s}} \leq NC_s$. Sufficient (and nearly necessary) conditions for equality are that either (i) $Q_i = c_{max}$ for all i or (ii), for all u, $\bar{L}(u)$ and T are conditionally independent given $(Z,X>u,T<c_{max})$. (b) $NC_s \leq NC_{full}$. Sufficient (and nearly necessary) conditions for equality are that either (i) $Q_i = c_{max}$ for all i, or (ii), for all u, $Var[T \mid Z,\bar{L}(u),X>u,T<c_{max}]=0$ in which case we call L(u) a perfect surrogate. (c) There exists $r_s \equiv r_s[\Delta,\ell n(X^*)]$ such that $NC(r_s, w_{op}^{r_s}, c^*) = NC_s^{c^*}$. In general, r_s is a solution to an integral equation which has a closed form solution only in special cases. (d) In particular, if L(u) is a perfect surrogate, $r_s = r_{pf} \equiv r_{pf}[\Delta,\ell n(X^*)] = \Delta r_{pf,1}(\ell nT) + (1-\Delta)r_{pf,2}(\ell nc^*)$. r_{pf} depends on the alternative. Specifically $r_{pf,1}^{AF}(u) = \partial \ell nf_{\ell nT}(u)/\partial u$, $r_{pf,1}^{PH}(u) = 1 - \int_{-\infty}^{u} \lambda_{\ell nT}(u)du$, $r_{pf,2}(u) = E[r_{pf,1}(\ell nT) \mid \ell nT>u]$, so $r_{pf,2}^{AF}(u) =$

$-\lambda_{\ell n T}(u)$ and $r^{PH}_{pf,2}(u) = -\int_{-\infty}^{u} \lambda_{\ell n T}(u)du$. r^{AF}_{pf} is proportional to r^{PH}_{pf} if $\bar{F}_0(u)$ is Weibull. (e) There exists $r_{\bar{s}} \equiv r_{\bar{s}}[\Delta, \ell n(X^*)]$ such that $NC(r_{\bar{s}},0,c^*)$ equals $NC_{\bar{s}}^{c^*}$. In particular, if $\lambda_Q(u \mid Z) = \lambda_Q(u)$, then for m = 1,2,

$$r_{\bar{s},m}(u) = r_{pf,m}(u)K(e^u) + \int_0^{e^u} r_{pf,2}(\ell nx)\lambda_Q(x)K(x)dx.$$

(f) Any RAL modified Buckley-James test of (2.3) is asymptotically equivalent to [i.e. has the same influence function as] some test $\psi(r,w^r_{op},c^*)$.

Since r_s does not exist in closed form, it is rather difficult to construct an adaptive test with non-centrality parameter $NC_s^{c^*}$. However, because $r_{\bar{s}}$ exists in closed form, we can use Lemma (3.1) to construct a simple adaptive ADF test $\psi(\hat{r}_{\bar{s}},w,c^*)$ of (2.3) that incorporates surrogate marker data and is guaranteed to be more powerful than the log-rank test against AF (or PH) alternatives if $\bar{F}_0(u)$ is in a specified parametric family $\bar{F}(u;\theta)$, such as the Weibull family. [Note, if $\bar{F}_0(u)$ is Weibull with $\lambda_0(u) = au^b$, then, under (2.3), $r^{AF}_{pf,2}(u) = -\lambda_{\ell n T}(u) = -ae^{u(b+1)}$ and $r^{AF}_{pf,1}(u) = (b+1) - \lambda_{\ell n T}(u)$ for Weibull parameters $\theta=(a,b)$.] We construct $\hat{r}_{\bar{s}}$ as follows.

Obtain "maximum likelihood" estimates of θ [and, thus, of $r_{pf,m}(u)$, for m = 1,2] based on the data (X_i, τ_i),i=1,⋯,n, with $c^* \equiv c_{max}$ under the assumption of independent censoring. Let $\hat{K}^{\dagger}(u)$ be the Kaplan-Meier estimate of K(u) and $\hat{\lambda}^{\dagger}_Q(u)$ be the Nelson-estimator of $\lambda_Q(u)$ based on all the data, i.e., based on (X_i, τ_i),i=1,⋯,n, with $c^* = c_{max}$. Let $\hat{r}_{\bar{s},1}(u)$ and $\hat{r}_{\bar{s},2}(u)$ be as in Theorem 3.2e except K(u), $\lambda_Q(u)$, and $r_{pf,m}(u)$ are replaced by these estimates. Then, as is argued in Appendix 3,

Lemma 3.2: $\psi(\hat{r}_{\bar{s}},w,c^*)$ will be asymptotically N(0,1) under (2.3) and (2.4c) even if the model for $\bar{F}_0(u)$ is misspecified. Further, if the parametric model for $\bar{F}_0(u)$ is correctly specified, $\lambda_Q(u) = \lambda_Q(u \mid Z)$ and Eqs. (3.5a)-(3.5b) hold with $R = R_{\bar{s}}$, then, $NC(\hat{r}_{\bar{s}},w,c^*) = NC(r_{\bar{s}},w,c^*) > NC(r_{\bar{s}},0,c^*) = NC_{\bar{s}}^{c^*}$. Thus, as $c^* \to c_{max}$, the non-centrality parameter of our adaptive test will eventually exceed that of the optimal log-rank test based on all the data. By Lemma 3.1, a high-dimensional w in (3.2) would lead to further gains in power.

3c. Simulation Experiment

The use of marker data can lead to particularly large increases of power in trials in which (1) new subjects are still being enrolled at the time of the first interim analysis, (2) the observed failures are concentrated amongst subjects who enrolled early in the study and (3) there exists a marker whose values soon after enrollment are good predictors of subsequent failure-time. Excellent candidates for trials in which the proposed methods can lead to substantial increases in power are the ACTG pediatric AIDS protocols, such as protocol 152 and 154. In these protocols, an infant's "failure time" is defined to be the time at which the infant's development falls below a prespecified level. For example, an infant is considered to have "failed" when the infant's head circumference first falls below the second percentile for age. Because of strong tracking in the rate of head growth, estimated rates of growth based on measurements made in the first several months after enrollment should be highly predictive of subsequent "failure." Thus, we assumed there was an interim analysis conducted exactly one year after initial enrollment into the trial began and patient accrual was uniform over the year, so that the censoring variable was uniformly distributed on (0,1). We assumed that no subject failed in the first six months after his/her enrollment. After six months, in our simulations under the null hypothesis (2.3), failure time in both arms was exponentially distributed with a hazard of 3.22. We used a PH model with β_0=-.35 under the alternative. To keep the analysis simple, we chose, as a surrogate, a single time-independent marker measured immediately after initiation of treatment. In the first set of trials, we unrealistically assumed that the surrogate was perfectly correlated with (and in fact equal to) subsequent failure time. In a second set of trials we made the more realistic assumption that the value of the surrogate for subject i was obtained by multiplying his/her failure time by an independent uniform random variable on (.75, 1.33). The results reported in each row of Table 1 are based on

TABLE 1. Results of a Simulation Experiment

Row	Surrogate Strength	Scoring Function r	Artificial End of F/U c* in yrs	W(u) Cox Model Covariates	Tail Method	Actual Rejection Rate under Null in %	Asymptotic Relative Efficiency
1	Perf	$r_{\bar{s}}$	.9	W_{op}^{r}	3	4.0	236
2	Perf	$r_{\bar{s}}$	1.0	W_{op}^{r}	3	4.0	237
3	Perf	$r_{\bar{s}}$	.9	Ind	3	3.5	148
4	Perf	$r_{\bar{s}}$	1.0	Ind	3	3.5	165
5	Mod or Perf	$r_{\bar{s}}$	.9	0	3	5.0	94
6	Mod or Perf	$r_{\bar{s}}$	1.0	0	3	7.0	100
7	Mod	$r_{\bar{s}}$	.9	W_{op}^{r}	3	6.0	158
8	Mod	$r_{\bar{s}}$	1.0	W_{op}^{r}	3	6.0	163
9	Mod	$r_{\bar{s}}$	.9	Ind	3	5.5	135
10	Mod	$r_{\bar{s}}$	1.0	Ind	3	5.0	141
11	Mod	$\hat{r}_{\bar{s}}$	1.0	0	3, 4	7.0	102
12	Mod	$\hat{r}_{\bar{s}}$	1.0	Ind	2, 3, 4	5.0	141
13	Mod	$\hat{r}_{\bar{s}}$	1.0	0	1	14.0	-
14	Mod	$\hat{r}_{\bar{s}}$	1.0	Ind	1	17.5	-
15	Mod	$\hat{r}_{\bar{s}}$	1.0	0	2	11.5	-

200 realizations. Each realization represented a trial with 300 subjects in each treatment arm. The column labeled "rejection rate" is the actual rejection rates of the nominal 5% tests $|\psi(r,w,c^*)|>1.96$ based on simulations conducted under the null hypothesis. Column 7 gives the estimated asymptotic relative efficiency (ARE) of our tests compared to the standard log-rank test based on the ratios of the square of the average Z-values (under the alternative), where ARE $\equiv NC(r,w,c^*)/NC_{LR}$. The log-rank test is based on follow-up through $c_{max} = 1$ year.

Column 2 characterizes the surrogate strength as perfect or moderate [uniform (.75, 1.33) noise]. Columns 3, 4, and 5 describe r,w, and c^* for our tests $\psi(r,w,c^*)$. w_{op}^r is the optimal covariate [depending on r and c^*] defined in Section 3b. The covariate "Ind" is given by (3.5a) with $W^*(u)$ a time-independent dichotomous covariate recording whether a subject's surrogate is above the population median, and thus is not optimal. "None" implies that no covariates were used in model (3.2) to estimate K(u) and thus refers to the test $\psi(r,0,c^*)$.

The numbers in column 5 refer to modifications to $\psi(r,0,c^*)$ to prevent large values of $R_i/\hat{K}_i(X_i)$as $X_i \to 1$ when $c^*=c_{max}=1$. The number "1" represents the unmodified $\psi(r,0,c^*)$. Modification 2 replaces $R_i \equiv r(\Delta_i,X_i^*)$ by $R_i - r(\Delta_i,1)$. In modification 3, if the last subject, say j, at risk in a given treatment arm is Q-censored (i.e., $Q_j=X_j$), subject j is treated as a failure at X_j in the analysis. In modification 4, c^* is replaced by the minimum of c^* and the earliest time t at which in either treatment arm only one subject remains at risk.

In rows 1 and 2 we obtain striking ARE's of 236 and 237 while preserving the nominal α-level under the null hypothesis (2.3). The relative efficiencies of 148 and 165 in rows 3 and 4 demonstrate the loss of efficiency from using a non-optimal w. Rows 5-6 demonstrate the further loss of efficiency when we fail to use the surrogate data. As predicted by our theoretical calculations, the ARE is equal to that of the log-rank test when $c^*=1.0$, although, even using modification 3, the rejection rate (RR) of 7% slightly exceeds the nominal. Rows 7-15 present the results of more realistic simulation experiments that use a surrogate with a moderate correlation with failure time. The ARE's of our tests in rows 7-10 are considerably less than when we had a perfect surrogate, although the ARE of 163 in row 8 is certainly great enough to be important.

In rows (11)-(12), we observe, that, as predicted by theory, $\psi(\hat{r}_{\bar{s}},w,c^*)$ performs similarly to $\psi(r_{\bar{s}},w,c^*)$. [In computing $\hat{r}_{\bar{s}}$, we took the guarantee period of .5 years as known, but estimated the Weibull parameters a=3.22 and b=1.] Again at $c^*=1.0$, the actual 7% RR of $\psi(\hat{r}_{\bar{s}},0,c^*)$ under tail methods (3) and (4) slightly exceeds the nominal. In contrast, the RR with $c^*=1.0$ is equal to the nominal for $\psi(\hat{r}_{\bar{s}},w,c^*)$ with w the indicator covariate, provided we use a tail modification. The slight elevation of the RR in row 11 is attibutable to the standard error estimator of $S\{\hat{D}(0)\}$ underestimating the actual standard error (data not shown). In contrast, the RR of 14% and 17.5% in rows (13) and (14) of the unmodified $S\{\hat{D}(0)\}$ reflects the fact that the unmodified $S\{\hat{D}(0)\}$ is not centered and clearly demonstrates the need for modification. On the other hand, the elevation of the

RR in row (15) is solely due to bias in the standard error estimate when using method (2). [Choosing c^* to be .9 (or less) guarantees the actual α-level equals the nominal (whether or not a modification is used), although some efficiency is thereby lost.] Finally, we note that if one uses modification (3) or (4), it is an algebraic fact that the value of $\psi(r,0,c^*)$ does not further depend on whether modification (2) is also used.

3d. Estimation of the Mean Model

For dichotomous Z, we consider a saturated model (2.2c) for the mean of Δ. Specifically, $g(Z,\beta_0) = \beta_{0,0}+\beta_{0,1}Z$, with $\beta_0=(\beta_{0,0},\beta_{0,1})'$. Define $\hat{D}(\beta) = (\Delta-g(Z,\beta),Z\{\Delta-g(Z,\beta)\})'$, and $W^{\Delta}_{op}(u) = \{Z-E(Z)\}E[\Delta \mid \bar{L}(u),Z,Y(u)=1]$. Let $\hat{\beta}^0$ and $\hat{\beta}^w$ be the estimators based on $\hat{K}^0_i$ and $\hat{K}^w_i$. In Section (4c), we prove

Lemma 3.3: Under (2.4c), (a) $\hat{\beta}^{w^{\Delta}_{op}}$ and $\hat{\beta}^0$ are semiparametric efficient based on the c^*-observed and c^*-$\overline{\text{sur}}$ data respectively; (b) $\hat{\beta}^0$ is never more and is usually less efficient than $\hat{\beta}^w$; and (c) as an estimator of $\beta_{0,0}$ = pr[$T<c^* \mid Z=0$], $\hat{\beta}^0_0$ is asymptotically equivalent to the usual Kaplan-Meier estimator $1-\Pi\{1-[\Sigma_j I(Z_j=0)Y_j(X_i)]^{-1}\}$, with the product over the set $\{i;\ \tau_i=1,Z_i=0,\Delta_i=1\}$ of observed failures in treatment arm Z=0, so $\hat{\beta}^w_0$ usually improves upon the Kaplan-Meier estimator by using surrogate marker data.

3e. Definition of $S\{\hat{D}(\beta)\}$ and $S^{(F)}\{\tilde{D}(\beta)\}$

With one exception, the estimating functions $S^{(F)}\{\tilde{D}(\beta)\}=\Sigma_i\tilde{D}_i(\beta)$ are not new. We review their definitions here both (a) to fix notation and (b) because the $S\{\hat{D}(\beta)\}$ are closely related to the $S^{(F)}\{\tilde{D}(\beta)\}$.

Let h(Z), h(u,Z), and $\theta(u,Z)$ be fixed functions taking values in R^p. For any V(u) define $\tilde{E}[V(u)] \equiv n^{-1}\Sigma_i V_i(u)$ and $\hat{E}\{V(u)\} \equiv n^{-1}\Sigma_i(\tau_i/\hat{K}_i)V_i(u)$.

For model (2.2c), $\tilde{D}^{MN}(\beta) \equiv \hat{D}^{MN}(\beta) \equiv \tilde{D}^{MN}(\beta,h) \equiv h(Z)\{\Delta-g(Z,\beta)\}$.

For model (2.2a), $\tilde{D}^{PH}(\beta) \equiv \tilde{D}^{PH}(\beta,h) \equiv \int_0^\infty dN_T(u)\{h(u,Z) - \tilde{\mathscr{L}}^{PH}(u,\beta,h)\}$, where $N_T(u) = I[X^*\le u,\Delta=1]$, $X^*=\min(T,c^*)$, $\tilde{\mathscr{L}}^{PH}(u,\beta,h) = \tilde{E}[e^{\beta'Z}I\{X^*>u\}h(u,Z)]/\tilde{E}[I\{X^*>u\}e^{\beta'Z}]$. Ritov and Wellner (1988) discuss the estimating functions $S^{(F)}\{\tilde{D}^{PH}(\beta)\}$. $\hat{D}^{PH}(\beta)$ replaces $\tilde{\mathscr{L}}$ by $\hat{\mathscr{L}}$ in the definition of $\tilde{D}^{PH}(\beta)$, where $\hat{\mathscr{L}}$ is defined like $\tilde{\mathscr{L}}$ but with $\hat{E}$ replacing $\tilde{E}$. Note $\tau\hat{D}^{PH}(\beta)$, in contrast to $\tau\tilde{D}^{PH}(\beta)$, can be computed from the c^*-observables (2.1).

For model (2.2b) we consider two choices of $\tilde{D}(\beta)$, namely, $\tilde{D}^{AF1}(\beta)$ and $\tilde{D}^{AF2}(\beta)$. $\tilde{D}^{AF1}(\beta) \equiv \tilde{D}^{AF1}(\beta,h) \equiv \int_{-\infty}^\infty dN_{\epsilon(\beta)}(u)\{h(u,Z) - \tilde{\mathscr{L}}^{AF1}(u,\beta,h)\}$ where $N_{\epsilon(\beta)}(u) = I[\nu(\beta) \le u,\Delta=1]$, $\nu(\beta) = \min\{\epsilon(\beta),\mu(\beta)\}$, $\mu(\beta) = \ell n c^* + \beta'Z$, $\epsilon(\beta) = \ell nT+\beta'Z$, and $\tilde{\mathscr{L}}^{AF1}(u,\beta,h) = \tilde{E}[I\{\nu(\beta)>u\}h(u,Z)]/\tilde{E}[I\{\nu(\beta)>u\}]$. Tsiatis (1990), Ritov (1990), Wei, Ying, and Lin (1990), and Kalbfleisch and Prentice (1980) discuss the estimating function $S^{(F)}\{\tilde{D}^{AF1}(\beta)\}$. $\hat{D}^{AF1}(\beta)$ replaces $\tilde{\mathscr{L}}$ by $\hat{\mathscr{L}}$ in the definition of $\tilde{D}^{AF1}(\beta)$.

We define $\tilde{D}^{AF2}(\beta) \equiv \hat{D}^{AF2}(\beta) \equiv \tilde{D}^{AF2}(\beta,h,\theta) \equiv [\int_{-\infty}^\infty dN_{\epsilon(\beta)}(u)\{h(u,Z) - \tilde{\mathscr{L}}^{AF2}(u,\beta,h)\}] -$

$\int_{-\infty}^{\infty} du I\{\nu(\beta)\geq u\}\{\theta(u,Z) - \tilde{\mathscr{L}}^{AF2}(u,\beta,\theta)\}$, where $\tilde{\mathscr{L}}^{AF2}(u,\beta,h) = \tilde{E}[I\{\mu(\beta)>u\}h(u,Z)]/\tilde{E}[I\{\mu(\beta)>u\}]$. $S^{(F)}\{\tilde{D}^{AF2}(\beta)\}$ is a new estimating function for β_0 in the AF model with a fixed known potential censoring time c^*. In practice, one would select $\theta(u,Z)$ to be a step function so the integral could be easily evaluated. Note $\tau_i\tilde{D}_i^{AF2}(\beta)$, in contrast to $\tau_i\tilde{D}_i^{AF1}(\beta)$, can be computed from the c^*-observables (2.1) since $\mu(\beta)$, in contrast to $\nu(\beta)$, is always observed.

$\tilde{D}^{AF2}(\beta,h,\theta)$ has a particularly simple form when Z is a dichotomous (0,1) variable. Using the fact that, for dichotomous Z, we can uniquely represent any function g(u,Z) as $g_1(u)Z+g_0(u)$, define $r_2(u) = -\int_{-\infty}^{u}\theta_1(x)dx$, and $r_1(u) = h_1(u) + r_2(u)$. Then $\tilde{D}^{AF2}(\beta,h,\theta) = \tilde{D}^{AF2}(\beta,r) \equiv R(\beta)[Z-\tilde{E}(Z)]$ where $R(0)=\{\Delta r_1(\ell nT)+(1-\Delta)r_2\{\ell n(c^*)\}$ as in Section 3b; and, more generally, $R(\beta) \equiv r\{\Delta(\beta),X^*(\beta)\} \equiv \Delta(\beta)r_1[\epsilon(\beta)] + \{1-\Delta(\beta)\}\{r_2[c^*(\beta)]\}$, where $c^*(\beta) \equiv \min(\ell nc^*, \ell nc^*+\beta)$, $X^*(\beta) \equiv \min[\epsilon(\beta),c^*(\beta)]$ and $\Delta(\beta) = I[\epsilon(\beta)<c^*(\beta)]$. We call $\Delta(\beta)$ an artificial (full-data) censoring indicator since if either (a) $\beta<0$ and Z=1 or (b) $\beta>0$ and Z=0, $\Delta(\beta)$ need not equal Δ. Note $\Delta(\beta)$ is discontinuous in β and $\Delta(0)=\Delta$.

3f. Asymptotic Distribution Theory for $\tilde{\beta}$ and $S^{(F)}\{\tilde{D}(\beta_0)\}$

$S\{\hat{D}(\beta)\}$ and $S^{(F)}\{\tilde{D}(\beta)\}$ are discontinuous in β under model the AF model (2.2b). In this instance, we define $\hat{\beta}$ and $\tilde{\beta}$ to be solutions to $n^{-1/2}\, S\{\hat{D}(\beta)\} = o_p(1)$ and $n^{-1/2}S^{(F)}\{\tilde{D}(\beta)\} = o_p(1)$, and one would compute $\hat{\beta}$ and $\tilde{\beta}$ respectively by minimizing $S\{\hat{D}(\beta)\}'S\{\hat{D}(\beta)\}$ and $S^{(F)}\{\tilde{D}(\beta)\}'S^{(F)}\{\tilde{D}(\beta)\}$ in a neighborhood of β_0. It will be useful to discuss the asymptotic distribution of $\tilde{\beta}$ based on the c^*-full data (3.4) before that of $\hat{\beta}$ based on c^*-observed data (2.1). Except for $\tilde{\beta}^{AF2}$, the following results are not new. Let $\mathscr{L}$ be defined like $\tilde{\mathscr{L}}$ but with a true expectation replacing $\tilde{E}$. Let $D(\beta)$ be defined like $\tilde{D}(\beta)$ except with $\mathscr{L}$ replacing $\tilde{\mathscr{L}}$.

<u>Theorem 3.3</u>: Under regularity conditions, for $|\beta-\beta_0|$ of $O(n^{-1/2})$, under the models (2.2), $n^{-1/2}S^{(F)}\{\tilde{D}(\beta_0)\} = n^{-1/2}\Sigma_i U_f + o_p(1)$,

$$n^{-1/2}S^{(F)}\{\tilde{D}(\beta)\} = n^{-1/2}S^{(F)}\{\tilde{D}(\beta_0)\} - \kappa(\beta-\beta_0) + o_p(1) \quad , \tag{3.6}$$

$n^{-1/2}S^{(F)}\{\tilde{D}(\beta_0)\}$ and $n^{1/2}(\tilde{\beta}-\beta_0)$ are asymptotically normal with mean 0 and asymptotic variance $Var(U_f)$ and $\kappa^{-1}Var(U_f)\kappa^{-1/}$ respectively .

Under model (2.2c), $U_f^{MN} \equiv U_f^{MN}(h) = D^{MN}(\beta_0,h) = \tilde{D}^{MN}(\beta_0,h)$, $\kappa^{MN} = E[U_f^{MN}(h)U_f^{MN}(h_{op}^{MN})'] = E[h(Z)\partial g(Z,\beta_0)/\partial\beta']$, $h_{op}^{MN}(Z) = [\partial g(Z,\beta_0)/\partial\beta]\{Var[\varsigma\mid Z]\}^{-1}$, $\varsigma \equiv \Delta-g(Z,\beta_0)$.

Under (2.2a), $U_f^{PH} \equiv U_f^{PH}(h) = \int_0^{\infty} dM_T(u)\{h(u,Z) - \mathscr{L}^{PH}(u,\beta_0,h)\}$, where $dM_T(u) = dN_T(u) - \lambda_0(u)I\{X^*\geq u\}du$, and $\lambda_0(u) = -\partial\ell n\bar{F}_0(u)/\partial u$. Set $h_{op}^{PH}(u,Z) = Z$. $\kappa^{PH} \equiv \kappa^{PH}(h) = E[U_f^{PH}(h)U_f^{PH}(h_{op}^{PH})'] = E[\int_0^{\infty}\lambda_0(u)I(X^*\geq u)\Phi^{PH}(h)\Phi^{PH}(h_{op}^{PH})'du] = E[\int_0^{\infty}dN_T(u)\{\mathscr{L}^{PH}(u,\beta_0,hh_{op}^{PH/}) - \mathscr{L}^{PH}(u,\beta_0,h)\mathscr{L}^{PH}(u,\beta_0,h_{op}^{PH})'\}]$, where

$\Phi(h) \equiv h(u,Z) - \mathcal{L}(u,\beta_0,h)$ and $\mathcal{L}(u,\beta_0,h_1h_2')$ is a pxp matrix.

Under (2.2b), $U_f^{AF1} \equiv U_f^{AF1}(h) = \int_{-\infty}^{\infty} dM_\epsilon(u)\{h(u,Z) - \mathcal{L}^{AF1}(u,\beta_0,h)\}$, $\kappa^{AF1} \equiv \kappa^{AF1}(h) =$ $E[U_f^{AF1}(h)U_f^{AF1}(h_{op}^{AF})'] = E[\int_{-\infty}^{\infty} dN_\epsilon(u)\{\mathcal{L}^{AF1}(u,\beta_0,hh_{op}^{AF}{}') - \mathcal{L}^{AF1}(u,\beta_0,h)\mathcal{L}^{AF1}(u,\beta_0,h_{op}^{AF})'\}]$, $\epsilon \equiv \epsilon(\beta_0)$, $\nu \equiv \nu(\beta_0)$, $dM_\epsilon(u) = dN_\epsilon(u) - I\{\nu \geq u\}\lambda_\epsilon(u)du$, $\lambda_\epsilon(u)$ is the hazard function of ϵ, and $h_{op}^{AF}(u,Z) = Z\partial \ell n\lambda_\epsilon(u)/\partial u$.

Further, $U_f^{AF2} \equiv U_f^{AF2}(h,\theta) = D^{AF2}(\beta_0) - E[D^{AF2}(\beta_0) \mid Z]$, $\kappa^{AF2} \equiv \kappa^{AF2}(h,\theta) =$ $E[U_f^{AF2}(h,\theta)U_f^{AF2}(h_{op}^{AF},\theta_{op}^{AF})']$ where $\theta_{op}^{AF}(u,Z) = Z\partial\lambda_\epsilon(u)/\partial u$. If $Z\in\{0,1\}, U_f^{AF2} \equiv U_f^{AF2}(r) =$ $\{R(\beta_0)-E[R(\beta_0)]\}\{Z-E[Z]\}$.

Remarks: For $\tilde{\beta}^{AF1}$ and $\tilde{\beta}^{PH}$, Theorem 3.3 is a direct consequence of Anderson and Gill (1982), Ritov and Wellner (1988), Tsiatis (1990), and Ritov (1990), among others. For $\tilde{\beta}^{MN}$, Theorem 3.3 is standard. The key step in the proof for $\tilde{\beta}^{AF2}$ is given in Theorem A.2 of Appendix 3. A key identity is based on the following lemma which can be proved by integration by parts.

Lemma 3.4: Any $U_f^{AF1}(h)$ can be written as $U_f^{AF2}(h,\theta)$ with $\theta(u,Z) \equiv \lambda_\epsilon(u)\, h(u,Z)$. Conversely any $U_f^{AF2}(h,\theta)$ can be written as $U_f^{AF1}(g)$ for g=h+h** with $h^{**}(u,Z) \equiv$ $\int_u^\mu \{h^*(x,Z) - \mathcal{L}^{AF2}(u,\beta_0,h^*)\}\bar{F}_\epsilon(x)dx/\bar{F}_\epsilon(u)$, $\mu = \ell nc^* + \beta_0'Z$, $\bar{F}_\epsilon(u) \equiv pr(\epsilon > u)$, and $h^*(u,Z) \equiv$ $\lambda_\epsilon(u)h(u,Z) - \theta(u,Z)$. In particular $U_f^{AF2}(h_{op}^{AF},\theta_{op}^{AF}) = U_f^{AF1}(h_{op}^{AF})$.

3.g. Asymptotic Distribution Theory for $\hat{\beta}$ and $S\{\hat{D}(\beta_0)\}$

The next theorem states that the asymptotic distribution of $\hat{\beta}$ and $S\{\hat{D}(\beta_0)\}$ differ from that of $\tilde{\beta}$ and $S^{(F)}\{\tilde{D}(\beta_0)\}$ only in that the asymptotic variance of $n^{-1/2}S\{\hat{D}(\beta_0)\}$ exceeds that of $n^{-1/2}S^{(F)}\{\tilde{D}(\beta_0)\}$ by the non-negative quantity $Var(U_{mis}) - Var(U_{rec})$, where (a) $Var(U_{mis})$ represents additional uncertainty attributable to observing the c*-observed data rather than the c*-full data, if, when computing $\hat{K}_i(X_i)$, we had used the true value α_0 rather than $\hat{\alpha}$ in formulas (3.3a) and (3.3b), and (b) $Var(U_{rec})$ represents the part of the additional uncertainty that can be recovered by estimating α_0 by $\hat{\alpha}$, even were α_0 known. Define $N_Q(x) = I[X \leq x, \tau = 0]$, $Y(u) = I[X \geq u]$, $M_Q(x) \equiv N_Q(x) - \int_0^x \lambda_Q[u \mid \bar{L}(u),Z]Y(u)du$. $M_Q(x)$ is, by (2.4a), a subject-specific martingale with respect to the filtration $\mathbf{F}(u)$ that records $[c^*,T,Z,\bar{L}(\min(T,u)), \{N_Q(x), 0 \leq x \leq u\}]$. For a random H(u), define $\Phi^Q(H,u,s) \equiv H(u) -$ $\mathcal{L}^Q(H,u,s)$ with $\mathcal{L}^Q(H,u,s) \equiv E\{H(u)Y(u)e^{\alpha_0'W(u)}I[S^*=s]\}/E\{Y(u)e^{\alpha_0'W(u)}I[S^*=s]\}$. Set $\Gamma(H)$ $\equiv \int_0^\infty dM_Q(u)\Phi^Q(H,u,S^*)$.

Theorem 3.4: Under regularity conditions, for $|\beta - \beta_0|$ of $O(n^{-1/2})$, under models (2.2), (3.2), and (2.4a), if $K(X) > \sigma > 0$ for some σ w.p.1

$$n^{-1/2}S\{\hat{D}(\beta_0)\} = n^{-1/2}\Sigma_i U_i + o_p(1)\ , \tag{3.7}$$

$$n^{-1/2}S\{\hat{D}(\beta)\} = n^{-1/2}S\{\hat{D}(\beta_0)\} - \kappa(\beta-\beta_0) + o_p(1)\ , \tag{3.8}$$

$$n^{1/2}(\hat{\beta}-\beta_0) = n^{-1/2}\Sigma_i\kappa^{-1}U_i + o_p(1) \quad , \tag{3.9}$$

$n^{-1/2}S\{\hat{D}(\beta_0)\}$ and $n^{1/2}(\hat{\beta}-\beta_0)$ are asymptotically normal with mean 0 and asymptotic variances $Var(U) = Var(U_f) + Var(U_{mis}) - Var(U_{rec})$ and $\kappa^{-1}Var(U)\kappa^{-1\prime}$; $U = U_f - U_{mis}+U_{rec}$; $U_{mis}=\Gamma(H_{mis})$, with $H_{mis}(u) = U_f/K(u)$ except that $H^{AF2}_{mis}(u) = D^{AF2}(\beta_0)/K(u)$; and $U_{rec} = \rho(U_{mis},U_w) \equiv E[U_{mis}U_w{}'](E[U_wU_w{}'])^{-1}U_w$, where $U_w = \Gamma(W)$ with W the function W(u). [$\rho(A,B)$ is, by definition, the predicted value from the population least squares regression of A on B.] Further $Var[U_{rec}] = E[U_{mis}U_w{}']\{E[U_wU_w{}']\}^{-1}E[U_wU_{mis}{}']$ and $E[U_{mis}U_{rec}{}'] = E[U_{rec}U_{rec}{}']$ where $E[\Gamma(H_1)\Gamma(H_2)'] = E[\int_0^\infty dN_Q(u)\{\mathscr{L}^Q(H_1H_2{}',u,S^*) - \mathscr{L}^Q(H_1,u,S^*)\mathscr{L}^Q(H_2,u,S^*)'\}]$ when $H_1(u)$, $H_2(u)$ are $\mathbf{F}(u)$ predictable. Note $E[U_fU_{mis}{}'] = E[U_fU_{rec}{}']=0$ since U_f is $\mathbf{F}(0)$-predictable. $NC\{\hat{D}(\beta)\}$, the non-centrality parameter of the test based on $S\{\hat{D}(\beta)\}$, is $\{\kappa^{-1}Var(U)\kappa^{-1\prime}\}^{-1}$.

A sketch of the proof of Theorem 3.4 is given in Section (3i). Consistent estimators for the asymptotic variance of $\hat{\beta}$ and $S\{\hat{D}(\beta_0)\}$ are given in Section 3j.2.

3h. <u>Fundamental Identities</u>:

The following identities are fundamental to our results. In particular, as described below, they motivated the estimating function $S\{\hat{D}(\beta)\}$.

$$1/K(t) = 1/K(u) + \int_u^t dx\lambda_Q(x)/K(x) \tag{3.10a}$$

$$\tau/K(t) = 1/K(u) - (1-\tau)/K(t) + \int_u^t dx\lambda_Q(x)/K(x) \tag{3.10b}$$

$$Y(u)\tau/K(X) = Y(u)/K(u) - \int_u^\infty dM_Q(x)/K(x) \tag{3.10c}$$

$$\tau/K(X) = 1 - \int_0^\infty dM_Q(x)/K(x) \tag{3.10d}$$

If $K(X)>\sigma>0$ w.p.1, and (2.4a) holds,

$$E[\tau h(T,Z)/K(X)] = E[h(T,Z)] \quad , \tag{3.10e}$$

and

$$E[h(T,Z,\bar{L}(u))/K(u) \mid \bar{L}(u),Z,Y(u)=1] = E[\tau h(T,Z,\bar{L}(u))/K(X) \mid \bar{L}(u),Z,Y(u)=1] \; . \tag{3.10f}$$

(3.10a) is proved by integration from which (3.10b) follows. (3.10c) and (3.10d) follow immediately from (3.10b). (3.10e) follows by multiplying both sides of (3.10d) by h(T,Z) taking expectations, and noting that $\int_0^t h(T,Z)M_Q(x)/K(x)$ is a mean zero $\mathbf{F}(t)$-adapted martingale since $h(T,Z)/K(x)$ is $\mathbf{F}(x)$-predictable and K(x) is bounded away from zero. (3.10f) follows similarly from (3.10c) except we take conditional expectations.

Since $E[D(\beta_0)]=0$ and $D(\beta_0)$ is a function of T,Z and c^* only, it follows from (3.10e) that $E[\tau D(\beta_0)/K(X)]=0$, proving the fundamental lemma of Section (3b). Thus we would expect, under suitable regularity conditions, that $\hat{\beta}$ solving $0 = S\{\hat{D}(\beta)\}$ will be an asymptotically normal and unbiased estimator of β_0 since $\hat{D}(\beta)$ is an $n^{1/2}$-consistent estimator of $D(\beta)$ and $\hat{K}_i(X_i)$ is an $n^{1/2}$-consistent estimator $K_i(X_i)$. In this paper, we do not specify these regularity conditions in detail.

3.i Sketch of the Derivation of Asymptotic Distribution Theory for $\hat{\beta}$ and $S\{\hat{D}(\beta_0)\}$

Our derivation of Theorem 3.4 is based on the following algebraic identity which is a sample analogue of (3.10d). For any B_i, $\Sigma_i \tau_i B_i/\hat{K}_i = \Sigma_i B_i - \Sigma_i \Psi_i(\hat{H}_i)$ with $\hat{H}_i(u) \equiv B_i/\hat{K}_i(u)$, $\Psi(H) \equiv \int_0^\infty dN_Q(u)\{H(u) - \tilde{\mathscr{L}}^Q(H,u,S^*)\}$ and $\tilde{\mathscr{L}}^Q(H,u,s) \equiv \tilde{E}[H(u)Y(u)e^{\hat{\alpha}' W(u)} I(S^*=s)]/\tilde{E}[Y(u)e^{\hat{\alpha}' W(u)} I(S^*=s)]$. But Robins (1991) shows by expanding around α_0, that, if $B=b(T,Z,c^*)$ and $K(X)>\sigma>0$, $n^{-1/2}\Sigma_i\Psi_i(\hat{H}_i) = n^{-1/2}\Sigma_i\Psi_i(H_i) + o_p(1)$ for $H_i(u) = B_i/K_i(u)$. But, under (2.4a), $n^{-1/2}\Sigma_i\Psi_i(H_i)$ is exactly the Cox partial likelihood score test of the true hypothesis $\theta_0=0$ in the correctly specified PH model for Q that adds $\theta_0'H(u)$ to $\alpha_0'W(u)$ in (3.2). But, Anderson and Gill (1982), Ritov and Wellner (1988), and Begun et al., (1983) show $n^{-1/2}\Sigma_i\Psi_i(H_i) = n^{-1/2}\Sigma_i U_{\theta|\alpha,i} + o_p(1)$ where $U_{\theta|\alpha} \equiv \Gamma(H) - \rho(\Gamma(H),\Gamma(W))$ is the Cox efficient score for θ. Hence, we have proved

Lemma 3.5: If (3.2) holds, $K(X)>\sigma>0$ w.p.1, and $B=b(T,Z,c^*)$, $n^{-1/2}\Sigma_i \tau_i B_i/\hat{K}_i \equiv n^{1/2}\hat{E}[B] = n^{-1/2}\Sigma_i[B_i - U_{\theta|\alpha,i}] + o_p(1)$. Further B_i is uncorrelated with $U_{\theta|\alpha,i}$ since B_i is $\mathbf{F}(0)$-predictable.

We establish (3.7) for the mean model (2.2c) by applying Lemma (3.5) to $n^{-1/2}S(\hat{D}(\beta_0)) \equiv n^{1/2}\hat{E}[D(\beta_0)]$. The key to establishing (3.7) for the PH model (2.2a) is the identity $n^{-1/2}S\{\hat{D}^{PH}(\beta_0)\} = n^{1/2}\hat{E}[\int_0^\infty dM_T(u)\{h(u,Z)-\hat{\mathscr{L}}^{PH}(u,\beta_0,h)\}]$. Using this identity one can show that, under regularity conditions, $n^{-1/2}S\{\hat{D}^{PH}(\beta_0)\} = n^{1/2}\hat{E}\{U_f^{PH}\}+o_p(1)$, although, the proof of this result is non-standard in that $\hat{\mathscr{L}}^{PH}(u,\beta_0,h)$ depends on the data obtained at times past u. Lemma (3.5) is then applied to $n^{1/2}\hat{E}\{U_f^{PH}\}$. A similar argument establishes (3.7) for $n^{-1/2}S\{\hat{D}^{AF1}(\beta_0)\}$.

We establish (3.7) for $n^{-1/2}S\{\hat{D}^{AF2}(\beta_0,r)\}$ with Z dichotomous from the identity $n^{-1/2}S\{\hat{D}^{AF2}(\beta_0,r)\} = n^{1/2}\hat{E}[D^{AF2}(\beta_0,r)] + \hat{E}\{R(\beta_0)\}n^{1/2}\tilde{E}[Z-E(Z)]$. We then apply Lemma (3.5) to $n^{1/2}\hat{E}[D^{AF2}(\beta_0,r)]$ and note $\hat{E}\{R(\beta_0)\} \xrightarrow{P} E[R(\beta_0)]$. Establishing Eq. (3.7) for $n^{-1/2}S\{\hat{D}^{AF2}(\beta_0)\}$ with Z non-dichotomous is discussed in Appendix (3).

Having established (3.7), we can establish (3.8) and (3.9) for the PH model and the mean model by a Taylor series expansion around β_0. Since, for the AF model, $S\{\hat{D}(\beta)\}$ is non-differentiable, one would establish (3.8) and (3.9) using the approach of Pakes and Pollard (1989) and Huber (1981) for non-differentiable m-estimators. κ is shown to be as stated in Theorems 3.3 and 3.4 in Section (4b).

3j. Variance Estimation

3j.1 c*-full data

To fix notation, we review previously derived estimators $\tilde{\text{Var}}\{U_f(\tilde{\beta})\}$ and $\tilde{\kappa}(\tilde{\beta})$ of $\text{Var}(U_f)$ and κ when c*-full data is available. $\tilde{\text{Var}}\{U_f(\beta)\} \equiv \tilde{E}[\tilde{\Omega}(\beta)]$, with $\tilde{\Omega}^{PH}(\beta,h) = \int_0^\infty dN_T(u)\tilde{J}^{PH}(u,\beta,h,h)$ and $\tilde{J}(u,\beta,h_1,h_2) \equiv \tilde{\mathcal{L}}(u,\beta,h_1h_2{}') - \tilde{\mathcal{L}}(u,\beta,h_1)\tilde{\mathcal{L}}(u,\beta,h_2)'$ (Kalbfleisch and Prentice, 1980); $\tilde{\Omega}^{AF1}(\beta,h) = \int_{-\infty}^\infty dN_{\epsilon(\beta)}(u)\tilde{J}^{AF1}(u,\beta,h,h)$; $\tilde{\Omega}^{MN}(\beta) = \tilde{D}^{MN}(\beta)\tilde{D}^{MN}(\beta)'$; and $\tilde{\Omega}^{AF2}(\beta) = \tilde{U}_f^{AF2}(\beta)\tilde{U}_f^{AF2}(\beta)'$, where $\tilde{U}_{f,i}^{AF2}(\beta) \equiv \tilde{U}_{f,i}^{AF2}(\beta,h,\theta) = \tilde{D}_i^{AF2}(\beta,h,\theta) - \sum_{j=1}^n n^{-1}P_{ji}(\beta,h,\theta)$ and $P_{ji}(\beta) \equiv P_{ji}(\beta,h,\theta) = [\int_{-\infty}^\infty dN_{\epsilon_j(\beta)}(u)\{h(u,Z_i) - \tilde{\mathcal{L}}^{AF2}(u,\beta,h)\}] - \int_{-\infty}^\infty du I\{\epsilon_j(\beta)>u\}I\{\mu_i(\beta)>u\}[\theta(u,Z_i) - \tilde{\mathcal{L}}^{AF2}(u,\beta,\theta)]$ since, by the independence of ϵ and $Z, n^{-1}\sum_{j=1}^n P_{ji}(\beta_0)$ is consistent for $E[D^{AF2}(\beta_0) \mid Z_i]$. $\tilde{\text{Var}}\{U_f^{AF2}(\beta)\}$ is a new estimator.

If Z is dichotomous (0,1), we use $\tilde{\text{Var}}(U_f^{AF2}(\beta)) \equiv \tilde{\text{Var}}\{U_f^{AF2}(\beta,r)\} = \tilde{E}[\{R(\beta) - \tilde{E}[R(\beta)]\}^2]\tilde{E}[\{Z-\tilde{E}(Z)\}^2]$.

Next define $\tilde{\kappa}^{PH}(\beta,h) = \tilde{E}[\int_0^\infty dN_T(u)\tilde{J}^{PH}(u,\beta,h,h_{op}^{PH})]$ and $\tilde{\kappa}^{MN}(\beta,h) = \tilde{E}[h(Z)\partial g(Z,\beta)/\partial\beta']$. Since h_{op}^{AF} and θ_{op}^{AF} depend on the derivative of the unknown density $\lambda_\epsilon(u)$ which we may wish to avoid estimating, we set $\tilde{\kappa}^{AF1}(\beta)$ and $\tilde{\kappa}^{AF2}(\beta)$ to be the matrix of symmetric partial numerical derivatives of the corresponding $n^{-1}S^{(F)}\{\tilde{D}(\beta)\}$ with respect to β' based on a step size of $O(n^{-1/2})$ (Pakes and Pollard, 1989; Robins and Tsiatis, 1991). Eq. (3.6) implies $\tilde{\kappa}(\tilde{\beta})$ is consistent for κ.

3j.2 c*-Observed Data

We shall provide consistent estimators $\hat{\text{Var}}\{U(\hat{\beta})\}$ and $\hat{\kappa}(\hat{\beta})$ of Var(U) and κ depending on $\hat{\beta}$. In particular, the estimate $[\hat{\text{Var}}\{U^{AF2}(0)\}]^{1/2}$, for dichotomous Z and a given $r \equiv r(\Delta(0),X^*(0)) \equiv r[\Delta,\ell n(X^*)]$, is the denominator of the test statistic $\psi(r,w,c^*)$ of Section (3b) used in the simulations in Section (3c).

It follows from the decomposition of Var(U) in Section 3g that $\hat{\text{Var}}\{U(\hat{\beta})\}$ can be based on consistent estimators of $\text{Var}(U_f)$ and $E[\Gamma(H_1)\Gamma(H_2)']$ for $H_1(u),H_2(u)\in\{H_{mis}(u),W(u)\}$.

A consistent estimator $\hat{\text{Var}}\{U_f(\hat{\beta})\}$ of $\text{Var}(U_f)$ is $\hat{\text{Var}}[U_f(\hat{\beta})] \equiv \hat{E}\{\hat{\Omega}(\hat{\beta})\}$ where $\hat{\Omega}(\beta)$ is defined like $\tilde{\Omega}(\beta)$ except $\hat{J}$ replaces $\tilde{J}$ (by replacing $\tilde{\mathcal{L}}$ with $\hat{\mathcal{L}}$), n^{-1} is replaced by $n^{-1}\tau_j/\hat{K}_j$ in the definition of $\tilde{U}_{f,i}^{AF2}(\beta)$, $P_{ji}(\beta)$ is unchanged and, for dichotomous Z, $\hat{\text{Var}}[U_f^{AF2}(\beta,r)] = \hat{E}[\{R(\beta)-\hat{E}[R(\beta)]\}^2]\,\tilde{E}[\{Z-\tilde{E}(Z)\}^2]$. $\hat{\kappa}^{PH}(\beta)$ and $\hat{\kappa}^{MN}(\beta)$ are defined as above with $\hat{E}$ and $\hat{J}$ replacing $\tilde{E}$ and $\tilde{J}$. Finally $\hat{\kappa}^{AF1}(\beta)$ and $\hat{\kappa}^{AF2}(\beta)$ are numerical partial derivative matrices of the corresponding $n^{-1}S\{\hat{D}(\beta)\}$.

$E[\Gamma(H_1)\Gamma(H_2)']$ can be consistently estimated by $\tilde{E}[\int_0^\infty dN_Q(u)\,\{\hat{\mathcal{L}}^Q(\hat{H}_1\hat{H}_2{}',u,S^*) - \hat{\mathcal{L}}^Q(\hat{H}_1,u,S^*)\hat{\mathcal{L}}^Q(\hat{H}_2,u,S^*)'\}$ where $\hat{H}_1(u)$ and $\hat{H}_2(u)$ are consistent for $H_1(u)$ and $H_2(u)$; $\hat{\mathcal{L}}^Q(H,u,s) = B(H,u,s)/B(\mathbf{1},u,s)$ with $\mathbf{1}(u)=1$, and $B(H,u,s) =$

$\sum_{i=1}^{n}\{\tau_i\hat{K}_i(u)/\hat{K}_i(X_i)\}Y_i(u)H_i(u)e^{\hat{\alpha}'W_i(u)}I[S_i^*=s]$. Note $\hat{\mathscr{L}}^Q(H,u,s)$ is consistent for $\mathscr{L}^Q(H,u,s)$ by Eq. (3.10f) and the weak law of large numbers.

Thus, it only remains to provide consistent estimators $\hat{H}_{mis}(u,\hat{\beta})$ for the $H_{mis}(u)$ as follows. $\hat{H}^{AF2}_{mis}(u,\hat{\beta}) = \hat{D}^{AF2}(\hat{\beta})/\hat{K}(u)$. $H^{MN}_{mis}(u,\hat{\beta}) = \hat{D}^{MN}(\hat{\beta})/\hat{K}(u)$. $\hat{H}^{PH}_{mis}(u,\hat{\beta}) \equiv \{\hat{K}(u)\}^{-1}\hat{U}^{PH}_f(\hat{\beta})$ where $\hat{U}^{PH}_f(\beta) \equiv \hat{U}^{PH}_f(\beta,h) = \int_0^\infty [dN_T(u) - I(X^*\geq u)e^{\beta'Z}\hat{\lambda}_0(u)du]\{h(u,Z) - \hat{\mathscr{L}}^{PH}(u,\beta,h)\}$ and $\hat{\lambda}_0(u) \equiv \hat{\lambda}_0(u,\beta) \equiv \hat{E}[dN_T(u)]/\hat{E}[e^{\beta'Z}I(X^*\geq u)]$. $\hat{H}^{AF1}_{mis}(u,\hat{\beta}) = \{\hat{K}(u)\}^{-1}\,\hat{U}^{AF1}_f(\hat{\beta})$, where $\hat{U}^{AF1}_f(\beta) = \int_{-\infty}^{\infty}[dN_{\epsilon(\beta)}(u) - I[\nu(\beta)\geq u]\hat{\lambda}_{\epsilon(\beta)}(u)du]\{h(u,Z) - \hat{\mathscr{L}}^{AF1}(u,\beta,h)\}$ with $\hat{\lambda}_{\epsilon(\beta)}(u) = \hat{E}[dN_{\epsilon(\beta)}(u)]/\hat{E}[I\{\nu(\beta)\geq u\}]$.

3k. Efficiency of $\hat{\beta}$

Write $\hat{\beta}^* \equiv \hat{\beta}^*(h)$ for an estimator $\hat{\beta}(h)$ solving $S\{\hat{D}(\beta,h)\}=0$ in which the components of $W^{mis}(u) \equiv E[H_{mis}(u) \mid \bar{L}(u),Z,Y(u)=1]$ are included as additional covariates in a correctly specified model (3.2). $\hat{\beta}^*$ is not a feasible estimator since it depends on unknown population quantities. Hence, given preliminary consistent estimates $\hat{\beta}$, $\hat{K}(u)$, $\hat{H}_{mis}(u,\hat{\beta})$, define for each u separately, $\hat{W}^{mis}(u)$ to be a subject's predicted value from a non-parametric (e.g., kernel) regression of $[\hat{K}(u)\tau\hat{H}_{mis}(u,\hat{\beta})/\hat{K}(X)]$ on $\bar{L}(u)$ and Z given $Y(u)=1$. Eq. (3.10f) implies $\hat{W}^{mis}(u)$ is consistent for $W^{mis}(u)$ under appropriate smoothness and regularity conditions guaranteeing the consistency of kernel regression estimators. Let $\hat{\hat{\beta}}^*(h)$ be a feasible second stage estimate in which the components of $\hat{W}^{mis}(u)$ are included as additional covariates in a correctly specified model (3.2). In Section (4c), we prove

Theorem 3.5: Subject to regularity conditions, under the semiparametric model defined by a model (2.2), (2.4a) and a particular model (3.2), the limiting distributions of $\hat{\beta}^*(h)$ and $\hat{\hat{\beta}}^*(h)$ are identical with asymptotic variance less than or equal to that of $\hat{\beta}(h)$ with $\hat{K}_i$ used in calculating $\hat{\beta}(h)$ based on any correctly specified model (3.2). Further, there exists h_{eff} such that the asymptotic variance of $\hat{\beta}^*(h_{eff})$ equals the semiparametric variance bound based on the c*-observed data (2.1). Theorem (3.5) also holds for the estimators $\hat{\beta}^{AF2}$ with (h,θ) replacing h and (h_{eff},θ_{eff}) replacing h_{eff}. [Also, if, in computing $\hat{\beta}(h)$, we replaced $\hat{K}_i$ by an appropriate completely non-parametric estimate, then, even without imposing (3.2), any $\hat{\beta}(h)$ would be asymptotically equivalent to $\hat{\beta}^*(h)$.]

Remark: Since $\bar{L}(u)$ is the history of a complex process up to u, the "feasible estimators" $\hat{\hat{\beta}}^*(h)$ are largely of theoretical interest, since, non-parametric estimation of $W^{mis}(u)$ is not practical due to the "curse of dimensionality." Thus, in practice, $\hat{W}_{mis}(u)$ would be replaced by the predicted value from the fit of a regression model.

4. Semiparametric Efficiency and Estimation in Missing Data Models

4a. Main Theorems

In Theorem 4.1, we provide representations for (a) the efficient score and (b) the influence function of a regular, asymptotically linear estimator in an arbitrary (discrete time) semiparametric model with the data missing at random. We specialize to the case of monotone missingness in Theorem (4.2) and extend our results to continuous time censoring processes in Theorem (4.3). Finally, in Theorems (4.4)-(4.6), we specialize Theorem (4.3) to the models (2.2a)-(2.2c). We start with a review of the theory of semiparametric efficiency bounds that borrows heavily from the survey paper of Newey (1990) and the monograph of Bickel, Klassen, Ritov, and Wellner (BKRW, 1991).

Suppose the data consists of n independent copies of an observed random variable V. Let $Lik(\beta,\theta; V)$ be the likelihood for a single subject in a semiparametric model indexed by a parameter $\beta \epsilon R^p$ of interest and a nuisance parameter θ taking values in some infinite dimensional set. Let (β_0,θ_0) index the distribution generating V. Define a regular parametric submodel to be a regular (fully) parametric model with parameters (β,η) and likelihood $Lik(\beta,\eta\, ; V)$ with true values β_0,η_0, where the "sub" prefix refers to the fact that for each η the distribution $Lik(\beta,\eta\, ; V)$ equals a distribution $Lik(\beta,\theta\, ; V)$ allowed by the semiparametric model. Define the nuisance tangent space Λ to be the mean square closure of the set of all random vectors bS_η, where S_η is the score for η in some regular parametric submodel (usually, $S_\eta = \partial \ell n L(\beta_0,\eta_0\, ; V)/\partial\eta$) and b is a conformable constant matrix with p rows; i.e. $\Lambda = \{A \epsilon R^p : E[\, \| A \|^2] < \infty$, and there exists $b_j\, S_{\eta,j}$ with $\lim_{j\to\infty} E[\, \| A - b_j S_{\eta,j} \|^2] = 0\}$, where each b_j is a matrix of constants and $\| A \| = A'A$. We shall consider Λ as a subset of the Hilbert space of p×1 random vectors H with inner product $E[H'H]$. In our examples, Λ is a linear subspace. The projection of any vector H on a closed linear space, such as Λ, exists and is the unique vector $\Pi(H \mid \Lambda)$ in Λ satisfying $E[\{H-\Pi(H \mid \Lambda)\}'A] = 0$ for all $A \epsilon \Lambda$. Π is the projection operator. Henceforth we restrict attention to random vectors H for which $E[H'H]<\infty$. The semiparametric variance bound for regular estimators of β_0 is, by definition, the supremum of the Cramer-Rao bounds for β_0 over all regular parametric submodels and equals the inverse of the variance of $S_{eff} \equiv S_\beta - \Pi[S_\beta \mid \Lambda]$ where S_β is the score for β (usually, $S_\beta = \partial \ell n L(\beta_0,\theta_0\, ; V)/\partial\beta$). S_{eff} is called the efficient score (Begun et al. 1983) and $Var[S_{eff}]$ is called the SIB.

We shall need the following definitions.

Def: (a) A test statistic $n^{-1/2}S(\beta)$ or (b) an estimator $\hat{\beta}$ is asymptotically linear at β_0 in a semiparametric model $Lik(\beta,\theta\, ; V)$ with influence function $D=d(V)$ if (a) $n^{-1/2}S(\beta_0) = n^{-1/2}\Sigma_i D_i + o_p(1)$ or (b) $n^{1/2}(\hat{\beta}-\beta_0) = n^{-1/2}\Sigma_i D_i + o_p(1)$, where $E(D_i) = 0$ and $E(D_i'D_i) < \infty$ whatever be θ_0. D_i may depend on (β_0,θ_0). We shall impose the additional condition that $E_{\beta,\eta}[D_i]$ is continuous at (β_0,η_0) for each regular parametric submodel.

Def: A local data generating process (LDGP) at (β_0,θ_0) in a regular parametric submodel is one in which the data is generated according to $Lik(\beta_n,\eta_n\, ; V)$ with $n^{1/2}(\beta_n-\beta_0)$ and $n^{1/2}(\eta_n-\eta_0)$ bounded.

Def: (a) A test statistic $n^{-1/2}S(\beta_0)$ or (b) an estimator is regular at β_0 if (a) the limiting distribution of $n^{-1/2}S(\beta_0)$ is the same for all LDGP $Lik(\beta_0,\eta_n\, ; V)$ in all regular parametric submodels or (b) the distribution of $n^{1/2}(\hat{\beta}-\beta_0)$ is the same for all LDGP Lik$(\beta_n,\eta_n;V)$ in all regular parametric submodels. The limiting distribution may depend on (β_0,θ_0).

Def: For any set $\mathscr{F}$ of random variables, let $\mathscr{F}_0$ be the subset with mean 0.

Lemma 4.1: In any semiparametric model, the influence function of any regular asymptotic linear (RAL) test of the true hypothesis $\beta=\beta_0$ is in $\Lambda_0^{\perp} \equiv (\Lambda^{\perp})_0$, where "$\perp$" denotes an orthogonal complement. The influence function of any RAL estimator of β_0 is in $\Lambda_{0*}^{\perp} \equiv \{A\in\Lambda_0^{\perp}; E[AS_\beta'] = I_{p\times p}\}$. Theorem (2.2) in Newey (1990a) implies Lemma 4.2 for RAL estimators. Theorem A.1 in Appendix 1 implies the lemma for RAL tests.

We now specialize the above general results to "full" and "missing data" models. Henceforth let $V = (V^{(1)},\cdots,V^{(K)})'$ be a multivariate random variable with each $V^{(k)}$ univariate and

$$L^{(F)}(\beta,\theta;V) \tag{4.1}$$

be the likelihood for a single subject when V is fully observed in a semiparametric model indexed by $\beta\in R^p$ and an infinite dimensional nuisance parameter θ and let $S_\beta^{(F)}$, $\Lambda^{(F)}$, and $S_{eff}^{(F)}$ be the score for β, the nuisance tangent space, and the efficient score in the full data model. Suppose now V may not be fully observed and let $R = (R^{(1)},\cdots,R^{(K)})$ be a random variable such that $R^{(k)}=1$ if $V^{(k)}$ is observed and $R^{(k)}=0$ otherwise. R has 2^K possible realizations (r). Let $V_{(r)}$ be the vector of observed components of V when R=(r) so that the observable random variables are $(R,V_{(R)})$. Suppose the data is missing at random (Rubin, 1976) i.e.,

$$\text{pr}[R=r \mid V] = \text{pr}[R=r \mid V_{(r)}] \equiv \pi(r,V_{(r)}) \equiv \pi(r) \tag{4.2}$$

and that

$$\pi(\mathbf{1}) > \sigma > 0 \text{ with probability one} \tag{4.3}$$

where $\mathbf{1}$ is the K-vector of 1's. Often it is assumed

$$\pi(r,V_{(r)})\in\{\pi(r,V_{(r)};\gamma) : \gamma\in \boldsymbol{\Upsilon}\} \tag{4.4}$$

where, for each γ, $\Pi(r,V_{(r)};\gamma)$ is a density for pr[R=r | V] satisfying (4.2) and $\boldsymbol{\Upsilon}$ may be infinite dimensional. Henceforth let S_β,Λ, and S_{eff} be the score for β, the nuisance tangent space, and the efficient score in the missing data model defined by (4.1)-(4.4) based on the data $(R,V_{(R)})$.

Define the tangent space for model (4.4) to be $\Lambda^{(3)} = \{A^{(3)}\in R^p$; there exists $b_j\, S_{\psi,j}$ with $\lim_{j\to\infty} E[\,\| A^{(3)}-b_jS_{\psi,j}\|^2] = 0\}$ where S_ψ is the score at the truth for a regular parametric submodel $\pi(r,V_{(r)},\psi)$ of the model $\pi(r,V_{(r)},\gamma)$ satisfying $\pi(r,V_{(r)},\psi_0) = \pi(r,V_{(r)})$ for some ψ_0 and the b_j are constant matrices. Define $\Lambda^{(2)} \equiv \{A^{(2)} = a^{(2)}(R,V_{(R)})\in R^P$; $E[A^{(2)} \mid V] = 0\}$.

In Appendix 2, we prove

Lemma 4.2: (a) $\Lambda^{(3)} \subseteq \Lambda^{(2)}$. (b) If the model (4.4) is completely non-parametric in the sense that it is unrestricted except for the condition $\Sigma_r \pi(r, V_{(r)}; \gamma) = 1$, then $\Lambda^{(2)} = \Lambda^{(3)}$.

Let $H = h(V)$ and $D = d(V)$ be generic functions of V. Define the operators $\mathbf{g}(H)$, $\mathbf{m}(H)$, and $\mathbf{u}(H)$ by $\mathbf{g}(H) = \Sigma_r I(R=r)E[H \mid V_{(r)}]$, $\mathbf{m}(H) = \Sigma_r \pi(r)E[H \mid V_{(r)}]$, $\mathbf{u}(H) = \pi(\mathbf{1})^{-1} I(R=\mathbf{1})H$ where the sums are over the 2^K possible value of r. Here and throughout bold lower case letters represent operators. Define $\Lambda^{(1)} = \{\mathbf{g}(A^{(F)}) : A^{(F)} \in \Lambda^{(F)}\}$. In Appendix 2, we prove

Theorem (4.1): In the missing data model defined by (4.1)-(4.4), (a) $S_\beta = \mathbf{g}(S_\beta^{(F)})$ is the score for β, (b) $\Lambda^{(1)}$ and $\Lambda^{(3)}$ are mutually orthogonal closed linear spaces; (c) $\Lambda = \Lambda^{(1)} \oplus \Lambda^{(3)}$ is the nuisance tangent space where $\oplus$ denotes the direct sum of two spaces, (d) $\Lambda_0^\perp = \{A_0^\perp \equiv A_0^\perp(D, A^{(2)}) \equiv \mathbf{u}(D) + A^{(2)} - \Pi[\mathbf{u}(D) + A^{(2)} \mid \Lambda^{(3)}] ; D \in \Lambda_0^{(F),\perp}, A^{(2)} \in \Lambda^{(2)}\}$ where $\Lambda_0^{(F),\perp} \equiv (\Lambda^{(F),\perp})_0$. (e) $E[A_0^\perp(D, A^{(2)}) S_\beta{}'] = E[D S_\beta^{(F)}{}'] = E[D S_{eff}^{(F)}{}']$ for $D \in \Lambda_0^{(F),\perp}$. Further $\Lambda_{0*}^\perp \equiv \{A \in \Lambda_0^\perp ; E[A S_\beta{}'] = I_{p\times p}\}$ is the set $\{\mathbf{u}(D) + A^{(2)} - \Pi[\mathbf{u}(D) + A^{(2)} \mid \Lambda^{(3)}] ; D \in \Lambda_{0*}^{(F),\perp}, A^{(2)} \in \Lambda^{(2)}\}$, where $\Lambda_{0*}^{(F),\perp} \equiv \{D \in \Lambda_0^{(F),\perp} ; E[D S_\beta^{(F)}{}'] = I_{p\times p}\}$. (f) The efficient score $S_{eff} \equiv S_\beta - \Pi[S_\beta \mid \Lambda]$ is given by $\mathbf{u}(D_{eff}) - \Pi[\mathbf{u}(D_{eff}) \mid \Lambda^{(2)}]$ with D_{eff} the unique D in $\Lambda^{(F),\perp}$ solving $\Pi[\mathbf{m}^{-1}(D) \mid \Lambda^{(F),\perp}] = S_{eff}^{(F)}$, where $\mathbf{m}^{-1}(D)$ is the unique H solving $\mathbf{m}(H) = D$. (g) $\Pi[\mathbf{u}(D) \mid \Lambda^{(2)}] = \mathbf{u}(D) - \mathbf{g}[\mathbf{m}^{-1}(D)]$, and thus $S_{eff} = \mathbf{g}[\mathbf{m}^{-1}(D_{eff})]$. We now indicate the importance of this rather abstract theorem.

By Lemma (4.1), it follows that part e (d) of Theorem (4.1) says that the influence function of any RAL estimator (test) in the missing data model (4.1)-(4.4) is obtained by first applying the operator $\mathbf{u}$ to the influence function of an RAL estimator (test) in the full data model $L^{(F)}(\beta, \theta ; V)$, adding an arbitrary element of $\Lambda^{(2)}$, and finally subtracting off the projection on the tangent space $\Lambda^{(3)}$ of the model (4.4) for the missingness process. Part f of the Theorem says that the efficient score for β_0 does not depend on $\Lambda^{(3)}$ and thus on the model (4.4) for the missingness process.

In a general missing data model, the characterization of S_{eff} given in part f of the theorem may be of little practical help since, for a given H, $\mathbf{m}^{-1}(H)$ and $\Pi[\mathbf{u}(H) \mid \Lambda^{(2)}]$ may not exist in closed form. However, $\Pi[\mathbf{u}(H) \mid \Lambda^{(2)}]$ and $\mathbf{m}^{-1}(H)$ exist in closed form when there is a monotone missing data pattern, where a monotone missing data pattern is defined as follows. Suppose we have a sequence of sets $S_0, S_1, \cdots, S_M$ satisfying $\phi = S_0 \subsetneq S_1 \subsetneq S_2 \cdots \subsetneq S_M \equiv \{1, \cdots, K\}$. For $1 \le m \le M+1$, let $\bar{V}_m$ be the vector comprised of the components $V^{(k)}$ of V for $k \in S_{m-1}$. For $0<m<M+1$, let $R_m = 1$ if $\bar{V}_{m+1}$ is fully observed and $R_m=0$ otherwise. Let $R_0=1$. Note $R_m=1$ if $R_{m+1}=1$. We say that the missing data pattern is monotone if whenever $R_{m+1}=0$ and $R_m=1$, $\bar{V}_{m+1}$ constitutes all the observed elements of V. Let $\pi_0=1$ and, for $m>1$, $\pi_m = P[R_m=1 \mid R_{m-1}=1, \bar{V}_m]$. Set $\bar{\pi}_m = \prod_{j=1}^{m} \pi_m$ and $\bar{\pi}_0=1$.

Theorem 4.2: If missingness is monotone in the missing data semiparametric model (4.1)-(4.4),

$E[D \mid \bar{V}_m] = E[D \mid \bar{V}_m, R_{m-1}=1]$; $\mathbf{g}(D) = R_M D + \sum_{m=1}^{M} (R_{m-1}-R_m) E[D \mid \bar{V}_m]$;

$\mathbf{m}(D) = \bar{\pi}_m D + \sum_{m=1}^{M} (1-\pi_m)\bar{\pi}_{m-1} E[D \mid \bar{V}_m]$; $\mathbf{u}(D) = D + \sum_{m=1}^{M} (R_m - \pi_m R_{m-1})\bar{\pi}_m^{-1} D$;

$\mathbf{m}^{-1}(D) = D + \mathbf{v}(D)$, where $\mathbf{v}(D) = \sum_{m=1}^{M} (1-\pi_m)\bar{\pi}_m^{-1}\{D - E[D \mid \bar{V}_m]\}$;

$\Lambda^{(2)} = \{A^{(2)} = \sum_{m=1}^{M} (R_m - \pi_m R_{m-1}) a(\bar{V}_m)$ for some $a(\bar{V}_m) \epsilon R^P\}$;

$\Pi[\mathbf{u}(D) \mid \Lambda^{(2)}] = \sum_{m=1}^{M} (R_m - \pi_m R_{m-1})\bar{\pi}_m^{-1} E[D \mid \bar{V}_m]$; and D_{eff} is the unique $D \epsilon \Lambda^{(F),\perp}$ solving

$$S_{eff}^{(F)} = D + \Pi[\mathbf{v}(D) \mid \Lambda^{(F),\perp}] \tag{4.5}$$

<u>Continuous Time Models</u>:

An example of monotone missingness occurs when missingness is solely due to right censoring by a random variable Q^* where Q^* has a discrete sample space $\{t_1, t_2, \cdots, t_M, t_{M+1}\}$ with $\bar{V}_{M+1} \equiv V$ being the full-data and $\bar{V}_m$ the observed data for a subject with $Q^* = t_m$. Suppose now $V \equiv \bar{V}(C) = \{V(t)\,;\, 0 \le t \le C\}$ is a continuous time process, Q^* is a continuous censoring variable, and C may be random. We obtain a continuous time version of Theorem (4.2) by taking limits as $\Delta t \equiv t_{m+1} - t_m \to 0$. Specifically we assume a full data model (4.1) for $V \equiv \bar{V}(C)$ but we only observe for each subject

$$C, \min(Q^*, C), \tau = I(C \le Q^*), \bar{V}\{\min(Q^*, C)\} \ . \tag{4.6}$$

We assume

$$\lambda_{Q^*}[u \mid \bar{V}(C)] = \lambda_{Q^*}[u \mid \bar{V}(u)\} \ , \tag{4.2$'$}$$

$$K^*(C) > \sigma > 0 \text{ with probability one } , \tag{4.3$'$}$$

and

$$\lambda_{Q^*}[u \mid \bar{V}(u)] \epsilon \{\lambda_{Q^*}[u \mid \bar{V}(u)\,;\gamma]\,;\gamma \epsilon \boldsymbol{\Upsilon}\} \ . \tag{4.4$'$}$$

Here $K^*(t) = \exp\{-\int_0^t \lambda_{Q^*}[u \mid \bar{V}(u)]du\}$. A continuous time version of Theorem (4.2) follows by making the formal identifications; $(1-\pi_m) = \lambda_{Q^*}(t_m \mid \bar{V}(t_m))\,\Delta t + o(\Delta t)\,;\, \bar{\pi}_m = K^*(t_m), R_m = 1 - N_{Q^*}(t_m)$, where $N_{Q^*}(t) = I[Q^* \le t]$. Specifically, define $M_{Q^*}(u) = N_{Q^*}(u) - \int_0^u \lambda_{Q^*}[t \mid \bar{V}(t)] Y^*(t) dt$, with $Y^*(t) = I[Q^* \ge t]$. Then we have

<u>Theorem 4.3</u>: In the semiparametric model defined by (4.1), (4.2$'$)-(4.4$'$), Theorem (4.1) and Eq. (4.5) continue to hold with $D = d\{\bar{V}(C)\}$, $E[D \mid \bar{V}(u), Y^*(u)=1] = E[D \mid \bar{V}(u)]$,

$\mathbf{g}(D) = \tau D + (1-\tau) E[D \mid \bar{V}(Q^*)]$;

$\mathbf{m}(D) = K^*(C) D + \int_0^C du\, E[D \mid \bar{V}(u)]\{K^*(u)\} \lambda_{Q^*}[u \mid \bar{V}(u)]$

$\mathbf{v}(D) = \int_0^C du \lambda_{Q^*}[u \mid \bar{V}(u)]\{K^*(u)\}^{-1}\{D - E[D \mid \bar{V}(u)]\}$

$\mathbf{m}^{-1}(D) = D+\mathbf{v}(D)$, $\mathbf{u}(D) \equiv \tau D/K^*(C) = D - \int_0^C dM_{Q^*}(u)\{K^*(u)\}^{-1}D$

$\Lambda^{(2)} = \{A^{(2)} \equiv \int_0^C dM_{Q^*}(u)a\{\bar{V}(u)\}\ ;\ a\{\bar{V}(u)\}\in R^p\}$

$\Pi[\mathbf{u}(D) \mid \Lambda^{(2)}] = -\int_0^C dM_{Q^*}(u)\{K^*(u)\}^{-1}E[D \mid \bar{V}(u)]$;

$\Lambda^{(3)}$ is the closed linear span of the set $\{A^{(3)} \equiv bS_\psi\}$ where S_ψ is the score for a regular parametric submodel $\lambda_{Q^*}[u \mid \bar{V}(u);\psi]$ for model (4.4′) and b is a conformable matrix with p rows. Typically $S_\psi = \frac{\partial \ell n}{\partial \psi}\Big[\{\lambda_{Q^*}[Q^* \mid \bar{V}(Q^*),\psi]\}^{1-\tau}\exp\Big\{-\int_0^{\min(Q^*,C)}\lambda_{Q^*}[t \mid \bar{V}(t),\psi]\Big\}\Big]_{\psi=\psi_0}$.

We have as yet specified neither the full data $V=\bar{V}(C)$ or the semiparametric model (4.1). Let $C=c^*$ and, for $u\le c^*$, $\bar{V}(u) = \{c^*,Z,\bar{L}(\min(T,u)), TI(T<u), I(T<u)\}$ so that $V = \bar{V}(c^*)$ is the c^*-full data (3.4). Now define $Q^* = Q$ if $Q<X^*$, where $X^* \equiv \min(T,c^*)$, and $Q^* = \infty$ otherwise, so once a subject has failed un-Q-censored (i.e., $T<Q$), he cannot be Q^* censored (since $Q^* = \infty$). Then $\lambda_{Q^*}[u \mid \bar{V}(u)] = \lambda_Q[u \mid \bar{L}(u), Z]$ and $K^*(u) = K(u)$ for $u\le X^*$; $\lambda_{Q^*}[u \mid \bar{V}(u)]=0$ and $K^*(u) = K(X^*)$ for $X^*\le u<c^*$; and the observed data (4.6) is the c^*-observed data (2.1). Further censoring mechanism (A.2′) is equivalent to the censoring mechanism (2.4d) rather than (2.4a). Hence, we can restate Theorem (4.3) as follows

Theorem (4.4): With V and $\bar{V}(u)$ as just defined, if (4.1), (2.4d), (4.3′) and model (3.2) for censoring by Q hold, then Theorem (4.1) and Eq. (4.5) remain true with

$E[D \mid \bar{L}(u),Z,Y(u)=1] = E[D \mid \bar{L}(u),Z,X^*>u]$

$\mathbf{g}(D) = \tau D + (1-\tau)E[D \mid \bar{L}(Q),Z,X^*>Q]$,

$\mathbf{m}(D) = K(X^*)D + \int_0^{X^*} du\lambda_Q[u \mid \bar{L}(u),Z]K(u)E[D \mid \bar{L}(u),Z,Y(u)=1]$, $\mathbf{m}^{-1}(D) = D + \mathbf{v}(D)$,

$\mathbf{v}(D) = \int_0^{X^*} du\lambda_Q[u \mid \bar{L}(u),Z]\{K(u)\}^{-1}\{D-E[D \mid \bar{L}(u),Z,Y(u)=1]\}$,

$\Lambda^{(2)} = \{A^{(2)} \equiv \int_0^\infty dM_Q(u)a(\bar{L}(u),Z)\}$,

$\Pi[\mathbf{u}(D) \mid \Lambda^{(2)}] = -\int_0^\infty dM_Q(u)\{K(u)\}^{-1}E[D \mid \bar{L}(u),Z,Y(u)=1]$,

$\mathbf{u}(D) = D - \int_0^\infty dM_Q(u)\{K(u)\}^{-1}D$, and

$\Lambda^{(3)} = \{A^{(3)} = b_1\Gamma(W) + b_2\int_0^\infty dM_Q(u)r(u,S^*)\}$ where $r(u,S^*)$ is any vector-valued function and b_1 and b_2 are constant conformable matrices with p rows. This characterization of $\Lambda^{(3)}$ follows from the results of Ritov and Wellner (1988) on the PH model.

4b. Semiparametric Models (2.2a)-(2.2c)

We now consider the semiparametric models for $V = \bar{V}(c^*)$ induced by (2.2a)-(2.2c). The semiparametric missing data model for the c^*-observed data (2.1) defined by a model (2.2a)-(2.2c) and censoring mechanism (2.4d) can be shown to be identical to that defined by the corresponding model (2.2a)-(2.2c) and (2.4a) in the sense that, for each β, the allowable distributions for the c^*-observables (2.1) are the same. Hence, assuming (2.4d) rather than (2.4a) can make no inferential difference. In particular, S_{eff} will be the same function of the c^*-observables and their joint distribution. Hence, one can calculate S_{eff}

under (2.4d) and then re-express S_{eff} in terms of the c^*-observables and their distribution to obtain S_{eff} under (2.4a). The following Theorem can be proved using results in Chamberlain (1987), Ritov and Wellner (1988), and Robins and Rotnitzky (1992).

<u>Theorem 4.5</u>: In the "full-data" semiparametric model for the c^*-full data V induced by (2.2a)-(2.2c), (a) $\Lambda^{(F)} = \Lambda_1^{(F)} \oplus \Lambda_2^{(F)} \oplus \Lambda_3^{(F)}$; (b) $\Lambda_1^{(F)}, \Lambda_2^{(F)}$, and $\Lambda_3^{(F)}$ are mutually orthogonal closed subspaces; (c) $\Lambda_1^{(F)} = \{A_1 = a_1(Z)\,;\, E[A_1]=0\}$; (d) $\Lambda_3^{(F)} = \{A_3 = a_3(V)\,;\, E[A_3 \mid Z, X^*, \Delta]=0\}$ except $\Lambda_3^{(F),MN} = \{A_3 = a_3(V)\,;\, E[A_3 \mid Z, \Delta]=0\}$; (e) $\Lambda_2^{(F),MN} = \{A_2^{(F)} = r(\varsigma, Z)\,;\, E[A_2^{(F)} \mid Z]=0,$ $E[\varsigma A_2^{(F)\prime} \mid Z]=0\}$; $\Lambda_2^{(F),AF} = \{ \int_{-\infty}^{\infty} r(u) dM_\epsilon(u) \}$; and $\Lambda_2^{(F),PH} = \{ \int_0^\infty r(u) dM_T(u) \}$; (f) $\Lambda_0^{(F),\perp} =$ $\{U_f\}$ with $U_f^{AF} \equiv U_f^{AF1}$. [Note, however, $\{U_f^{AF1}\} = \{U_f^{AF2}\}$ by Lemma 3.4.]; (g) $S_{eff}^{(F)} = U_f(h_{op})$.

To use Eq. (4.5) to find D_{eff}, we need to be able to compute $\Pi(D \mid \Lambda_0^{(F),\perp})$ for $D=d(V)$. One can prove the following using the definition of a projection.

<u>Theorem 4.6</u>: $\Pi(D \mid \Lambda_0^{(F),\perp,AF}) = U_f^{AF1}(h)$ with $h(u,Z) = \mathbf{r}\{d^{*,AF}(u,Z)\}$ with $d^{*,AF}(u,z) =$ $E[D \mid \Delta=1, \epsilon=u, Z]$ if $u<\mu$ and $d^{*,AF}(u,Z) = E[D \mid \Delta=0, \epsilon>u, Z]$ if $u \geq \mu$, where $\mathbf{r}^{AF}\{g(u,Z)\} \equiv$ $g(u,Z) - E[g(\epsilon,Z) \mid Z, \epsilon>u]$ as in Ritov and Wellner (1988).

$\Pi[D \mid \Lambda_0^{(F),\perp,PH}] = U_f^{PH}(h)$ with $h(u,Z) = \mathbf{r}^{PH}\{d^{*,PH}(u,Z)\}$, with $d^{*,PH}(u,Z) =$ $E[D \mid \Delta=1, T=u, Z]$ if $u<c^*$ and $d^{*,PH}(u,Z) = E[D \mid \Delta=0, T>c^*, Z]$ if $u \geq c^*$, where $\mathbf{r}^{PH}\{g(u,Z)\} =$ $g(u,Z) - E[g(T,Z) \mid Z, T>u]$.

$\Pi[D \mid \Lambda_0^{(F),\perp,MN}] = U_f^{MN}(h)$, with $h(Z) = E[D\varsigma \mid Z]\{E[\varsigma^2 \mid Z]\}^{-1}$.

We next demonstrate that the influence function U of $n^{-1/2} S\{\hat{D}(\beta_0)\}$ and the influence function $-\kappa^{-1}U$ of $\hat{\beta}$ given in Theorem (3.4) satisfy, as they must, (d) and (e) of Theorem (4.1). Using the projection arguments for the Cox model in Ritov and Wellner (1988), for any $D=d(V)$, $\mathbf{u}(D) - \Pi[\mathbf{u}(D) \mid \Lambda^{(3)}] = D - \Gamma(D/K) + \rho[\Gamma(D/K), \Gamma(W)]$ where D/K is the function $D/K(u)$. For any $A^{(2)} \equiv \int_0^\infty dM_Q(u) H^{(2)}(u)$ with $H^{(2)}(u)$ being $\mathbf{F}(u)$-predictable, $A^{(2)} - \Pi[A^{(2)} \mid \Lambda^{(3)}] = \Gamma(H^{(2)}) - \rho[\Gamma(H^{(2)}), \Gamma(W)]$. Thus, from their definitions, $U \equiv U_f - U_{mis}$ $+ U_{rec} = \mathbf{u}(U_f) + A_{mis}^{(2)} - \Pi[\mathbf{u}(U_f) + A_{mis}^{(2)} \mid \Lambda^{(3)}]$ where $A_{mis}^{(2)} = 0$ except that $A_{mis}^{(2),AF2} =$ $-\int_0^\infty dM_Q(u) E[D^{AF2}(\beta_0) \mid Z]/K(u) \in \Lambda^{(2)}$ and $U_f \in \Lambda_0^{(F),\perp}$ as required by part d of Theorem 4.1. Similar arguments show $\kappa^{-1}U$ satisfies part (e) of Theorem (4.1). Since, under regularity conditions, $\hat{\beta}$ will be RAL, Lemma (4.2) and Theorem (4.1) imply that the value of κ in Theorem (3.4) must be $E[U_f S_{eff}^{(F)\prime}]$.

4c. <u>Proof of Lemmas 3.1 and 3.3 and Theorem 3.5</u>

Consider any two correctly specified models (a) and (b) for the censoring process satisfying (3.2) with the first model nested within the second, so that $\Lambda^{3,(a)} \subseteq \Lambda^{3,(b)} \subseteq \Lambda^{(2)}$ in obvious notation. Then we have

<u>Theorem (4.7)</u>: For a given $\hat{D}(\beta)$, $\kappa^{-1}\mathrm{Var}(U^{(a)})\kappa^{\prime -1} \geq \kappa^{-1}\mathrm{Var}(U^{(b)})\kappa^{-1\prime} \geq \kappa^{-1}\mathrm{Var}\{\mathbf{u}(U_f) -$ $\Pi[\mathbf{u}(U_f) \mid \Lambda^{(2)}]\}\kappa^{-1\prime} = \mathrm{Var}^A\{n^{1/2}(\hat{\beta}^* - \beta_0)\}$. The last inequality is strict unless $\Pi[\mathbf{u}(U_f) +$

$A^{(2)}_{mis} \mid \Lambda^{(3),b}] = \Pi[\mathbf{u}(U_f) + A^{(2)}_{mis} \mid \Lambda^{(2)}]$ with probability one which will occur if $W^{mis}_{(u)}$ is included in (3.2) for model b.

<u>Proof</u>: For any A and closed subspace Ω, it is straightforward to show $\mathrm{Var}[A-\Pi(A \mid \Omega)] = \mathrm{Var}(A) - \mathrm{Var}\{\Pi(A \mid \Omega)\}$. Further $\mathrm{Var}\{\Pi[A \mid \Omega^{(1)}]\} \leq \mathrm{Var}\{\Pi[A \mid \Omega^{(2)}]\}$, in the positive definite sense, if $\Omega^{(1)} \subseteq \Omega^{(2)}$ with strict inequality unless the projections are equal with probability one. The theorem follows by setting $A = \mathbf{u}(U_f) + A^{(2)}_{mis}$, noting $\hat{D}(\beta)$ determines A, and calculating the projections using the formulae given in Theorem (4.4) and in the last subsection.

Except for the asymptotic equivalence of $\hat{\beta}^*$ and $\hat{\hat{\beta}}^*$ discussed in Appendix 3, Theorem (3.5) now follows from Theorems (4.1f), (4.5), and (4.7) with h_{eff} defined by $D_{eff} = U_f(h_{eff})$. Part (a) of Lemma (3.1) follows from Theorem (4.7). Part (b) follows by noting, by Theorem 3.4, $\mathrm{Var}(U_{rec}) \neq 0 \Leftrightarrow E[\Gamma(H_{mis})\Gamma(W)']\neq 0$ and then evaluating this expectation under (2.4c). Part (a) of Lemma (3.3) follows from Theorem (4.7) and part (e) of Theorem (4.1), upon noting that $\Lambda^{(F),\perp,MN}_{0^*}$ has a single member when Z is dichotomous (since all functions h(Z) lead to the same estimate $\tilde{\beta}^{MN}$). Part (b) is a corollary of Theorem (4.7). Finally part (c) follows from the fact that the usual KM-estimator is known to be the non-parametric MLE based on the c^*-$\overline{sur}$-data.

4d. <u>Semiparametric Efficient Estimation</u>

For the models (2.2a)-(2.2c), the solution $D_{eff} \equiv U_f(h_{eff})$ to Eq. (4.5) does not generally exist in closed form except for the mean model (2.2c).

Specifically, since, by Theorem (4.5f), $\Lambda^{(F),\perp,MN}_0 = \{h(Z)\varsigma\}$, we can write $D_{eff} \equiv h_{eff}(Z)\varsigma$. Theorems (4.5g) and (4.6) and Eq. (4.5) imply $S^{(F)}_{eff} = [\partial g(Z,\beta_0)/\partial\beta]\{E[\varsigma^2 \mid Z]\}^{-1}\varsigma = h_{eff}(Z)\varsigma + E[\mathbf{v}(D_{eff})\varsigma \mid Z]\{E[\varsigma^2 \mid Z]\}^{-1}\varsigma$. But, by Theorem (4.4), $E[\mathbf{v}(D_{eff})\varsigma \mid Z] = h_{eff}(Z)P^{MN}(Z)$ where $P^{MN}(Z) = E\{\int_0^{X^*} du\lambda_Q[u \mid \bar{L}(u),Z]\{K(u)\}^{-1}\mathrm{Var}[\varsigma \mid \bar{L}(u),Z,Y(u)=1] \mid Z\}$. Hence $h^{MN}_{eff}(Z) = [\partial g(Z,\beta_0)/\partial\beta]\{E[\varsigma^2 \mid Z] + P^{MN}(Z)\}^{-1}$. Thus, as argued in Appendix 3, $\hat{\hat{\beta}}^*(\hat{h}_{eff})$ is a semiparametric efficient adaptive estimator where $\hat{h}_{eff}(Z)$ is estimated by applying (nested) non-parametric regression estimates to each conditional expectation in the RHS of the identities $E[\varsigma^2 \mid Z] = E[\tau\varsigma^2/K(X) \mid Z]$ and $P^{MN}(Z) =$

$E[\int_0^\infty dN_Q(u)\{E[\tau\varsigma^2/K(X) \mid \bar{L}(u),Z,Y(u)=1] - [K(u)]E^2[\tau\varsigma/K(X) \mid \bar{L}(u),Z,Y(u)=1]\} \mid Z]$ with K(X) and ς replaced by preliminary estimates with $E^2(A) = [E(A)]^2$.

When, as in the PH or AF model, Eq. (4.5) does not admit a closed form solution but $\Lambda^{(F),\perp}_0$ can be explicitly exhibited as $\{U_f(h)\}$, one can obtain semiparametric efficient estimators via "linear combinations of estimating functions" as in Newey (1992). Specifically, let $g_1(u,Z), g_2(u,Z),\cdots$ be an infinite complete basis sequence of real-valued functions in the sense that if $\iint_{u\in[0,c_{max}]} h^2(u,z)dudF(z)<\infty$, then there exists constants c_j such that $\lim_{k\to\infty} \iint_{u\in[0,c_{max}]}[h(u,z) - \sum_{j=1}^{k} c_j g_j(u,z)]^2 dudF(z)$ is zero. Sequences of polynomials (i.e., power series) in (u,Z) are known to be complete. Let $g^{[k]}(u,Z)$ be the K-vector $(g_1,\cdots,g_k)'$, $k>p$, $\beta_0 \epsilon R^p$. Following Newey (1992b), $\hat{h}^{[k]}(u,Z) = \hat{b}^{[k]}g^{[k]}(u,Z)$ is the estimated optimal linear combination of the $g^{[k]}$ in the sense that $\mathrm{Var}^A\{n^{1/2}[\hat{\hat{\beta}}^*(\hat{h}^{[k]})-\beta_0]\} \le \mathrm{Var}^A\{n^{1/2}[\hat{\hat{\beta}}^*(bg^{[k]})-\beta_0]\}$ for any p×k matrix b. Here $\hat{b}^{[k]}$ is the p×k matrix $\hat{\kappa}(\hat{\beta},g^{[k]})'\{\hat{\mathrm{V}}\mathrm{ar}[U(\hat{\beta},g^{[k]})]\}^{-1}$, $\hat{\beta}$ is a preliminary estimate, and e.g., $\hat{\kappa}(\hat{\beta},g^{[k]})$ is $\hat{\kappa}(\hat{\beta},h)$ with $g^{[k]}$ replacing h, and, finally, the covariates W(u) used in computing $\hat{\mathrm{V}}\mathrm{ar}[U(\hat{\beta},g^{[k]})]$ include $\hat{W}^{mis}(u)$ as defined in Section 3k except based on $\hat{H}_{mis}(u,\hat{\beta},g^{[k]})$ rather than $\hat{H}_{mis}(u,\beta) \equiv \hat{H}_{mis}(u,\beta,h)$. It follows from the completeness of $g_1,g_2,\cdots$ and the optimality (for fixed k) of $\hat{\hat{\beta}}^*(\hat{h}^{[k]})$ that the asymptotic variance of $\hat{\hat{\beta}}^*(\hat{h}^{[k]})$ can be made arbitrarily close to that of $\hat{\beta}^*(h_{eff})$ (and thus to the semiparametric variance bound) for k sufficiently large, provided the linear operator mapping h to U(h) is mean-square continuous. In fact, Newey (1992b) shows that, under regularity conditions, if we let the dimension $k \equiv k(n)$ increase with sample size n at an appropriate rate $\hat{\hat{\beta}}^*(\hat{h}^{[k(n)]})$ is a feasible semiparametric efficient estimator. The results obtained in this subsection will be largely of theoretical rather than practical interest due to the "curse of dimensionality."

4e. Proof of Theorem 3.2 and an Efficient Koul et al. Estimator

We first prove part e of Theorem 3.2. Ritov and Wellner (1988) proved that, for the AF model, the efficient score $S_{eff,\bar{s}}$ based on the c^*-$\overline{sur}$-data, under the independent censoring mechanism (2.4b), is

$$\int_{-\infty}^{\infty}\{dN_\epsilon^*(u) - \lambda_\epsilon(u)\, I(\omega>u)du\}b^{AF}(u,Z) \tag{4.8}$$

with $N_\epsilon^*(u) = I\{\omega \le u, \Delta\tau = 1\}$ and $b^{AF}(u,Z) = \{d\ell n\lambda_\epsilon(u)/du\}\{Z-E[Z \mid \omega>u]\}$, and $\omega = \ell nX+\beta_0 Z$ where $\bar{s}$ in $S_{eff,\bar{s}}$ specifies the c^*-$\overline{sur}$-data. Suppose we could find a function $\rho(\Delta,\nu,Z)$ of the c^*-full data (3.4) (i.e., of V) such that $S_{eff,\bar{s}} = \mathbf{g}[\rho(\Delta,\nu,Z)]$ with $\mathbf{g}(\cdot)$ as in Theorem 4.4. Then since, by Theorem (4.1g), $S_{eff,\bar{s}} = \mathbf{g}\{\mathbf{m}^{-1}(D_{eff,\bar{s}})\}$, we would obtain $D_{eff,\bar{s}} = \mathbf{m}\{\rho(\Delta,\nu,Z)\}$ without solving (4.5). Now, since $D_{eff,\bar{s}} \in \Lambda_0^{(F),\perp}$ and, for dichotomous Z, $\Lambda_0^{(F),\perp} = \{U_f^{AF2}(r)\}$, we can write $D_{eff,\bar{s}} = U_f^{AF2}(r_{\bar{s}}) \equiv D^{AF2}(\beta_0,r_{\bar{s}}) = R_{\bar{s}}(\beta_0)\{Z-E(Z)\}$ with $E[R_{\bar{s}}(\beta_0)] = 0$. Hence $\hat{\beta}_{\bar{s}}$ solving $0 = S\{\hat{D}^{AF2}(\beta,r_{\bar{s}})\} = \Sigma_i \tau_i R_{\bar{s},i}(\beta)\{Z_i-\hat{E}(Z)\}/\hat{K}_i^0$ is semiparametric efficient in the absence of surrogate marker data. We call $\hat{\beta}_{\bar{s}}$ the efficient Koul et al., estimator due to its similarity to the estimator in Koul et al. (1981). (Even if Q is independent of Z, we must divide by the Z-specific KM estimator $\hat{K}_i^0$ rather than the marginal KM estimator for censoring to "fully project on $\Lambda^{(2)}$" and thus to be efficient.) If Z is not dichotomous, $D_{eff,\bar{s}}$ will not be a multiplicatively separable function of (Δ,ν) and Z.

A calculation shows Eq. (4.8) equals $\mathbf{g}\{\rho(\Delta,\nu,Z)\}$ for $\rho(\Delta,\nu,Z) = \Delta\, \boldsymbol{\ell}\, b^{AF}(\epsilon,Z) + (1-\Delta)b^{*,AF}(\mu,Z)$ where $\boldsymbol{\ell}\, b^{AF}(\epsilon,Z) = b^{AF}(\epsilon,Z) + b^{*,AF}(\epsilon,Z)$, and

$b^{*,AF}(u,Z) = -\int_{-\infty}^{u} \lambda_\epsilon(x) b^{AF}(x,Z)dx = E[\ell\, b^{AF}(\epsilon,Z) \mid Z, \epsilon>u]$ (Ritov and Wellner, 1988). Hence, by Theorem (4.4), $D_{eff,\bar{s}} = \mathbf{m}\{\rho(\Delta,\nu,Z)\} = K(X^*)\rho(\Delta,\nu,Z) + \int_0^{X^*} \lambda_Q[u \mid Z]K(u)b^{*,AF}\{\ell nu+\beta_0 Z,Z\}$. Specializing to dichotomous Z and $\beta_0 = 0$, $\rho(\Delta,\nu,Z) = [\Delta r^{AF}_{pf,1}(\ell nT) + (1-\Delta)\, r^{AF}_{pf,2}(\ell n(c^*))][Z-E(Z)]$ and $b^{*,AF}(u,Z) = r^{AF}_{pf,2}(u)[Z-E(Z)]$. It follows that $r^{AF}_{\bar{s}}(u)$ is as given in part e of Theorem (3.2) if $\lambda_Q[u \mid Z] = \lambda_Q(u)$. To proof part e for the PH model, we note $S^{PH}_{eff,\bar{s}}$ at β_0=0 is given by (4.8) with $\epsilon=\ell nT$ and with $b^{PH}(u,Z) \equiv Z-E[Z \mid \ell nX>u]$ replacing b^{AF} (Ritov and Wellner, 1988).

Theorem (3.2a) follows by noting $NC_s^{c^*} = NC_{\bar{s}}^{c^*}$ if and only if $D_{eff,\bar{s}}$ still solves (4.5) given data on L(u). That is, by Theorem (4.4), if and only if $\int_0^{X^*} du\{E[D_{eff,\bar{s}} \mid \bar{L}(u),Z,Y(u)=1] - E[D_{eff,\bar{s}} \mid Z,Y(u)=1]\} = 0$ w.p.1. Theorem (3.2c) follows from Theorem 4.4 and the relations $NC_s^{c^*} = NC_{full}^{c^*} \Leftrightarrow S_{eff} = S^{(F)}_{eff} \Leftrightarrow S^{(F)}_{eff} = D_{eff} \Leftrightarrow \mathbf{v}(D_{eff}) = 0$ and $\mathbf{u}(D_{eff}) - \Pi[\mathbf{u}(D_{eff}) \mid \Lambda^{(2)}] = D_{eff}$ w.p.1. Theorem (3.2d) then follows from $S^{(F)}_{eff} = D_{eff}$ and the expressions for $S^{(F),PH}_{eff}$ and $S^{(F),AF}_{eff}$ at β_0=0, e.g., as given in Ritov and Wellner (1988).

To prove part (f), note that, by Lemma (4.1) and Theorem (4.1d), the influence function of any RAL modified Buckley-James test of the hypothesis β_0=0 in model (2.2b), under (2.2c), must be of the form $\mathbf{u}(D) - \Pi[\mathbf{u}(D) \mid \Lambda^{(2)}]$ for some $D \in \Lambda_0^{(F),\perp}$ [since the fact that the test does not use the information that $\lambda_Q[u \mid \bar{L}(u),Z] = \lambda_Q(u \mid Z)$ implies that it must be RAL even if only (2.4d) rather than (2.4c) were imposed, and, thus, $\Lambda^{(3)}$ were $\Lambda^{(2)}$]. But, by Theorems (4.5f) and (4.7), this is exactly the influence function of the test $\psi(r, w^r_{op}, c^*)$ with $r[\Delta,\ell n(X^*)] = D/\{Z-E(Z)\}$.

Appendix 1

Theorem (A.1): A RAL test statistic $n^{-1/2}S(\beta_0)$ at β_0 with influence function D is asymptotically normal with mean $c_1 E[DS_{eff}']$ and variance $E[DD']$ under a LDGP $Lik(\beta_n,\eta_n;V)$ in any regular parametric submodel such that $n^{1/2}(\beta_n-\beta_0)\rightarrow c_1, n^{1/2}(\eta_n-\eta_0)\rightarrow c_2$. Further, $D\in\Lambda_0^{(F),\perp}$. The NC parameter, $E[D'S_{eff}]\{E(DD')\}^{-1}E[DS_{eff}']$, is less than or equal to $E[S_{eff}S_{eff}']$ with equality if $D = S_{eff}$.

Proof: $n^{-1/2}S(\beta_0) = \Sigma_i D_i + o_p(1)$ under any LDGP by contiguity of a LDGP to the fixed process (β_0,η_0). Hence, $n^{-1/2}S(\beta_0) = n^{-1/2}\Sigma_i[D_i - E_n(D_i)] + n^{1/2}E_n(D_i)$ where E_n is an expectation with respect to the density $Lik(\beta_n,\eta_n,V)$. But, $n^{-1/2}\Sigma_i[D_i-E_n(D_i)] \xrightarrow{d} N\{0,E(DD')\}$ under the LDGP as discussed in the proof of Theorem 2.2 of Newey (1990a). Also, as in Newey (1990a, Theorem 2.2), by regularity of the submodel, $n^{1/2}E_n[D] = E[DS_\beta']c_1 + E[DS_\eta']c_2 + o(1)$. Consider now a LDGP with $\beta_n = \beta_0$, i.e., $c_1 = 0$. Then, by the assumption of regularity of $n^{-1/2}S(\beta_0)$, $E[DS_\eta'] = 0$ for all regular submodels so $D\in\Lambda_0^{(F),\perp}$. Hence $E[DS_\beta'] = E[DS_{eff}']$ and the theorem follows by Slutsky's theorem and the fact that the maximum possible value of the NC parameter is $E[S_{eff}S_{eff}']$ by the Cauchy-Schwartz inequality.

Appendix 2

Proofs of Lemma (4.1a) and Theorems (4.1) and (4.2)

Proof of Lemma (4.1a): $\Lambda^{(3)} \subseteq \Lambda^{(2)}$ since (a) any score S_ψ of (4.4) is in $\Lambda^{(2)}$ by the conditional mean zero property of scores, and (b) $\Lambda^{(2)}$ is closed since it is the inverse image of the closed set of {0} under the continuous mapping $E\{\cdot \mid V\}$. To prove (4.1b), note for any bounded $A^{(2)}$ in $\Lambda^{(2)}$, the submodel $\pi(R,V_{(R)})(1+\psi' A^{(2)})$ defined on a sufficiently small open ball around $\psi_0=0$ is regular with score $A^{(2)}$ by Lemma C.4 of Newey (1990b). But any function in $\Lambda^{(2)}$ can be approximated in mean square by bounded functions.

Proof of Theorem 4.1: Part (a) follows from Lemma A5.5 of BKRW. Proof of (b): Since $\Lambda^{(3)} \subset \Lambda^{(2)}$, in order to prove $\Lambda^{(3)} \perp \Lambda^{(1)}$, it suffices to show that for any $H=h(V)$,

$$E[\mathbf{g}(H)A^{(2)\prime}]=0 \tag{A.1}$$

But $E[\mathbf{g}(H)A^{(2)\prime}] = E\{E[\mathbf{g}(H)A^{(2)\prime} \mid V]\} = E[\Sigma_r \pi(r,V_{(r)})E[H \mid V_{(r)}]a^{(2)}(r,V_{(r)})'] = E\{HE[A^{(2)} \mid V]'\}=0$, where the 2nd identity is by (4.2), and the third by iterated expectations. Now $\Lambda^{(3)}$ is closed by definition. To prove $\Lambda^{(1)}$ is closed, we note that $\Lambda^{(1)}$ is the image of the closed set $\Lambda^{(F)}$ under the linear operator $\mathbf{g}(\cdot)$. Hence it suffices to show $\mathbf{g}(\cdot)$ has a continuous inverse. But $\mathbf{g}(\cdot)$ has a continuous inverse by part a) of proposition A1.5 of BKRW (1991) since $E[\| \mathbf{g}(\Lambda^{(F)}) \|^2] \geq E[I(R=\mathbf{1})A^{(F)\prime}A^{(F)}] = E[\pi(\mathbf{1})A^{(F)\prime}A^{(F)}] \geq \sigma E[\| A^{(F)} \|^2]$ where the equality is by (4.2) and the final inequality by (4.3). To prove part (c), it follows from Lemma A5.5 of BKRW and Lemma C.4 of Newey (1990b) that Λ is the closure of $\{\mathbf{g}(A^{(F)}) + A^{(3)}; A^{(3)} \in \Lambda^{(3)}, A^{(F)} \in \Lambda^{(F)}\}$ which is $\Lambda^{(1)} \oplus \Lambda^{(3)}$ by part (b).

To prove (d) and (e), note that if $B \in \Lambda_0^{\perp}$, then for some D^{**}, $B=D^{**}-\Pi[D^{**} \mid \Lambda]$ with $E[D^{**}]=0$. Hence, by $\Lambda^{(1)} \perp \Lambda^{(3)}$, $B = D^{*}-\Pi[D^{*} \mid \Lambda^{(3)}]$, where $D^{*} \equiv D^{**}-\Pi[D^{**} \mid \Lambda^{(1)}] \in \Lambda_0^{(1),\perp}$. Now, for any $H^{*}=h^{*}(R,V_{(R)})$, (4.2) implies

$$H^{*}-\mathbf{u}(H) \in \Lambda^{(2)} \quad , \tag{A.2a}$$

with $H \equiv E[H^{*} \mid V]$. Thus,

$$E[H^{*}\mathbf{g}(A^{(F)\prime})] = E[\mathbf{u}(H)\mathbf{g}(A^{(F)})'] = E[HA^{(F)\prime}] \tag{A.2b}$$

where the first equality is by (A.2a) and $\Lambda^{(1)} \perp \Lambda^{(2)}$, and the second uses (4.2). Hence,

$$H^{*} \in \Lambda_0^{(1),\perp} \Rightarrow H \in \Lambda_0^{(F),\perp} \quad . \tag{A.3}$$

Now write $D^{*} = \mathbf{u}(D) + \{D^{*}-\mathbf{u}(D)\}$, and note $D \equiv E[D^{*} \mid V] \in \Lambda_0^{(F),\perp}$ and $D^{*}-\mathbf{u}(D) \in \Lambda^{(2)}$ by (A.2a), proving part (d). Further, part (e) follows from $E[BS_\beta'] = E[B\mathbf{g}(S_\beta^{(F)})'] = E[\mathbf{u}(D)\mathbf{g}(S_\beta^{(F)})'] = E[DS_\beta^{(F)\prime}]$ where the second equality uses (A.2b) and $\Lambda^{(1)} \perp \Lambda^{(3)}$ and the third uses (4.2).

Proof of part (f): $S_{eff} = S_\beta-\Pi[S_\beta \mid \Lambda] = S_\beta-\Pi[S_\beta \mid \Lambda^{(1)}] \in \Lambda_0^{(1),\perp}$ since, by (A.1), $S_\beta \equiv \mathbf{g}(S_\beta^{(F)}) \perp \Lambda^{(3)}$. Since $\Pi[S_\beta \mid \Lambda^{(1)}] = \mathbf{g}(A_\beta^{(F)})$ for some $A_\beta^{(F)} \in \Lambda^{(F)}$ by $\Lambda^{(1)}$ closed, $S_{eff} =$

$\mathbf{g}(A_\beta) = \mathbf{u}\{\mathbf{m}(A_\beta)\} + [\mathbf{g}(A_\beta)-\mathbf{u}\{\mathbf{m}(A_\beta)\}]$, where $A_\beta = S_\beta^{(F)}-A_\beta^{(F)}$. Since for any $H^*\equiv h^*(V)$, $E[\mathbf{g}(H^*) \mid V] = \mathbf{m}(H^*)$ and, so by (A.2a), $\mathbf{g}(H^*)-\mathbf{u}[\mathbf{m}(H^*)]\in\Lambda^{(2)}$, it follows from (A.1), that

$$\mathbf{u}[\mathbf{m}(H^*)]-\mathbf{g}(H^*) = \Pi[\mathbf{u}\{\mathbf{m}(H^*)\} \mid \Lambda^{(2)}] \ . \tag{A.4}$$

Letting $D_{eff} = \mathbf{m}(A_\beta)$, we have $S_{eff} = \mathbf{u}(D_{eff}) - \Pi[\mathbf{u}(D_{eff}) \mid \Lambda^{(2)}]$, with $D_{eff}\in\Lambda_0^{(F),\perp}$ by (A.3).

To show D_{eff} is the unique $D\in\Lambda_0^{(F),\perp}$ solving

$$\Pi[\mathbf{m}^{-1}(D) \mid \Lambda_0^{(F),\perp}] = S_\beta^{(F)}-\Pi[S_\beta^{(F)} \mid \Lambda_0^{(F),\perp}] \ , \tag{A.5}$$

we first show $\mathbf{m}^{-1}(D)$ is well-defined, i.e., for any $D=d(V)$, there exists a unique $H=h(V)$ such that $D=\mathbf{m}(H)$. To do so, we note $D=\mathbf{m}(H)$ is equivalent to $\pi(\mathbf{1})^{-1}D=H+\pi(\mathbf{1})^{-1}\Sigma_{r\neq 1}\pi(r)E[H \mid V_{(r)}] \equiv (\mathbf{i}+\mathbf{k})H$, where $\mathbf{i}$ is the identity operator and $\mathbf{k}=\pi(\mathbf{1})^{-1}\Sigma_{r\neq 1}\Pi(r)E[\cdot \mid V_{(r)}]$. Thus, by the Riesz theory for type 2 Fredholm operator equations $D=\mathbf{m}(H)$ will have a unique solution H if the bounded linear operator $\mathbf{k}$ is compact and the null space of $(\mathbf{i}+\mathbf{k})$ is $\{0\}$. [Note $\mathbf{k}$ is bounded since $\pi(\mathbf{1})^{-1}$ is bounded by (4.3)]. But $E(\cdot \mid V_{(r)})$ is Hilbert-Schmidt and, thus compact (BKRW, Section A.4). Since the product of two bounded linear operators is compact if either is compact, $\mathbf{k}$ is compact (Kress, 1990). To show that if $H\neq 0$, $(\mathbf{i}+\mathbf{k})(H)\neq 0$, write $E[\{\pi(\mathbf{1})\}H\}'[\mathbf{i}+\mathbf{k}](H)] = E\{H'[\Sigma_r\pi(r)E(H \mid V_{(r)})]\} \geq E[\pi(\mathbf{1}) \| H \|^2] \geq \sigma E[\| H \|^2]$ where the last inequality uses (4.3). Now, D_{eff} satisfies (A.5) by $\mathbf{m}^{-1}(D_{eff}) = S_\beta^{(F)}-A_\beta^{(F)}, A_\beta^{(F)}\in\Lambda^{(F)}$.

To show uniqueness of D_{eff}, suppose (A.5) holds for $\tilde{D}\in\Lambda_0^{(F),\perp}, \tilde{D} \neq D_{eff}$. Then $\mathbf{m}^{-1}(\tilde{D}) = S_\beta^{(F)} + \tilde{A}^{(F)}$ for $\tilde{A}^{(F)} \neq -A_\beta^{(F)}$, $\tilde{A}^{(F)}\in\Lambda^{(F)}$. Hence, $-\mathbf{g}(\tilde{A}^{(F)}) \neq \Pi[\mathbf{g}(S_\beta^{(F)}) \mid \Lambda^{(1)}]$, so, by definition, $A^{(F)}$ exists such that $0 \neq E[\mathbf{g}(S_\beta^{(F)}+\tilde{A}^{(F)})\mathbf{g}(A^{(F)})']$, i.e., $0 \neq E[\mathbf{u}\{\mathbf{m}(S_\beta^{(F)}+\tilde{A}^{(F)})\}\mathbf{g}(A^{(F)})'] = E[\mathbf{u}(\tilde{D})\mathbf{g}(A^{(F)})'] = E[\tilde{D}A^{(F)}{}']$ so $\tilde{D}$ is not in $\Lambda_0^{(F),\perp}$ where we have used (A.2b) and (4.2). Thus we have proved uniqueness by contradiction.

Proof of (g): g is just (A.4) above.

Proof of Theorem 4.2: If missingness is monotone and (4.2) holds, (a)-(d) of Theorem 4.2 follow directly from their definitions. Part (f) follows from the fact that under monotone missingness $f[R \mid V] = \prod_{m=1}^{M} \{f(R_m \mid R_{m-1}=1,\overline{V}_m)\}^{R_{m-1}}$ with parametric submodels $\prod_{m=1}^{M} \{f(R_m \mid R_{m-1}=1,\overline{V}_m;\psi)\}^{R_{m-1}}$. It is straightforward to show (f) implies (g). Part (e) can be shown to hold by induction. Part (h) follows from part (c) of Theorem (4.2) and part (f) of Theorem 4.1.

Appendix 3:

In this Appendix, we further discuss the asymptotic distribution of some of our proposed tests and estimators. Under sufficient regularity and smoothness conditions, the estimators $\hat{\hat{\beta}}^*(h)$ of Section 3k and $\hat{\hat{\beta}}^*(h_{eff}^{MN})$ of Section 4c, the test statistic $\psi(\hat{r}_{\bar{s}},w,c^*)$ of Lemma 3.2, and the adaptive test $\psi(r,\hat{w}_{op}^r,c^*)$ discussed following Lemma 3.1 will be RAL.

If RAL, the equality of the limiting distributions of (a) $\hat{\hat{\beta}}^*(h)$ with that of $\hat{\beta}^*(h)$ and (b) $\hat{\hat{\beta}}^*(\hat{h}_{eff})$ with that of $\hat{\beta}^*(h_{eff})$ are a consequence of Proposition (3) in Newey (1992a) since, in Newey's notation, $E[M(z) \mid v]=0$. Newey's Proposition (3) gives conditions under which substituting estimates for an unknown population quantity has no effect on the limiting distribution of a statistic. Similarly, if RAL, Newey's Proposition (3) implies that (a) the limiting distribution of $\psi(\hat{r}_{\bar{s}},w,c^*)$ is as stated in Lemma (3.2), and (b) that $\psi(r,\hat{w}^r_{op},c^*)$ has the same limiting distribution as $\psi(r,w^*,c^*)$ where w^* is the limit of $\hat{w}^r_{op}$.

We now sketch the derivation of the limiting distribution (3.7) of $n^{-1/2}S\{\hat{D}^{AF2}(\beta_0)\}$. By the identity in the first paragraph of Section (3e), $n^{-1/2}S\{\hat{D}(\beta_0)\} = n^{-1/2}\Sigma_i\tilde{D}_i(\beta_0) - n^{-1/2}\Sigma_i\Psi_i(\hat{H}_{1i})$ with $\hat{H}_{1i}(u) = \tilde{D}_i(\beta_0)/\hat{K}_i(u)$, where we have suppressed the AF2 superscript. Proposition (3) of Newey (1992a) implies that $n^{-1/2}\Sigma_i\Psi_i(\hat{H}_{1i}) = n^{-1/2}\Sigma_i\Psi_i(\hat{H}_{2i}) + o_p(1)$ where $\hat{H}_{2i}(u) = D_i(\beta_0)/\hat{K}_i(u)$. Thus, by Lemma 3.5, (3.7) follows from

Theorem A.2: Under (2.2b) $n^{-1/2}S^{(F)}\{\tilde{D}(\beta_0)\} \equiv n^{-1/2}\Sigma_i\tilde{D}_i(\beta_0) = n^{-1/2}\Sigma_iU_f + o_p(1)$ (A.6) with $U_f = D(\beta_0)-E[D(\beta_0) \mid Z]$.

Sketch of Proof: With $\tilde{D}(\beta_0) = \tilde{D}(\beta_0,h,\theta)$, $n^{-1/2}S^{(F)}\{\tilde{D}(\beta_0)\} = A_1+A_2$, where $A_1 = n^{1/2}\tilde{E}\{\int_{-\infty}^{\infty}dM_\epsilon(u)[h(u,Z)-\tilde{\mathscr{L}}(u,\beta_0,h)]\}$, $A_2 = n^{1/2}\tilde{E}\{\int_{-\infty}^{\infty}duI(\nu>u)[h^*(u,Z)-\tilde{\mathscr{L}}(u,\beta_0,h^*)]\}$, with $h^*(u,Z) = \lambda_\epsilon(u)h(u,Z) - \theta(u,Z)$. Now, by a standard argument using Lenglart's inequality, $A_1=A_{11}+o_p(1)$ with $A_{11} = n^{1/2}\tilde{E}[\int_{-\infty}^{\infty}dM_\epsilon(u)[h(u,Z)-\mathscr{L}(u,\beta_0,h)]]$. Further $A_2=A_{21}+A_{22}+A_{23}$, where, with $\mathscr{L}(u) \equiv \mathscr{L}(u,\beta_0,h^*)$, $\tilde{\mathscr{L}}(u) \equiv \tilde{\mathscr{L}}(u,\beta_0,h^*)$, $\bar{F}_\nu(u) = \bar{F}_\mu(u)\bar{F}_\epsilon(u)$ with $\bar{F}_\mu(u) = pr(\mu>u)$, $A_{21} = n^{1/2}\tilde{E}[\int_{-\infty}^{\infty}duI(\nu>u)\{h^*(u,Z)-\mathscr{L}(u)\}]$, $A_{22} = n^{1/2}\int_{-\infty}^{\infty}du\bar{F}_\nu(u)[\mathscr{L}(u)-\tilde{\mathscr{L}}(u)]$, $A_{23} = \int_{-\infty}^{\infty}du\, n^{1/2}\tilde{E}[I\{\nu>u\}-\bar{F}_\nu(u)][\mathscr{L}(u)-\tilde{\mathscr{L}}(u)]$. Now, under regularity conditions, A_{23} is $o_p(1)$ since $\int_{-\infty}^{\infty}n^{1/2}\tilde{E}[I\{\nu>u\}-\bar{F}_\nu(u)]$ is $O_p(1)$ by the CLT, and $\mathscr{L}(u)-\tilde{\mathscr{L}}(u)$ converges to zero uniformly in u. Thus $n^{-1/2}S^{(F)}\{\tilde{D}(\beta_0)\} = n^{-1/2}\Sigma_iD_i(\beta_0)+A_{22}+o_p(1)$ since $A_{11}+A_{21} = n^{-1/2}\Sigma_iD_i(\beta_0)$. Thus, it remains to show $A_{22} = -n^{-1/2}\Sigma_iE[D_i(\beta_0) \mid Z_i] + o_p(1)$.

Since $\tilde{\mathscr{L}}(u) = \tilde{\mathscr{L}}_1(u)/\tilde{\mathscr{L}}_2(u)$ with $\tilde{\mathscr{L}}_1(u) = \tilde{E}[I\{\mu>u\}h^*(u,Z)]$, $\tilde{\mathscr{L}}_2(u) = \tilde{E}[I\{\mu>u\}]$, a Taylor expansion around $\mathscr{L}_1(u)$ and $\mathscr{L}_2(u)$ gives $A_{22} = A_{22}^* + o_p(1)$, where

$$A_{22}^* = n^{1/2}\int_{-\infty}^{\infty}du\bar{F}_\epsilon(u)\bar{F}_\mu(u)\left\{\frac{\tilde{\mathscr{L}}_1(u)-\mathscr{L}_1(u)}{\bar{F}_\mu(u)} + \frac{\mathscr{L}_1(u)[\tilde{\mathscr{L}}_2(u)-\bar{F}_\mu(u)]}{[\bar{F}_\mu(u)]^2}\right\} =$$

$$-n^{1/2}\int_{-\infty}^{\infty}du\bar{F}_\epsilon(u)[\tilde{\mathscr{L}}_1(u) - \tilde{\mathscr{L}}_2(u)\mathscr{L}_1(u)/\bar{F}_\mu(u)] = -n^{-1/2}\Sigma_i\int_{-\infty}^{\infty}du\bar{F}_\epsilon(u)I\{\mu>u\}$$

$$\left\{h^*(u,Z) - \frac{E[h^*(u,Z)I(\mu>u]}{E[I(\mu>u)]}\right\} = -n^{-1/2}\Sigma_iE[D_i(\beta_0) \mid Z_i].$$ A more abstract approach to the proof of Theorem A.2 is to use Theorem 4.3 in Newey (1990a).

Appendix 4

In this Appendix, we assume that (i) time to potential end of follow-up Q_1 and T are dependent due to rapid secular changes in the prognosis of patients at entry into the trial over the calendar year of enrollment; but (ii), in contrast with Section 2, we no longer assume that these changes can be fully explained by secular changes in the covariates recorded in $\bar{L}(u)$. Hence (2.4a) will be false when, as in Section 2, $Q \equiv \min(Q_1,Q_2)$.

Redefine Q to be Q_2, time to initiation of prophylaxis, and, for notational convenience, write time to potential end of follow-up Q_1 as C, and include C in $\bar{L}(u)$ for each $u, u \geq 0$. (Note C is observed even for subjects failing prior to C.) Replace c^* by C wherever c^* occurs in the paper, so, for example, henceforth, $X = \min(T,Q,C)$; $X^* = \min(T,C)$; $\Delta = I(T<C), \tau = I(X \neq Q)$, and the C-observed and C-full data are given by (2.1) and (3.4) with C replacing c^*. With Q,X, and $\bar{L}(u)$ so redefined, (2.4a) will again hold under the assumptions concerning initiation of prophylaxis given in the introduction (provided data on PCP history was recorded in $\bar{L}(u)$). If Z were the treatment indicator, by Z assigned completely at random, we would have

$$C \amalg Z \quad . \tag{A.7}$$

Consider the conditional AF models

$$\text{pr}[T>t \mid Z,C] = \bar{F}_0(e^{\beta_0' Z} t \mid C) \quad , \tag{A.8}$$

and the more general model

$$\text{pr}[T>t \mid Z,C] = \bar{F}_0[e^{g(Z,C,\beta_0)} t \mid C] \tag{A.9}$$

where $g(Z,C,\beta) = 0$ if $Z = 0$ or $\beta = 0$, $g(\cdot,\cdot,\cdot)$ is a fixed real-valued function, and $\bar{F}_0(t \mid c)$ is, for each c, an unspecified survival function. If $g(Z,C,\beta_0)$ depends on C, we say there is a C-treatment interaction. We will show how to generalize the AF2 method of estimation to consistently estimate the parameters of (A.9) from the C-full data under (A.7) and from the C-observed data under (2.4a) and (A.7) [provided K(X) is bounded away from zero]. Only the (generalized) AF2-method allows one to obtain asymptotically normal and unbiased estimators of β_0 from the C-observed data without having to use non-parametric methods to estimate the conditional law of either $\epsilon = \ell nT+g(Z,C,\beta_0)$ or Z given C. Furthermore, since (A.8) and (A.7) together imply the marginal AF model (2.2b), it follows that the parameters of (2.2b) can still be estimated when (A.8), (A.7), and (2.4a) [as redefined] hold.

Let, h(u,Z,C), and θ(u,Z,C) be fixed functions taking values in R^p. Redefine $\tilde{D}^{AF2}(\beta) \equiv \hat{D}^{AF2}(\beta) \equiv \tilde{D}^{AF2}(\beta,h,\theta) \equiv [\int_{-\infty}^{\infty} dN_{\epsilon(\beta)}(u)\{h(u,Z,C) - \tilde{\mathscr{G}}^{AF2}(u,\beta,h,C)\}] - \int_{-\infty}^{\infty} du I\{\nu(\beta) \geq u\}\{\theta(u,Z,C) - \tilde{\mathscr{G}}^{AF2}(u,\beta,\theta,C)\}$, where $N_{\epsilon(\beta)}(u) = I[\nu(\beta) \leq u, \Delta=1]$, $\nu(\beta) = \min\{\epsilon(\beta),\mu(\beta)\}$, $\mu(\beta) = \ell nC + g(Z,C,\beta)$, $\epsilon(\beta) = \ell nT+g(Z,C,\beta)$, $\tilde{\mathscr{G}}^{AF2}(u,\beta,h,c) = \tilde{E}[I\{\ell nc+g(Z,c,\beta)>u\}h(u,Z)]/\tilde{E}[I\{\ell nc+g(Z,c,\beta)>u\}]$ and $\tilde{E}[I\{\ell nc+g(Z,c,\beta)>u\}] =$

$n^{-1}\sum_{i=1}^{n} I\{\ell nc+g(Z_i,c,\beta)>u\}$. $\tilde{D}^{AF2}(\beta,h,\theta,C)$ has a particularly simple form when Z is a dichotomous (0,1) variable. Specifically $\tilde{D}^{AF2}(\beta,h,\theta) \equiv \tilde{D}^{AF2}(\beta,r) = R(\beta)[Z-\tilde{E}(Z)]$, where $R(\beta) \equiv r\{\Delta(\beta),X^*(\beta),C\} \equiv \Delta(\beta)r_1[\epsilon(\beta),C] + \{1-\Delta(\beta)\}\{r_2[C(\beta),C]\}$, $C(\beta) \equiv \min(\ell nC, \ell nC+g(1,C,\beta))$, $X^*(\beta) \equiv \min[\epsilon(\beta),C(\beta)]$, $\Delta(\beta) = I[\epsilon(\beta)<C(\beta)]$, $r_2(u,C) = -\int_{-\infty}^{u}\theta_1(x,C)dx$, $r_1(u,C) = h_1(u,C) + r_2(u,C)$, and, e.g., we write $\theta(u,Z,C)$ as $Z\theta_1(u,C) + \theta_0(u,C)$.

Noting (A.7) and (A.9) imply that

$$(\epsilon,C)\coprod Z \tag{A.10}$$

with $\epsilon \equiv \epsilon(\beta_0)$, we see that $\tilde{\mathscr{L}}^{AF2}(u,\beta_0,h,c)$ converges in probability under (A.7) and (A.9), to $\mathscr{L}^{AF2}(u,\beta_0,h,c) \equiv E[h(u,Z,C) \mid \mu>u,C=c] = E[h(u,Z,C) \mid \nu>u,C=c]$ with $\mu=\mu(\beta_0)$ and $\nu=\nu(\beta_0)$.

With the above redefinitions, Theorem (3.3) [under model (A.9) and (A.7)] and Theorem (3.4) [under model (A.9), (A.7), and (2.4a)] remain true for $\tilde{\beta}^{AF2}$ and $\hat{\beta}^{AF2}$, with $U_f^{AF2} \equiv U_f^{AF2}(h,\theta) = D^{AF2}(\beta_0) - E[D^{AF2}(\beta_0) \mid Z]$, $\kappa^{AF2} \equiv \kappa^{AF2}(h,\theta) = E[U_f^{AF2}(h,\theta)U_f^{AF2}(h_{op}^{AF},\theta_{op}^{AF})']$ where $h_{op}^{AF}(u,Z,C) = Z\partial\ell n\lambda_\epsilon(u \mid C)/\partial u$ and $\theta_{op}^{AF}(u,Z,C) = Z\partial\lambda_\epsilon(u \mid C)/\partial u$. If $Z\epsilon\{0,1\}$, $U_f^{AF2} \equiv U_f^{AF2}(r) = \{R(\beta_0)-E[R(\beta_0)]\}\{Z-E[Z]\}$. The formulae for the variance estimates in Section 3j.1 and 3j.2 need to be modified only in that, now, $P_{ji}(\beta) \equiv P_{ji}(\beta,h,\theta) \equiv [\int_{-\infty}^{\infty} dN_{\epsilon_j(\beta)}(u)\{h(u,Z,C_j) - \tilde{\mathscr{L}}^{AF2}(u,\beta,h,C_j)\}] - \int_{-\infty}^{\infty} du I\{\epsilon_j(\beta)>u\}I\{\ell nC_j+g(Z_i,C_j,\beta)>u\}[\theta(u,Z_i,C_j) - \tilde{\mathscr{L}}^{AF2}(u,\beta,\theta,C_j)]$. If Z is dichotomous, we use $\tilde{V}ar(U_f^{AF2}(\beta)) \equiv \tilde{V}ar\{U_f^{AF2}(\beta,r)\} = \tilde{E}[\{R(\beta) - \tilde{E}[R(\beta)]\}^2]\tilde{E}[\{Z-\tilde{E}(Z)\}^2]$.

Since (C,Z) is ancillary, $S_{eff}^{(F)} = U_f^{AF2}(h_{op},\theta_{op})$ is the efficient score for β_0 based on the C-full data in model (A.9) whether or not (A.7) is true, and, if (A.7) is true, whether or not (A.7) is known to be true *a priori*. Similarly, the efficient score S_{eff} and the solution D_{eff} to Eq. (4.5) based on the C-observed data in the model defined by (A.9), $K(X)>\sigma>0$ w.p.1, and (2.4a) do not depend on whether the restriction (A.7) is true. Thus the restriction (A.7) was useful only in that it allowed us to consistently estimate β_0 of (A.9) without requiring non-parametric estimation of the conditional law of either ϵ or Z given C. $[A_0^{(F),\perp}$ equals $\{U_f^{AF2}(h,\theta) + a(Z,C) - E[a(Z,C) \mid C] - E[a(Z,C) \mid Z]\}$, where a(Z,C) is any mean 0 function of Z and C, when (A.7) is known, and equals $\{U_f^{AF2}(h,\theta)\}$ when (A.7) is true but not known, although D_{eff} is of the form $U_f^{AF2}(h,\theta)$ in either case.]

Acknowledgement: This work was in part supported by grants K04-ES00180, 5-P30-ES00002, and R01-ES03405 from NIEHS. We are indebted to W. Newey for many helpful conversations concerning the subject matter of this paper. Z. Ying also provided helpful advice.

Bibliography

Andersen, P.K. and Gill, R.D. (1982). Cox's regression model for counting processes: A large sample study. Ann. Statist. 10:1100-1120.

Begun, J.M., Hall, W.J., Huang, W.M. and Wellner, J.A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. Ann. Statist. 11:432-452.

Bickel, P., Klassen, C., Ritov, Y., and Wellner, J. (1990). Efficient and adaptive inferences in semiparametric models. (forthcoming)

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. J. Econometrics. 34:305-324.

Gray, R.J., and Tsiatis, A.A. (1989). The linear rank test for use when the main interest is in differences and cure rates. Biometrics 45:899-904.

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

Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York.

Kress, R. (1990). Linear integral equations. Springer-Verlag Berlin.

Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. Ann. Statist. 9:1276-88.

Lagakos, S., Lim, L., and Robins, J.M. (1990). Adjusting for early treatment termination in comparative clinical trials. Statist. Med. 9:1417-1424.

Newey, W.K. (1990a). Semiparametric efficiency bounds. J Appl. Econom. 5:99-135.

Newey, W.K. (1990b). Efficient estimation of Tobit models under conditional symmetry in W. Barnett, J. Powell, and G. Tauchen, (eds), Semiparametric and Non-parametric Methods in Econometrics and Statistics, Cambridge Univ. Press, Cambridge, pp. 291-336.

Newey, W.K. (1992a). The variance of semiparametric estimators. Econometrica (forthcoming).

Newey, W.K. (1992b). Efficient estimation of semiparametric models via moment restrictions, manuscript, MIT.

Pakes, A. and Pollard, D. (1989). Simulation in the asymptotics of optimization estimators. Econometrica 57:1027-1057.

Prentice, R.L. (1978). Linear rank tests with right censored data. Biometrika 65:167-179.

Ritov, Y. and Wellner, J.A. (1988). Censoring, martingales, and the Cox model. Contemp. Math. Statist. Inf. Stoch. Pro. (N.U. Prabhu, ed.) 80:191-220. Amer. Math. Soc.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. Ann. Statist. 18:303-328.

Robins, J.M. (1991). Estimating regression parameters in the presence of dependent censoring. (submitted)

Robins, J.M. and Tsiatis, A.A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Commun. Statist.-Theory Meth. 20:2609-2631.

Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1992). Analysis of semiparametric regression models for repeated outcomes under the presence of missing data. (submitted)

Robins, J.M., and Rotnitzky, A. (1992). Semiparametric efficiency in multivariate regression models with missing data. (submitted)

Robins, J.M., Zhao, L.P., Rotnitzky, A., and Lipsitz, S. (1991). Estimation of regression coefficients when a regressor is not always observed. (submitted).

Rubin, D.B. (1976). Inference and missing data. Biometrika 63:581-92.

Tsiatis, A.A. (1990). Estimating regression parameters using linear rank tests for censored data. Ann. Statist. 18:354-372.

Wei, L.J., Ying, Z., and Lin, D.Y. (1990). Linear regression analysis of censored survival data based on rank tests. Biometrika 77:845-852.

Departments of Epidemiology and Biostatistics, Harvard School of Public Health,
665 Huntington Ave., Boston, MA 02115.

bjr/t2e.cr