1993 Proceedings of the
Biopharmaceutical Section,
American Statistical Association.
pp. 24-33.

INFORMATION RECOVERY AND BIAS ADJUSTMENT IN PROPORTIONAL HAZARDS REGRESSION ANALYSIS OF RANDOMIZED TRIALS USING SURROGATE MARKERS

James M. Robins, Harvard School of Public Health

677 Huntington Ave., Boston, Massachusetts 02115

SUMMARY

A class of test and estimators for the parameters of the Cox proportional hazards model are proposed that use surrogate marker data to recover information lost to independent censoring and to adjust for bias due to dependent censoring in a randomized clinical trial. In the presence of independent censoring, our tests are (i) asymptotically distribution free under the null hypothesis of no treatment effect on survival, (ii) incorporate surrogate marker data, (iii) are guaranteed to be locally more powerful than the standard log-rank test and (iv) are shown to outperform the log-rank test in a small simulation study. We show that the optimal estimator in our class attains the semiparametric variance bound for our model in the sense of Begun et al. (1983). Since the optimal estimator depends on the unknown distribution of the data, we propose an adaptive estimator whose asymptotic variance may attain the variance bound. We also use our approach to obtain adaptive estimates of the treatment arm specific survival curves that are more efficient than the usual Kaplan-Meier estimates and may attain the semiparametric efficiency bound.

I. INTRODUCTION

Randomized clinical trials of the effect of a new treatment on mortality from a chronic fatal disease must be conducted over prolonged periods of time. It is important to be able to stop such trials the moment that it can be determined that the new treatment prolongs survival. The first goal of this paper is to develop, under an independent censoring assumption, 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 on survival.

When the null hypothesis of no effect of treatment arm on survival is rejected, the magnitude of the treatment effect is commonly estimated by estimating the parameter β_0 of the proportional hazards model

$$pr[T>t \mid Z] = \{\bar{F}_0(t)\}^{\exp(\beta_0 Z)} \qquad (1.1)$$

Here T is time to failure, Z is a dichotomous (0,1) treatment arm indicator, and $\bar{F}_0(t)$ is an unspecified baseline survivor function. The null hypothesis $\beta_0=0$ in (1.1) is equivalent to the usual "intention to treat null hypothesis"

$$pr[T>t \mid Z] = \bar{F}_0(t) \qquad (1.2)$$

of equality of treatment arm specific survival curves. We shall provide a class of estimators of β_0 that (a) incorporate surrogate marker data and (b) are guaranteed to be more efficient than the Cox partial likelihood estimator of β_0 and (c) can attain the semiparametric efficiency bound.

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 ACTG AIDS randomized trial 002 comparing the effect of a treatment A, high-dose AZT, with treatment B, low-dose AZT, on survival, over one-half of subjects failed to comply with the assigned treatment protocol and initiated treatment with a new non-randomized therapy, prophylaxis therapy for pneumocystis carinii pneumonia (PCP). To obtain some useful information regarding the benefits of high- vs low-dose AZT, we will regard subjects as censored at the time they initiate prophylaxis therapy. 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 and the Cox partial likelihood estimator of β_0 may be biased because (a) subjects in the low-dose AZT arm were more likely than subjects in the high-dose arm to initiate prophylaxis therapy (61% vs. 50%), and (b), within each arm, censoring is not independent of failure, i.e., death, because subjects at high risk of death (that is, subjects with recurrent episodes of PCP) were twice as likely to be censored (i.e., to initiate prophylaxis). Together (a) and (b) imply that censoring and failure are dependent. If, in this setting, the investigators conducting the trial 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 at t is unrelated to prognosis, our methods provide 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.

2. A FORMALIZATION OF THE PROBLEM

With time measured as time since randomization, defined for subject i, $i=1,\cdots,n$, in Trial 002

Q_i = censoring time = minimum of time to prophylaxis, to loss to follow-up, and to end of follow-up.

T_i = survival time in the absence of censoring

Z_i = the dichotomous treatment arm indicator.

$L_i(t)$ 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 and independent covariates such as PCP, CD4-count, gender and HIV-antigen level. The observable random vectors are

$$\{Z_i, X_i = \min(Q_i,T_i), \bar{L}_i(X_i), \tau_i = I[X_i = T_i]\}, i=1,\cdots,n$$

which we assume are independent and identically distributed. We shall assume the censoring mechanism satisfies

$$\lambda_Q(u \mid \bar{L}(u),Z,T) = \lambda_Q(u \mid \bar{L}(u),Z) \qquad (2.1)$$

where

$\lambda_Q(u \mid \cdot) \equiv \lim_{h \to 0} h^{-1} pr[u \le X < u+h, \tau=0 \mid X \ge u, \cdot]$ is the cause-specific hazard for censoring and we have suppressed the i subscript. That is, censoring and failure are independent given $\bar{L}(u)$ and Z. We shall also consider the consequences of imposing the additional assumption that

$$\lambda_Q(u \mid \bar{L}(u),Z) = \lambda_Q(u \mid Z) \qquad (2.2)$$

Eqs. (2.1) and (2.2) together imply the standard independent censoring assumption made in analyzing right censored, i.e.,

$$\lambda_T(u \mid Z, Q>u) = \lambda_T(u \mid Z) \qquad (2.3)$$

where $\lambda_T(u \mid \cdot) = \lim_{h \to 0} h^{-1} pr[u \le T \le u+h \mid T>u, \cdot]$ is the hazard for failure at u given the information in $\cdot$.

Our goal is to develop tests and estimators for the parameter β_0 of the proportional hazards model (1.1). When (2.2) holds in addition to (2.1) the usual Cox partial likelihood estimator of β_0 and the log-rank test, i.e., Cox score test of the hypothesis $\beta_0 = 0$, will be valid.

3. A NEW CLASS OF TESTS AND ESTIMATORS

3.1. Estimating the survival curve for censoring

In Section (3.2), we shall propose a modification to the usual Cox score equation that incorporates surrogate marker data. The key to our modified Cox score equation is an estimate $\hat{K}_i^w(X_i)$ of $K_i(X_i)$ with $K(t) \equiv \exp\{-\int_0^t \lambda_Q[u \mid \bar{L}(u),Z]du\}$. Our estimate is computed as follows. We suppose we have a correctly specified stratified time-dependent Cox model for censoring

$$\lambda_Q[u \mid \bar{L}_i(u),Z_i] = \lambda_{S_i^*(u)}(u)\exp[\alpha_0' W_i(u)] \qquad (3.1)$$

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, $S_i^*(u)$ is a discrete, possibly time-dependent, stratification variable depending on $\bar{L}_i(u)$ and Z_i that will, until Section 10, be set equal to Z_i, and the $\lambda_{S_i^*(u)}(u)$ 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

$$\hat{\lambda}_{S_j^*}(X_j) = (1-\tau_j) \times \left[\sum_{i=1}^n e^{\hat{\alpha}' \cdot W_i(X_j)} Y_i(X_j) I[S_i^*(X_j) = S_j^*(X_j)]\right]^{-1} \qquad (3.2a)$$

where $Y_i(u) = I[X_i \ge u]$ 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, $\hat{K}_i^w(u) \equiv$

$$\prod_{\{j;X_j \leq u, \tau_j = 0\}} \left[1 - \hat{\lambda}_{S_j^*}(X_j) e^{\hat{\alpha}' \cdot W_i(X_j)}\right]^{I[S_i^*(X_j) = S_j^*(X_j)]} . \quad (3.2b)$$

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

Eq. (2.1) and Eq. (2.2) together imply the Cox model (3.1) is guaranteed to be correctly specified with $\alpha_0=0$; nonetheless, in this setting, <u>the estimation of the coefficients α_0 that are known to be zero</u> is the key to recovery of information from the surrogate markers $L_i(u)$. Let $\hat{K}_i^0(u)$ be $\hat{K}_i^w(u)$ with $\hat{\alpha} \equiv 0$ in (3.2a) and (3.2b). $\hat{K}_i^0(u)$ is the <u>treatment-arm-specific</u> Kaplan-Meier estimator of the probability Q_i exceeds u.

3.2 <u>The Modified Cox Score Equation</u>

The Cox partial likelihood score statistic $\Sigma_i \tau_i [Z_i - \{\Sigma_j Y_j(X_i) e^{\beta Z_j} Z_j / \Sigma_j Y_j(X_i) e^{\beta Z_j}\}]$ can be written $S^0(\beta) = \Sigma_i [\tau_i \hat{K}_i^0(X_i)/\{\hat{K}_i^0(X_i)\}] \{Z_i - \bar{E}_i^0(\beta)\}$ where

$$\bar{E}_i^0(\beta) = \frac{\sum_j Z_j Y_j(X_i) e^{\beta Z_j} \hat{K}_j^0(X_i)/\{\hat{K}_j^0(X_i)\}}{\sum_j Y_j(X_i) e^{\beta Z_j} \hat{K}_j^0(X_i)/\{\hat{K}_j^0(X_i)\}} .$$

Now define $\bar{E}_i^w(\beta)$ like $\bar{E}_i^0(\beta)$ but with the $\hat{K}_j^0(X_i)$ in set braces replaced by $K_j^w(X_i)$ so

$$\bar{E}_i^w(\beta) = \frac{\sum_j Z_j Y_j(X_i) e^{\beta Z_j} \hat{K}_j^0(X_i)/\{\hat{K}_j^w(X_i)\}}{\sum_j Y_j(X_i) e^{\beta Z_j} \hat{K}_j^0(X_i)/\{\hat{K}_j^w(X_i)\}}$$

and define

$S^w(\beta) = \Sigma_i [\tau_i \hat{K}_i^0(X_i)/\{\hat{K}_i^w(X_i)\}][Z_i - \bar{E}_i^w(\beta)]$.

Thus the modified Cox score statistic $S^w(\beta)$ differs from the usual Cox score statistic $S^0(\beta)$ in that in 3 separate places $\hat{K}^w$ has been substituted for $\hat{K}^0$. Let $\bar{\beta}^w$ solve $S^w(\beta) = 0$ and $\bar{\beta}^0$ solve $S^0(\beta) = 0$. $\bar{\beta}^0$ is the usual Cox partial likelihood score estimator. We then have

<u>Proposition 1</u>: (a) Subject to regularity conditions, under (1.1), (2.1), and (3.1) there exists a solution $\bar{\beta}^w$ to $S^w(\beta) = 0$ such that $n^{1/2}(\bar{\beta}^w - \beta_0)$ is asymptotically normal with mean 0. (b) Further, if (2.2) also holds, so that $n^{1/2}(\bar{\beta}^0 - \beta_0)$ is also asymptotically normal with mean zero, then $Var^A\{n^{1/2}(\bar{\beta}^w - \beta_0)\} \leq Var^A\{n^{1/2}(\bar{\beta}^0 - \beta_0)\}$. (c) Furthermore, with α_j and $W_j(u)$, $j=1,\cdots,J$, representing the parameter and covariate vector in the j^{th} of J <u>nested</u> correctly specified Cox models (3.1) ordered by increasing dimension of $W_j(u)$, $Var^A[n^{1/2}(\bar{\beta}^{w_j} - \beta_0)] \leq Var^A[n^{1/2}(\bar{\beta}^{w_{j^*}} - \beta_0)]$ if $j^* < j$. (d) In addition, with $W_{opt}(u) \equiv [K(u)]^{-1} E[U_f(u) \mid \bar{L}(u), Z, Y(u)=1] = E[U(u) \mid \bar{L}(u), Z, Y(u) = 1]$ and $U_f(u)$ and $U(u)$ are defined in Sections (6) and (7), respectively,

$Var^A[n^{1/2}(\bar{\beta}^w - \beta_0)] \geq Var^A[n^{1/2}(\bar{\beta}^{w_{opt}} - \beta_0)]$ with strict inequality unless there exists a constant matrix b such that

$$bW(u) = W_{opt}(u) \quad . \qquad (3.3)$$

Proposition (1b) implies that any estimator $\bar{\beta}^w$ in our class is at least as efficient as the Cox partial likelihood estimator $\bar{\beta}^0$. Proposition (1c) implies that the efficiency with which we estimate β_0 can often be increased by adding additional time-dependent covariates to an already correctly specified Cox model (3.1) for censoring. However, Proposition (1d) implies that $Var^A[n^{1/2}(\bar{\beta}^{w_{opt}} - \beta_0)]$ provides a lower bound for the asymptotic variance of any estimator $\bar{\beta}^w$ in our class. $W_{opt}(u)$ depends on the joint distribution generating the data. The appendix considers adaptive estimation of $W_{opt}(u)$.

Let $\psi^0(0)$ be the standard log-rank test statistic, i.e., the Cox score test of the hypothesis $\beta_0 = 0$. Define $\psi^w(\beta) = n^{-1/2} S^w(\beta) / \Omega^w(\beta)^{1/2}$ where $\Omega^w(\beta)$ is the consistent estimator of $Var^A[n^{-1/2} S^w(\beta)]$ given in Section 7. Proposition (1b) implies that the power of $\psi^w(0)$ against local alternatives to the null hypothesis (1.2) is at least as great as that of the standard log-rank test $\psi^0(0)$ when (2.1) and (2.2) hold.

Our proof in Sec. 6 that $\bar{\beta}^w$ is asymptotically normal and unbiased for β_0 under (2.1) and (3.1) assumes that $\hat{\alpha}$ is $n^{1/2}$-consistent for the parameter α_0 of (3.1). However, results in Newey (1990) suggest $n^{1/4}$-consistency is sufficient which in principle would allow us to replace the estimate of K(u) based on model (3.1) by a nonparametric estimate (Robins and Rotnitzky, 1992), using for example kernel smoothing with appropriately chosen band width. In practice, the predictors of censoring by initiation of prophylaxis therapy Q_{1i} may be quite different from the predictors of censoring by loss to or end of follow-up Q_{2i}. In that case an investigator can specify Cox models for the separate cause-specific hazards corresponding to Q_{1i} and Q_{2i}. Then $\hat{K}_i^w(u)$ is obtained by multiplying together the cause-specific versions of the righthand side of Eq. (3.2b).

In reanalyzing Trial 002, we fit models of the form (3.1) separately for the cause-specific hazards of Q_{1i} and Q_{2i}. For both models W(u) was the vector of indicator (0,1) variables derived from $\bar{L}(u)$, (PCP1(u), PCP2(u), AZT(u), T4<20(u))$'$, where PCP1(u) = 1 if the subject has had one post-randomization episode of PCP prior to u; PCP2(u) = 1 if the subject has had two or more episodes of post-randomization PCP prior to u; AZT(u) = 1 if the subject is no longer receiving AZT at time u (usually due to toxicity); and T4<20(u) = 1 if a subject's T4 count at u is less than 20. We obtained the 95% confidence interval of .21 ± 1.96 (.32) based on $\hat\beta^w$ and of .22 ± 1.96 (.31) based on $\hat\beta^0$, indicating that the effect of dependent censoring and thus the bias in $\hat\beta^0$ was presumably small.

4. A SIMULATION STUDY UNDER INDEPENDENT CENSORING

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. Thus, we assumed (2.1) and (2.2), and thus (2.3), were true and 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 $\beta_0 = 0$, 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. For all u>0, the value of the surrogate $L_i(u)$ for subject i was obtained by adding to his/her failure time an independent uniform random variable on (-0.2, 0.2). The results reported in each row of Table 1 are based on 200 realizations. Each realization represented a trial with 300 subjects in each treatment arm.

Column 2 characterizes the covariate W(u) used in the Cox model (3.1). The covariate vector "2-levels" is W(u) = (W*(u),W*(u)Z)$'$ with W*(u) a time-independent dichotomous covariate recording whether a subject's surrogate is above the population median. The covariate vector "optimal" is $W_{opt}(u)$.

In columns 4 and 5 of Table 1, we compare the performance of the test statistics $\psi^0(\beta)$ and $\psi^w(\beta)$ by reporting the actual rejection rate of the nominal 5% tests that reject when the statistic exceeds 1.96 in absolute value. In column 6 of Table 1, we estimate the asymptotic relative efficiency (ARE) of $\psi^w(0)$ compared to $\psi^0(0)$ when β_0 = -.35 as the square of the ratio of the Monte Carlo average of $\psi^w(0)$ to that of $\psi^0(0)$. In columns 7 and 8, we estimate the asymptotic relative efficiency of $\hat\beta^w$ compared to $\hat\beta^0$. In column 7, we report the ratio of the Monte Carlo variance of $\hat\beta^0$ to that of $\hat\beta^w$. In column 8, we report the square of the ratio of the Monte Carlo interquartile range of $\hat\beta^0$ to that of $\hat\beta^w$. Table 1 demonstrates the efficiency advantage of $\hat\beta^w$ compared to $\hat\beta^0$ and the power advantage of $\psi^w(0)$ compared to $\psi^0(0)$ even when we use only the dichotomous covariate "2-levels."

5. RELATIONSHIP WITH PREVIOUS WORK

The estimators proposed in this paper are modifications of those proposed by Robins and Rotnitzky (1992). Robins (1993a) first generalized Robins and Rotnitzky (1992) by proposing an estimator $\hat\beta^w$ that differs from β^w only in that $E_i^w(\beta)$ was defined to be

$$\frac{\sum_j Z_j Y_j(X_i) e^{\beta Z_j} \hat{K}_j^0(X_i)\{\tau_j / \hat{K}_j^w(X_j)\}}{\sum_j Y_j(X_i) e^{\beta Z_j} \hat{K}_j^0(X_i)\{\tau_j / \hat{K}_j^w(X_j)\}}$$

Robins (1993a) showed that $\hat\beta^w$ was asymptotically normal and unbiased for β_0 under (1.1), (2.1), and (3.1) and was always more efficient than the Cox partial likelihood estimator $\hat\beta^0$ when (2.1) and (2.2) were both true. The asymptotic variance of the optimal estimators $\hat{\beta}^{w_{opt}}$ and $\bar{\beta}^{w_{opt}}$ are identical. However, if (3.3) and (2.2) are false, β^{w} will be asymptotically more efficient and have better small sample properties than β^{w}. It is for this reason we have recommended the use of β^{w} in this paper. The idea of using $\hat{K}^{w}_{j}(X_i)$ rather than $\hat{K}^{0}_{j}(X_j)/\tau_j$ in defining $E^{w}_{i}(\beta)$ was motivated (a) by the modification proposed by Malani (1992) to the estimators of Robins and Rotnitzky (1992), and (b) the estimators of Robins, Rotnitzky, and Zhao (1992) proposed in the context of estimating the mean function in longitudinal data with monotone missingness. Lagakos (1977), Finkelstein and Schoenfeld (1992) and Gray (1992) have taken a different approach to recovering information from surrogate marker data that applies the G-computation algorithm formula of Robins (1986, 1989). Fleming et al. (1993) proposed an estimator that requires a non-parametric regression of functions of T on $\bar{L}(u),Z,Y(u) = 1$ evaluated at the observed censoring times u. The Fleming et al. approach is closely related to the modified "Buckley-James" approach considered by Robins and Rotnitzky (1992, p. 305). Even under independent censoring (i.e., when (2.1) and (2.2) are true) the Lagakos (1977) and Finkelstein and Schoenfeld approach (1993) fail to provide an asymptotically distribution-free test of the null hypothesis (1.2). In addition, under independent censoring, the Gray, Fleming et al., Malani, and modified Buckley-James approach, although asymptotically distribution free, cannot be used with a highly multivariate $\bar{L}(u)$ with continuous components, since the use of non-parametric regression is not practical, due to the curse of dimensionality.

6. PROOF SKETCH

To avoid technical difficulties choose c^* such that $K(c^*)>\sigma>0$ with probability one for some small σ, say .001. Define $X^* = \min(T,c^*)$, $\Delta = I(X^*=T)$. Redefine $X = \min(X^*,Q)$ and $\tau = I(X\neq Q)$. We next define several random variables that are functions of T and Z only. Define the subject-specific counting process $N^*_T(u) = I[X^*\leq u,\Delta=1]$. Define $M^*_T(u) = N^*_T(u) - \int_0^u \lambda_0(x)e^{\beta_0 Z} I(X^*>x)dx$, where $\lambda_0(u) = -\partial \ell n \bar{F}_0(u)/\partial u$. $M^*_T(u)$ is a subject-specific martingale w.r.t. the filtration $\mathbf{F}_T(u) = \sigma[Z;\{I(X^*>x),N_T(x);\ x\leq u\}]$. Next $K^*(u) \equiv k^*(u,Z)$ is defined to be the probability limit of $\hat{K}^0(u)$ so $K^*(u) = K(u)$ if (2.1) and (2.2) hold. We next define several random variables that may depend on all the observables. Define $N_Q(x) = I[X\leq x, \tau=0]$, $N_T(x) = I\{X\leq x, \tau=1\}$, $Y(u) = I[X\geq u]$, $M_Q(x) \equiv N_Q(x) - \int_0^x \lambda_Q[u \mid \bar{L}(u),Z]Y(u)du$, $M_T(u) = N_T(u) - \int_0^u \lambda_0(x)e^{\beta_0 Z}Y(x)dx$. $M_Q(x)$ is, under (2.1), a subject-specific martingale with respect to the filtration $\mathbf{F}(u) = \sigma[T,Z,\bar{L}\{\min(T,u)\}, \{N_Q(x),0\leq x\leq u\}]$. $U_f(u,t) \equiv \int_u^t dM^*_T(x)K^*(x)[Z-\mathscr{L}(x)]$ where $\mathscr{L}(x) \equiv E[I(X^*>x)K^*(x)e^{\beta_0 Z}Z]/E[I(X^*>x)K^*(x)e^{\beta_0 Z}] = E[Y(x)e^{\beta_0 Z}ZR(x)]/E[Y(x)e^{\beta_0 Z}R(x)]$, where the last equality is by the fundamental identities (FI) of Sec. 8, and $R(x) = K^*(x)/K(x)$. $U_f(0,u)$ is a martingale process adapted to $\mathbf{F}_T(u)$ since $K^*(x)[Z-\mathscr{L}(x)]$ is $\mathbf{F}_T(x)$ predictable. Set $U_f(u) \equiv U_f(u,\infty)$, $U_f \equiv U_f(0)$. Define $\kappa = E\{\int_0^\infty dN^*_T(x)[Z-\mathscr{L}(x)]K^*(x)Z\} = E\left[\int_0^\infty dN_T(x)R(x)\{Z-\mathscr{L}(x)\}Z\right]$. Given model (3.1), for a random H(u), define $\mathscr{L}^Q(H,u,s) \equiv E\{H(u)Y(u)e^{\alpha_0' W(u)}I[S^*(u)=s(u)]\}/E\{Y(u)e^{\alpha_0' W(u)}I[S^*(u)=s(u)]\}$. Set $\Gamma(H) \equiv \int_0^\infty dM_Q(u)\{H(u) - \mathscr{L}^Q(H,u,S^*)\}$. Set $\bar{E}\{V(u)\} \equiv \sum_{i=1}^{n} V_i(u)/n$ for any V(u). Finally, to avoid technical difficulties associated with the tail of the censoring distribution, redefine

$$n^{-1/2}S^w(\beta) =$$

$$n^{1/2}\bar{E}[\int_0^\infty dN^*_T(u)I(Q>u)\hat{R}(u)\{Z - \bar{\mathscr{L}}(u,\beta)\}]$$

where $\hat{R}(u) = \hat{K}^0(u)/\hat{K}^w(u)$, $\bar{\mathscr{L}}(u,\beta) = \bar{E}[Ze^{\beta Z}Y(u)\hat{R}(u)]/\bar{E}[e^{\beta Z}Y(u)\hat{R}(u)]$. This agrees exactly with our previous definition of $n^{-1/2}S^w(\beta)$ whenever no failures occur after c^*. The proof of Proposition 1 relies on the following Theorem.

Theorem (6.1): Under regularity conditions, for $|\beta-\beta_0|$ of $O(n^{-1/2})$, under (1.1), (2.1), and (3.1),

$$n^{-1/2}S^w(\beta_0) = n^{-1/2}\Sigma_i U^w + o_p(1)\ , \quad (6.1)$$

$$n^{-1/2}S^w(\beta) = n^{-1/2}S^w(\beta_0) - \kappa(\beta-\beta_0) + o_p(1) \quad (6.2)$$

$$n^{1/2}(\hat{\beta}^w-\beta_0) = n^{-1/2}\Sigma_i \kappa^{-1}U^w + o_p(1)\ , \quad (6.3)$$

$n^{-1/2}S^w(\beta_0)$ and $n^{1/2}(\hat{\beta}^w-\beta_0)$ are asymptotically normal with mean 0 and asymptotic variances $Var(U^w)$ and $\kappa^{-1}Var(U^w)\kappa^{-1/}$ where

$$U^w = U_f - U_{mis}+U^w_{rec} \qquad (6.4)$$

$U_{mis}=\Gamma(H_{mis})$, with $H_{mis}(u) = U_f(u)/K(u)$; and $U^w_{rec} = \rho(U_{mis},U_w) \equiv E[U_{mis}U_w{}^{/}]\{E[U_wU_w{}^{/}]\}^{-1}U_w$, where $U_w = \Gamma(W)$ with W the function W(u).

Proof Sketch: To prove Eq. (6.1), we begin with a chain of algebraic identities. Since $Y(u) = I(X^*>u)I(Q>u)$, it follows that $n^{-1/2}S^w(\beta)$ equals

$$n^{1/2}\tilde{E}\{\int_0^\infty dM_T^*(u,\beta)I(Q>u)\hat{R}(u)[Z-\tilde{\mathscr{L}}(u,\beta)]\} \quad (6.5)$$

where $dM_T^*(u,\beta) = dN_T^*(u) - \lambda_0(u)I(X^*>u)e^{\beta Z}du$. Our key algebraic identity is that (6.5) equals

$$n^{1/2}\tilde{E}\{\int_0^\infty dM_T^*(u,\beta)\hat{K}^0(u)[Z-\tilde{\mathscr{L}}(u,\beta)]\} \qquad (6.6)$$

minus

$$n^{1/2}\tilde{E}\{\int_0^\infty dM_Q(u,\hat{\alpha})[\tilde{H}_{mis}(u,\beta) - \tilde{\mathscr{L}}^Q(\tilde{H}_{mis}(\beta),u,\hat{\alpha},S^*)]\} \qquad (6.7)$$

where $dM_Q(u,\hat{\alpha}) = dN_Q(u) - \lambda_{S^*(u)}(u)Y(u)e^{\hat{\alpha}'W(u)}du$, $\tilde{H}_{mis}(u,\beta) = \{\hat{K}^w(u)\}^{-1}\int_u^\infty dM_T^*(x,\beta)\hat{K}^0(x)[Z-\tilde{\mathscr{L}}(x,\beta)]$ and, for any H(u), $\tilde{\mathscr{L}}^Q(H,u,\hat{\alpha},s) =$

$$\frac{\tilde{E}[H(u)Y(u)e^{\hat{\alpha}'W(u)}I(S^*(u)=s(u))]}{\tilde{E}[Y(u)e^{\hat{\alpha}'W(u)}I(S^*(u)=s(u))]} .$$

With these identities in hand, our goal is now to establish that

$$n^{1/2}\tilde{E}\{\int_0^\infty dM_T^*(u)\hat{K}^0(u)[Z-\tilde{\mathscr{L}}(u)]\} = n^{1/2}\tilde{E}\{\int_0^\infty dM_T^*(u)K^*(u)[Z-\mathscr{L}(u)]\} + o_p(1) \quad , (6.8)$$

where $\tilde{\mathscr{L}}(u) \equiv \tilde{\mathscr{L}}(u,\beta_0)$, and that

$$n^{1/2}\tilde{E}[\int_0^\infty dM_Q(u,\hat{\alpha})\{\tilde{H}_{mis}(u) - \tilde{\mathscr{L}}^Q(\tilde{H}_{mis},u,\hat{\alpha},S^*)\}] = n^{1/2}\tilde{E}[\Gamma(H_{mis}) - \rho\{\Gamma(H_{mis}),\Gamma(W)\}] + o_p(1) \qquad (6.9)$$

from which Eq. (6.1) follows by the definition of U_f. Here $H_{mis}(u) = H_{mis}(u,\beta_0)$. We first establish (6.8). Now, by definition, $\hat{K}^0(u)\xrightarrow{P}K^*(u)$. Furthermore, $\tilde{\mathscr{L}}(u)\xrightarrow{P}\mathscr{L}(u)$, so

$$\hat{h}(u,Z) \equiv \hat{K}^0(u)[Z-\tilde{\mathscr{L}}(u)]\xrightarrow{P}h(u,Z) \equiv K^*(u)[Z-\mathscr{L}(u)] \quad . \qquad (6.10)$$

Eq. (6.8) would follow from (6.10) by standard martingale arguments if h(u,Z) were predictable w.r.t. $\mathbf{F}_{T,i}(u)$, $i=1,\cdots,n$. Now h(u,Z) is not predictable because of dependence on $\bar{L}(u)$. Nevertheless, under regularity conditions, it follows from Proposition 3 and Theorem 5.4 of Newey (1992) that (6.10) implies (6.8) since

$$E[\int_0^\infty dM_T(u)g(u,Z) \mid Z] = 0 \qquad (6.11)$$

for all g(u,Z). See Pugh, Robins, Lipsitz, and Harrington (1992) for regularity conditions and method of proof. We next establish (6.9). By expanding around α_0 in a Taylor series, we obtain that, under regularity conditions, the L.H.S. of (6.9) equals

$$n^{1/2}\tilde{E}[\int_0^\infty dM_Q(u)\{\tilde{H}_{mis}(u) - \tilde{\mathscr{L}}^Q(\tilde{H}_{mis},u,\alpha_0,S^*)\}] + \qquad (6.12)$$

$$n^{-1/2}\frac{\partial}{\partial\alpha'}[\text{L.H.S. of (6.9)}]\Big|_{\alpha=\alpha_0} n^{1/2}(\hat{\alpha}-\alpha_0) \quad (6.13)$$
$$+ o_p(1) \ .$$

Now $\{\tilde{H}_{mis}(u)-\tilde{\mathscr{L}}^Q(\tilde{H}_{mis},u,\alpha_0,S^*)\}\xrightarrow{P}H_{mis}(u) - \mathscr{L}^Q(H_{mis},u,S^*)$. Since $E[\int_t^\infty dM_Q(u)g(\bar{L}(u),Z,T) \mid \bar{L}(t),Z,T] = 0$ for all $g(\bar{L}(u),Z,T)$, Newey's (1992) results again imply that, under suitable regularity conditions, (6.12) - $\Gamma(H_{mis})=o_p(1)$. Further since $n^{1/2}(\hat{\alpha}-\alpha_0)= n^{1/2}\tilde{E}[E\{\Gamma(W)\Gamma(W)'\}^{-1}\Gamma(W)] + o_p(1)$ (Ritov and Wellner, 1988), then, in view of (6.13), (6.9) follows since $n^{-1/2}\frac{\partial}{\partial\alpha'}[\text{L.H.S. of (6.9)}] = E[\Gamma(H_{mis})\Gamma(W)'] + o_p(1)$. (Proof omitted.)

Under regularity conditions, (6.2) follows from a Taylor expansion of $n^{-1/2}S^w(\beta)$ around β_0 and the fact that $\partial n^{-1}S^w(\beta)/\partial\beta\xrightarrow{P}-\kappa$. (Proof omitted.) Eq. (6.3) is an immediate consequence of Eq. (6.2). The normal limiting distribution of $n^{-1/2}S^w(\beta_0)$ and $n^{1/2}(\hat{\beta}^w-\beta_0)$ follow from (6.1) and (6.3) by the central limit theorem and Slutsky's Theorem. The following is a corollary of Theorem (6.1).

Theorem (6.2): If (1.1), (3.1), (2.1), and (2.2) are true then (6.1) and (6.3) are true with $S^0(\beta_0)$ and $\hat{\beta}^0$ replacing $S^w(\beta_0)$ and $\hat{\beta}^w$ if we redefine U^w_{rec} to be zero.

Proof of Proposition 1: Proposition (1a) and (1b) follow immediately from Theorems (6.1) and (6.2). Proposition (1c) follows by Eq. (7.1) below and the fact that $Var(U^w_{rec})$ is the variance of the predicted value U^w_{rec} from the population linear regression of U_{mis} on U_w. Finally proposition (1d) follows from the fact, as shown in the proof of theorem (3.5) in Robins and Rotnitzky (1992), that, for any $W(u)$, $\mathrm{Var}(U^{w}_{rec}) \leq \mathrm{Var}(U^{w_{opt}}_{rec})$.

7. VARIANCE ESTIMATION

To derive a variance estimator we consistently estimate κ by

$$\hat{\kappa}(\hat{\beta}^{w}) = \hat{E}\left[\int_0^\infty dN_T(x)\hat{R}(x)\{Z-\hat{\mathscr{L}}(x,\hat{\beta}^{w})\}Z\right] .$$

The estimator of $\mathrm{Var}(U^{w})$ in (7.2) is derived as follows. Define $U(x) = \int_x^\infty dM_T(u)R(u)\{Z-\mathscr{L}(u)\}$, $U^{s}_{rec} = \int_0^\infty dM_Q(u)\mathscr{L}^{Q}(U,u,S^*)$, and $U_d \equiv \int_0^\infty dM_Q(u)H_{mis}(u)$. Under (2.1), it can be shown, using the FI of Sec. 8, that $U^{s}_{rec} = \int_0^\infty dM_Q(u)\mathscr{L}^{Q}(H_{mis},u,S^*)$ and $U(0) = U_f(0) - U_d$. Hence, by (6.4), we have $U^{w} \equiv U(0) + U^{s}_{rec} + U^{w}_{rec}$. Now, set $V = \mathrm{var}\{U(0)\}$, $V^{w}_{rec} \equiv \mathrm{var}(U^{w}_{rec}) = E\{U(0)U_w{}'\}\{E(U_wU_w{}')\}^{-1}E\{U_wU(0)\}$ by the FIs of Sec. 8, $V^{s}_{rec} = \mathrm{var}(U^{s}_{rec})$. We then have by the FIs

Proposition 2: $E(U^{s}_{rec}U_w{}') = 0$, $E(U(0)U^{w}_{rec}{}') = V^{w}_{rec}$, $E(U(0)U^{s}_{rec}) = -V^{s}_{rec}$. Hence

$$\mathrm{var}(U^{w}) = \mathrm{var}\{U(0) - U^{w}_{rec} + U^{s}_{rec}\} = V - V^{w}_{rec} - V^{s}_{rec}. \tag{7.1}$$

$\kappa^{-1}V\kappa^{-1\prime}$ is the asymptotic variance of the solution to $0 = S^{w}(\beta)$ if we replaced $\hat{K}^{w}(u)$ by $K(u)$ in defining $S^{w}(\beta)$. $\kappa^{-1}(V-V^{s}_{rec})\kappa^{-1\prime}$ is the asymptotic variance of the solution to $0 = \hat{S}^{w}(\beta)$ if we had replaced $\hat{\alpha}$ by α_0 in defining $\hat{K}^{w}(u)$. Hence $\kappa^{-1}V^{s}_{rec}\kappa^{-1\prime}$ is the savings in variance attributable to estimating the baseline stratum-specific hazards $\lambda_{S^*(u)}(u)$. $\kappa^{-1}V^{w}_{rec}\kappa^{-1\prime}$ is the additional savings in variance attributable to estimating the coefficients α_0 by $\hat{\alpha}$. Our estimator of $\mathrm{Var}(U^{w})$ is

$$\hat{\Omega}^{w}(\beta) = \hat{V}(\beta) - \hat{V}^{w}_{rec}(\beta) - \hat{V}^{s}_{rec}(\beta) \tag{7.2}$$

evaluated at $\hat{\beta}^{w}$ with $\hat{V}(\beta) = \hat{E}\{\hat{U}(\beta,0)^2\}$ where $\hat{U}(\beta,x) = \int_x^\infty d\hat{M}_T(u,\beta)\hat{R}(u)\{Z-\hat{\mathscr{L}}(u,\beta)\}$, $\hat{M}_T(u,\beta) = N_T(u) - \int_0^u \hat{\lambda}_0(x,\beta)Y(x)e^{\beta Z}dx$, and $\hat{\lambda}_0(x,\beta) = \hat{E}\{dN_T(x)/\hat{K}^{w}(x)\}/\hat{E}\{e^{\beta Z}Y(x)/\hat{K}^{w}(x)\}$. For any $H_1(u)$, $H_2(u)$, set $\Phi(H_1,H_2) = \hat{E}\left[\int_0^\infty dN_Q(u)\{\hat{\mathscr{L}}^{Q}(H_1H_2{}',u,\hat{\alpha},S^*) - \hat{\mathscr{L}}^{Q}(H_1,u,\hat{\alpha},S^*)\hat{\mathscr{L}}^{Q}(H_2,u,\hat{\alpha},S^*)'\}\right]$ and define $\hat{V}^{w}_{rec}(\beta) = \hat{\Phi}(\hat{U}(\beta),W)\{\hat{\Phi}(W,W)\}^{-1}\hat{\Phi}(W,\hat{U}(\beta))$, $\hat{V}^{s}_{rec}(\beta) = \int_0^\infty dN_Q(u)\{\hat{\mathscr{L}}^{Q}(\hat{U}(\beta),u,\hat{\alpha},S^*)\}^2$.

$\hat{U}(\hat{\beta}^{w},0)$ and $\hat{V}^{s}_{rec}(\hat{\beta}^{w})$ are consistent for $U(0)$ by the FIs and the fact that $\int_0^u dM_Q(x)\mathscr{L}^{Q}(U,x,S^*)$ is a martingale with respect to the filtration $\mathbf{F}^*(u) = \sigma[Z,\bar{L}\{\min(X,u)\}, \{N_Q(x),N_T(x);\ 0\leq x\leq u\}]$. The consistency of $\hat{V}^{w}_{rec}(\hat{\beta}^{w})$ for V^{w}_{rec} is proved similarly.

8. FUNDAMENTAL IDENTITIES

1) $1/K(t) = 1/K(u) + \int_u^t dx\lambda_Q(x \mid \bar{L}(x),Z)/K(x)$
2) $\tau/K(t) = 1/K(u) - (1-\tau)/K(t) + \int_u^t dx\lambda_Q(x \mid \bar{L}(x),Z)/K(x)$
3) $Y(u)\tau/K(X) = Y(u)/K(u) - \int_u^\infty dM_Q(x)/K(x)$
4) $\tau/K(X) = 1 - \int_0^\infty dM_Q(x)/K(x)$

If $K(X)>\sigma>0$ w.p.1, and (2.1) holds then, with $h()$ arbitrary,

5) $E[\tau h(T,Z)/K(X)] = E[h(T,Z)]$,
6) $E[h(T,Z,\bar{L}(u))/K(u) \mid \bar{L}(u),Z,Y(u)=1] = E[\tau h(T,Z,\bar{L}(u))/K(X) \mid \bar{L}(u),Z,Y(u)=1]$.
7) $E[\tau h(T,Z)Y(u)/K(X)]=E[h(T,Z)Y(u)/K(u)]$
8) $E[h(Z)Y(u)/K(u)] = E[h(Z)I(X^*>u)]$
9) $E\{h(T,Z)Y(u)/K(u) \mid \bar{L}(x),Z,Y(x)=1\} = E\{h(T,Z)I(X^*>u)/K(x) \mid \bar{L}(x),Z,Y(x)=1\}$.
10) $I(T>u)I(Q>u)/K(u) = I(T>u)\{1 - \int_0^u dM_Q(x)/K(x)\}$.
11) $E[U_f(x)\{K(x)\}^{-1} \mid \bar{L}(x),Z,Y(x)=1] = E[U(x) \mid \bar{L}(x),Z,Y(x)=1]$.

Remark: The first is proved by integration from which the second follows. The third and fourth follow immediately from the third. The fifth follows by multiplying both sides of the fourth 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. The sixth follows similarly from the third except we take conditional expectations. The seventh is an immediate consequence of the third. The eighth follows from the seventh and the fifth upon rewriting $\tau Y(u)$ as $\tau I(X^*>u)$. The ninth follows from the sixth and the third. The tenth follows by setting $c^* = u$ and then applying the fourth.

The eleventh follows from the ninth.

9. EFFICIENCY

Given a weight function $\omega(u)$, we can introduce the factor $\omega(X_i)$ or $\omega(u)$ as appropriate into the sum and integrals in Eqs. (3.3) and the definitions of $U_f(x)$, $U(x)$, and $U(\beta,x)$. In the absence of surrogate data, the most efficient estimator of β_0 uses the weighting function of $\omega(u) = 1$. In contrast, in the presence of surrogate data, the optimal estimator uses a weighting function $\omega_{opt}(u)$ that is non-constant in u. To make this precise, write $\beta^w(\omega)$, $U(u,\omega)$, $\hat{U}(\beta,x,\omega)$, $W_{opt}(u,\omega)$, $\psi^w(0,\omega)$ $S^w(\beta,\omega)$ to stress the dependence on ω. Let $\bar{\beta}(\omega) = \bar{\beta}^{w_{opt}(\omega)}(\omega)$, $\psi(0,\omega) = \psi^{w_{opt}(\omega)}(0,\omega)$. Results in Robins and Rotnitzky (1992) imply that the asymptotic variance of the estimator $\bar{\beta}(\omega_{opt})$ that uses $\omega_{opt}(u)$ attains the semiparametric variance bound in the semiparametric model characterized by (1.1) and (2.1). $\omega_{opt}(u)$ depends on the joint distribution of $(T,\bar{L}(T),Z,Q)$. Robins (1993b) shows that when both (2.1) and (2.2) hold so there is independent censoring, $\omega_{opt}(u)$ is the unique solution to the integral equation

$$\omega(u)=1 + \int_0^{\infty} dt\, \omega(t)h(u,t)dt \qquad (9.1)$$

where

$$h(u,t) = \{m(u)\}^{-1}E\left[\int_0^{\min(u,t)} dN_Q(x)g(u,\bar{L}(x),Z)\, g(t,\bar{L}(x),Z)\right] ,$$

$$m(u) = E[Y(u)\lambda_0(u)e^{\beta Z}\{Z - \mathscr{L}(u)\}^2] ,$$

$$g(u,\bar{L}(x),Z)du \equiv E[dM_T(u) \mid \bar{L}(x),Y(x)=1,Z]\{Z-\mathscr{L}(u)\} = \{\lambda_T(u \mid \bar{L}(x),Z) - \lambda_T(u \mid Z)\}pr[Y(u)=1 \mid \bar{L}(x),Y(x)=1,Z]\{Z-\mathscr{L}(u)\}du,$$

and $\lambda_T(u \mid \cdot) = \lim_{h\to 0} pr(u<X\leq u+h, \tau=1 \mid X\geq u,\cdot)$. The function $h(u,t)$ is called the kernel of the integral equation.

In the Appendix, we construct an estimate $\hat{\omega}_{opt}(u)$ of $\omega_{opt}(u)$ under (2.1) and (2.2) by using an integral equation equivalent to (9.1). That is,

$$\omega(u) = 1 + \{m(u)\}^{-1} E\left[\int_0^u dN_Q(x)g(u,\bar{L}(x),Z)U(x,\omega)\right] \qquad (9.2)$$

10. DISCUSSION

Suppose that, for a given choice of $W(u)$, Eq. (3.3) does not hold and thus β^w is less efficient than $\bar{\beta}^{w_{opt}}$. The efficiency of estimator β^w can in general be improved not only by adding covariates to the Cox model (3.1), but also by adding time-dependent stratification variables to $S_i^*(u)$ in model (3.1). It is straightforward to extend our results to allow Z and β_0 to be vector-valued provided Z is discrete. If Z has continuous components, then methods described in Robins and Rotnitzky (1992) can be used.

11. AN IMPROVEMENT ON THE KAPLAN-MEIER ESTIMATOR

In this final section, we provide an estimator of a treatment-arm specific survival curve that is more precise than the usual Kaplan-Meier estimator. We shall restrict attention to subjects in a particular treatment arm, say z. In the remainder of this Section, let n denote the number of subjects with Z = z, and let $\lambda_0(x)$ be the hazard of failure in arm Z = z. Furthermore to allow the reuse of our previous notation, let β now denote a time in $(0,\infty)$ rather than a parameter, regard the random variable Z as taking the value z with probability one, and set $\beta_0 = 0$ and X equal to $\min(T,Q,\beta)$ with $\tau = 1$ if X = t, $\tau = 0$ if X = Q and $\tau = -1$ if $X = \beta$. Using ideas in Robins and Rotnitzky (1992), Malani (1992) and Robins (1993a), consider the estimator $\hat{\Lambda}^w(\beta) =$

$$\sum_{i=1}^{n} I(X_i<\beta)\{\tau_i/\hat{K}_i^{w}(X_i)\}/\{\sum_{j=1}^{n} Y_j(X_i)/\hat{K}_j^{w}(X_i)\}$$

of the cumulative hazard $\Lambda(\beta) = \int_0^{\beta} \lambda_0(x)dx$ through time β in treatment arm z. Now redefine $n^{-1/2}S^w(\beta)$ to be $n^{1/2}\{\bar{\Lambda}^w(\beta)-\Lambda(\beta)\} \equiv n^{-1/2}$

$$\sum_{i=1}^{n} \int_0^{\beta} [\hat{K}_i^{w}(u)]^{-1}dM_{T,i}(u,0)/ \left\{\sum_{i=1}^{n}[\hat{K}_i^{w}(u)]^{-1}Y_i(u)/n\right\} .$$

Redefine

$$U(0) = \int_0^{\beta} [K(u)]^{-1}dM_T(u,0)/E\{[K(u)]^{-1}Y(u)\},$$

$\dot{U}(0,x) \equiv$

$$\int_x^\beta [\hat{K}_i^w(u)]^{-1} d\bar{M}_T(u,0) \Big/ \left\{ \sum_{i=1}^n [\hat{K}_i^w(u)]^{-1} Y_i(u)/n \right\}.$$

Then, by arguments exactly analogous to those given in Sections 6 and 7, we obtain

Proposition 4: Under (2.1) and (3.1), $n^{-1/2}\tilde{S}^w(\beta)$ is asymptotically normal with mean zero and variance $Var(U^w)$ that can be consistently estimated by $\Omega^w(0)$. Further, when (2.1) and (2.2) are true so that the usual Nelson estimator $\Lambda^0(\beta)$ is consistent for $\Lambda(\beta)$, $\Lambda^w(\beta)$ will always be at least as efficient as the Nelson estimator.

We obtain intervals for the survival curve $pr(T>\beta \mid Z=z)$ by exponentiating the limits of these intervals for $\Lambda(\beta)$ and inverting. When (2.2) holds, the (inefficient) interval for $pr(T>\beta \mid Z=z)$ corresponding to $\Lambda^0(\beta) \pm 1.96\, \hat{\Omega}^0(0) n^{-1/2}$ is asymptotically equivalent to the interval based on the standard Kaplan-Meier estimator. Using arguments similar to those used to prove the efficiency of $\beta(\omega_{opt})$ Robins and Rotnitzky (1992) show that the semiparametric variance bound for $\Lambda(\beta)$ is $Var^A\{n^{1/2}[\tilde{\Lambda}^{w_{opt}}(\beta) - \Lambda(\beta)]\}$.

BIBLIOGRAPHY

Andersen, P.K. & 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. & Wellner, J.A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**:432-452.

Finkelstein, E.M. & Schoenfeld, D.A. (1993). Analyzing survival in the presence of an auxiliary variable. (To appear, *Statist. Med.*).

Fleming, T.R., Prentice, R.L., Pepe, M.S. & Glidden, D. Surrogate and auxiliary end points in clinical trials, with potential applications in cancer and AIDS research. (To appear, *Statist. Med.*).

Gray, R. (1992). A kernel method for incorporating information on disease progression in the analysis of survival (unpublished manuscript).

Lagakos, S.W. (1977). Using auxiliary variables for improved estimates of survival time. *Biometrics* 33:399-404.

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

Malani, H.M. (1992). Modification of the redistribution to the right algorithm using disease markers. Technical Report 3, *Biostatistics Technical Reports*, Univ. of Calif., Berkeley.

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

Ritov, Y. & 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.

Robins, J.M. (1993a). Analytic Methods for estimating HIV-treatment and cofactor effects. In: *Methodological Issues of AIDS Behavioral Research.* pp. 213-287. Eds. Ostrow DG, Kessler R. NY: Plenum Publishing.

Robins, J.M. (1993b). A least squares approach to information recovery in proportional hazards regression using surrogate markers. (Submitted, *JASA*).

Robins, J.M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods -- application to control of the healthy worker survivor effect. *Math Modelling.* 7:1393-1512.

Robins, J.M. (1989). The control of confounding by intermediate variables. *Statist. Med.* **8**:679-701.

Robins, J.M., & Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In: *AIDS Epidemiology - Methodological Issues.* pp. 297-331. Eds: Jewell N, Dietz K, Farewell V. Birkhäuser: Boston, MA.

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

Appendix: Adaptive Estimation

We shall estimate $W_{opt}(u)$ by $\hat{W}_{opt}(u)$. To calculate $\tilde{\beta}^{\hat{w}_{opt}}$ we need only compute $\hat{W}_{opt}(u)$ at the observed censoring times u. At each censoring time u, let $\hat{W}_{opt,i}(u)$ be the predicted value for subject i from the least squares fit of $\bar{U}_j(\beta^w,u)$ on a fixed number of functions of $\bar{L}_j(u)$ and Z_j chosen by the analyst among subjects j with $\hat{S}_j^*(u) = S_i^*(u)$. Separate linear regression models may be fit at each censoring time u. As discussed in Robins and Rotnitzky (1992), (i) $\tilde{\beta}^{\hat{w}_{opt}}$ will have the same limiting distribution as

$\tilde{\beta}^{w_{opt}}$ if each of the regression models were correctly specified, and (ii) even if the models were misspecified, $\tilde{\beta}^{\hat{w}_{opt}}$ be asymptotically unbiased.

Denote by $\hat{\omega}_{opt}(u)$ the solution to (9.2) when the unknown population quantities on the right-hand side are replaced by the estimates: $\hat{m}(u) = \quad n^{-1}\Sigma_i Y_i(u)\hat{\lambda}_0^{sm}(u,\hat{\beta})e^{\hat{\beta}Z_i}\{Z_i-\bar{\mathscr{L}}(u,\hat{\beta})\}^2$, $\hat{E}[\int_0^u dN_Q(x)\hat{g}(u,\bar{L}(x),Z)\hat{U}(\beta,x,\omega)]$ where $\hat{\lambda}_0^{sm}(u,\beta)$ is a kernel-smoothed version of the usual Cox baseline hazard estimator $\hat{\lambda}_0(u,\beta)$ to insure that $\hat{m}(u)$ and thus $\hat{\omega}_{opt}(u)$ converges in probability and $\beta = \beta^w$. Hence, it only remains to define $\hat{g}(u,\bar{L}(x),Z)$. Since (i) $S^w(\omega)$ and $U(\beta,x,\omega)$ only depend on the value of $\omega(u)$ at the observed failure times u and (ii) $\hat{E}[\int_0^u dN_Q(x)g(u,\bar{L}(x),Z)U(\omega,x)]$ only depends on $g(u,\bar{L}(x),Z)$ at the observed censoring times x, we estimate $g(u,\bar{L}(x),Z)$ only at those (u,x), x<u, for which u is an observed failure time and x is an observed censoring time. Set $\hat{g}(u,\bar{L}(x),Z) = \{\hat{\lambda}_T^{sm}(u \mid \bar{L}(x),Z) - n^{-1}\Sigma_i\hat{\lambda}_T^{sm}(u \mid \bar{L}_i(x),Z_i)Y_i(u)I(Z_i=Z)/\Sigma_j Y_j(u)\}I(Z_i=Z)$ $\hat{pr}[Y(u)=1 \mid \bar{L}(x),Y(x)=1,Z]$ $\{Z-\bar{\mathscr{L}}(u,\beta)\}$, where $\hat{pr}[Y(u)=1 \mid \bar{L}(x),Y(x)=1,Z]$ is the predicted value from the fit of a logistic model for the probability that Y(u)=1 on functions of $\bar{L}(x)$ and Z among subjects with Y(x)=1 and $\hat{\lambda}_T^{sm}(u \mid \bar{L}(x),Z)$ is a kernel smoothed version of the Cox estimate of $\lambda_T(u \mid \bar{L}(x),Z)$ based on fitting, at each censoring time x, separate proportional hazards models

$$\lambda_T(u \mid \bar{L}(x),Z) = \lambda_0(u)e^{\eta' D}, u>x$$

where, in each model, $\lambda_0(u)$ is a completely unrestricted baseline hazard and D is a vector of functions of $\bar{L}(x)$ and Z. Since $\hat{\omega}_{opt}(u)$ is only defined at the observed failure times, the estimated version of (9.2) is a finite dimensional matrix equation which we now solve. Let $u_1,\cdots,u_K$ be the K ordered observed failure times and write $\hat{\omega}_{opt} = (\hat{\omega}_{opt}(u_1),\cdots,\hat{\omega}_{opt}(u_K))'$. Let $\hat{H}$ be the K×K matrix with ℓ,k element $\hat{H}_{\ell,k} = \{\hat{m}(u_k)\}^{-1}\hat{E}\left[\int_0^{u_k} dN_Q(x)\right.$ $\left. g(u_k,\bar{L}(x),Z)\{d\hat{M}_T(u_\ell,\hat{\beta})[Z-\bar{\mathscr{L}}(u_\ell,\hat{\beta})]I(u_\ell>x)\}\right]$. Let $\mathbf{1}_K$ be the column K-vectors of ones and $I_{K\times K}$ be the K×K identity matrix. Then

$$\hat{\omega}_{opt} = [I_{K\times K} + \hat{H}]^{-1}\mathbf{1}_K \qquad (9.3)$$

Having obtained $\hat{\omega}_{opt}(u)$ from (9.3) we proceed to estimate $\hat{W}_{opt}(u,\hat{\omega}_{opt})$ as above. Finally $\tilde{\beta}(\hat{\omega}_{opt})$ solves $S(\beta,\hat{\omega}_{opt}) \equiv S^{\hat{W}_{opt}(\hat{\omega}_{opt})}(\beta,\hat{\omega}_{opt}) = 0$. Under (2.1)-(2.2), $\tilde{\beta}(\hat{\omega}_{opt})$ will attain the semiparametric variance bound if the models used to adaptively estimate $\omega_{opt}(u)$ and $W_{opt}(u,\omega_{opt})$ were correctly specified. Even if they are misspecified, $\tilde{\beta}(\hat{\omega}_{opt})$ will be asymptotically normal and unbiased for β_0.

TABLE 1. Results of a Simulation Experiment

Statistic	W(u)	β_0	Actual Rejection Rate of $\psi(\beta_0)$ in %	Actual Rejection Rate of $\psi(0)$ in %	Estimated ARE of ψ^w	Median $\tilde{\beta}$	Estimated ARE of $\tilde{\beta}^w$ based on the Monte Carlo Variance of $\tilde{\beta}$	Estimated ARE of $\tilde{\beta}^w$ based on the Inter-quartile Range of $\tilde{\beta}$
S^0	-	0	4.5	4.5	-	.014	-	-
S^w	2-levels	0	5.0	5.0	-	.010	137	172
S^w	optimal	0	6.0	6.0	-	.011	162	183
S^0	-	-.35	4.5	60.0	-	-.352	-	-
S^w	2-levels	-.35	4.0	76.0	141	-.349	148	188
S^w	optimal	-.35	5.0	83.5	172	-.356	169	200