Lifetime Data Analysis, 7, 125–141, 2001

Using Auxiliary Time-Dependent Covariates to Recover Information in Nonparametric Testing with Censored Data

SUSAN MURRAY skmurray@umich.edu
Assistant Professor in the Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029

ANASTASIOS A. TSIATIS
Professor in the Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203

Received March 1, 1999; Revised September 5, 2000; Accepted October 13, 2000

Abstract. Murray and Tsiatis (1996) described a weighted survival estimate that incorporates prognostic time-dependent covariate information to increase the efficiency of estimation. We propose a test statistic based on the statistic of Pepe and Fleming (1989, 1991) that incorporates these weighted survival estimates. As in Pepe and Fleming, the test is an integrated weighted difference of two estimated survival curves. This test has been shown to be effective at detecting survival differences in crossing hazards settings where the logrank test performs poorly. This method uses stratified longitudinal covariate information to get more precise estimates of the underlying survival curves when there is censored information and this leads to more powerful tests. Another important feature of the test is that it remains valid when informative censoring is captured by the incorporated covariate. In this case, the Pepe-Fleming statistic is known to be biased and should not be used. These methods could be useful in clinical trials with heavy censoring that include collection over time of covariates, such as laboratory measurements, that are prognostic of subsequent survival or capture information related to censoring.

Keywords: Kaplan-Meier estimate, longitudinal, missing data, survival, two-sample test

1. Introduction

Information from auxiliary variables that relate to a clinical endpoint can be used to augment information provided by the clinical endpoint in treatment comparisons of overall, or marginal, survival. Some researchers have developed tests that model relationships between progression and survival within treatment specific marginal survival models. Gray (1994) considered a three-state model in which time to progression could influence survival following progression via kernel estimators. Finkelstein and Schoenfeld (1994) used a similar three state model in the semi-Markov case using parametric models for survival given progression.

Others have augmented censored failure information on overall survival with fewer assumptions on the relationship between prognostic covariates and outcome. Kosorok and Fleming (1993) combined linear rank statistics on primary and secondary endpoints. Malani (1995) suggested a covariate-based modification of Efron's redistribution to the right algorithm to estimate survival and adjust the score test from the Cox (1972) proportional hazards model. Robins and Rotnitzky (1992) presented inverse probability weighting strategies for survival estimation and testing, with weights based on models of the censoring distribution

as it related to certain prognostic covariates. Murray and Tsiatis (1996) constructed a nonparametric maximum likelihood estimator for survival involving conditional Kaplan-Meier estimates averaged over values of the covariate process. The weighted survival (WS) estimate described by Murray and Tsiatis is algebraically related to the estimate described by Malani for categorical time-dependent covariates and also related to the estimates of Robins and Rotnitzky when nonparametrically estimated inverse probability weights described in section 2.1 are applied.

These marginal survival methods differ from traditional conditional approaches to using covariates, which tend to be parametric or semi-parametric in nature and study survival according to particular risk factors rather than across the entire population of risk factors as in the marginal analyses. For instance the proportional hazards model may be used to adjust for confounding effects of imbalanced baseline covariates across treatment groups and gains efficiency in studying treatment effects, particularly when prognostic covariates are balanced across treatment groups. The Cox model is able to give correct inferences on these baseline-covariate-adjusted treatment effects when proportional hazards modeling assumptions are valid and informative censoring is correctly captured with respect to the baseline covariates in the model. However, the ability of this model to simultaneously adjust for dependent censoring and study treatment effect diminishes when there are relevant time-dependent covariates that are altered by treatment and can be considered treatment outcomes in their own right. Although time-dependent covariate Cox models are available, it is problematic to use them to study treatment effects with anything other than external time-dependent covariates as described by Kalbfleisch and Prentice (1980, pp. 122–126). The limitations occur since internal time-dependent covariates may be part of the causal pathway between the effect of treatment and survival; hence when included in the model may diminish or even remove legitimate treatment differences.

With marginal survival estimation methods, time-dependent covariates are used to augment overall survival information rather than to make a conditional adjustment. Murray and Tsiatis show that their nonparametric marginal WS estimation strategy, when applied with a prognostic stratified time dependent covariate, produces an estimate of the underlying survival distribution that has a smaller asymptotic closed form variance than the usual Kaplan-Meier (KM) estimate under uninformative censoring. When the stratified time dependent covariate also completely captures information related to the censoring mechanism, the survival estimate discussed by Murray and Tsiatis remains consistent while the standard KM survival estimate is subject to bias. Hence, the WS estimate has improved inferential properties as compared to the KM estimate while using fewer assumptions in relation to the censoring mechanism.

We propose to evaluate the difference between two integrated marginal survival curves as in Pepe and Fleming (1989, 1991), but with the covariate-augmented survival estimates proposed by Murray and Tsiatis used in place of the KM estimates used in the original statistic. Hence, the nonparametric aspect to the original test statistic is maintained, but efficiency is gained and bias reduced through use of the stratified time-dependent covariates in the WS estimates. Pepe and Fleming presented their statistic as an attractive alternative to

the logrank test against alternatives with nonproportional hazards. Since many estimates and tests in survival analysis are functionals of estimated survival probabilities, we expect that substitution of covariate-augmented survival estimates would lead to improved estimates and tests. Also, unlike the Pepe-Fleming (PF) tests, tests using the modified survival estimates will be valid when informative censoring can be completely explained by the auxiliary longitudinal covariate.

In section 2, we describe the WS estimate as proposed by Murray and Tsiatis noting required covariance calculations new to this manuscript. In section 3, we review the PF statistic and its asymptotic properties, derive the asymptotic closed form properties of the augmented PF statistic and describe asymptotic closed form gains in efficiency over the original PF test. In section 4 we present properties of the proposed statistic based on both closed form and finite sample simulation studies. Discussion follows in section 6.

2. Covariate-Augmented Weighted Survival Estimate

We first introduce the augmented survival estimate proposed by Murray and Tsiatis. Let T and C denote failure and censoring times and $X = \min(T, C)$ the observable event time. Let $S(t)$ and $H(t)$ denote survival and censoring survival functions of T and C and let $\lambda(t)$ denote the hazard function for T. Suppose a time-dependent covariate is observed on each subject at times $T_0^*, \ldots, T_s^*$, with s finite. The covariate may be continuous or categorical, but is here treated as categorical with a finite number of possible values. A continuous covariate may be stratified by using quantiles of its distribution if no biological breakpoints suggest themselves. To reduce notation we shall consider the covariates Z_i measured at times T_{i-1}^* to have $k+1$ possible values labeled $0, \ldots, k$. The estimation methods extend straightforwardly when the number of covariate strata varies at the different measuring times. One may define the covariate strata using multiple covariates or use definitions that specify how missing covariate data are handled. For instance in AIDS applications one might define Z_i as a categorical version of the last CD4 count observed for each patient at or before time T_{i-1}^*. Hence, if a patient misses a bloodwork appointment at time T_{i-1}^* his most recent measurement may be used to define his covariate strata at time T_{i-1}^*.

If we were to ignore this prognostic covariate information, we would estimate survival nonparametrically with the KM estimate (1958). But if censoring is present, Murray and Tsiatis have shown that one may extract information about survival from prognostic covariates without making additional assumptions about the data. Let θ_{i_1} be the probability that $Z_1 = i_1$ at T_0^* and let $\theta_{i_1 \ldots i_m}$ be the probability that a subject has $Z_m = i_m$ at T_{m-1}^*, conditional on the subject surviving at least to time T_{m-1}^* and previously having $Z_1 = i_1$ at T_0^*, $Z_2 = i_2$ at $T_1^*, \ldots,$ and $Z_{m-1} = i_{m-1}$ at T_{m-2}^*, and let $S_{i_1 \ldots i_m}(t)$ be the probability that a subject survives past time t, conditional on the subject surviving past time T_{m-1}^* and having $Z_1 = i_1$ at T_0^*, $Z_2 = i_2$ at $T_1^*, \ldots,$ and $Z_m = i_m$ at T_{m-1}^*, $m = 2, 3, \ldots, s+1$. Using similar subscripting conventions for the subjects with survival function, $S_{i_1 \ldots i_m}(t)$, define the corresponding censoring survival function to be $H_{i_1 \ldots i_m}(t)$ and the corresponding

failure hazard to be $\lambda_{i_1 \ldots i_m}(t)$, $m = 1, \ldots, s+1$. We may write $P(T > t) = S(t)$ as

$$\begin{cases} \displaystyle\sum_{i_1=0}^{k} S_{i_1}(t)\theta_{i_1} & 0 < t \le T_1^* \\ \displaystyle\sum_{i_1=0}^{k}\sum_{i_2=0}^{k} S_{i_1 i_2}(t)\theta_{i_1 i_2} S_{i_1}(T_1^*)\theta_{i_1} & T_1^* < t \le T_2^* \\ \vdots & \\ \displaystyle\sum_{i_1=0}^{k}\sum_{i_2=0}^{k}\cdots\sum_{i_{s+1}=0}^{k} S_{i_1 \ldots i_{s+1}}(t)\theta_{i_1 \ldots i_{s+1}} & \\ \qquad \times S_{i_1 \ldots i_s}(T_s^*)\theta_{i_1 \ldots i_s} \cdots S_{i_1 i_2}(T_2^*)\theta_{i_1 i_2} S_{i_1}(T_1^*)\theta_{i_1} & T_s^* < t. \end{cases}$$

Based on the above probability statement, the WS estimate uses all possible covariate information available at each time so that $WS(t) =$

$$\begin{cases} \displaystyle\sum_{i_1=0}^{k} \hat{S}_{i_1}(t)\frac{n_{i_1}}{n} & 0 < t \le T_1^* \\ \displaystyle\sum_{i_1=0}^{k}\sum_{i_2=0}^{k} \hat{S}_{i_1 i_2}(t)\frac{n_{i_1 i_2}}{n_{i_1 \cdot}} \hat{S}_{i_1}(T_1^*)\frac{n_{i_1}}{n} & T_1^* < t \le T_2^* \\ \vdots & \\ \displaystyle\sum_{i_1=0}^{k}\sum_{i_2=0}^{k}\cdots\sum_{i_{s+1}=0}^{k} \hat{S}_{i_1 \ldots i_{s+1}}(t)\frac{n_{i_1 \ldots i_{s+1}}}{n_{i_1 \ldots i_s \cdot}} & \\ \qquad \times \hat{S}_{i_1 \ldots i_s}(T_s^*)\frac{n_{i_1 \ldots i_s}}{n_{i_1 \ldots i_{s-1} \cdot}} \cdots \hat{S}_{i_1 i_2}(T_2^*)\frac{n_{i_1 i_2}}{n_{i_1 \cdot}} \hat{S}_{i_1}(T_1^*)\frac{n_{i_1}}{n} & T_s^* < t, \end{cases}$$

where $n_{i_1 \ldots i_m}$ is the number of people having $Z_1 = i_1$ at T_0^*, $Z_2 = i_2$ at $T_1^*, \ldots,$ and $Z_m = i_m$ at time T_{m-1}^*, $n_{i_1 \ldots i_{m-1} \cdot} = n_{i_1 \ldots i_{m-1} 0} + \cdots + n_{i_1 \ldots i_{m-1} k}$ and $\hat{S}_{i_1 \ldots i_m}(t)$ is the KM survival estimate at time t among those with past covariate values corresponding to $i_1, \ldots i_m$, $m = 1, 2, \ldots, s+1$. By estimating the separate conditional components of $S(t)$ we increase efficiency of estimation in the presence of uninformative censoring. Also, $WS(t)$ uses KM estimates that are estimated conditional on covariate strata so it remains consistent when informative censoring patterns are completely captured by the discretized covariate process. The usual KM estimate requires non-informative censoring to remain consistent. When a continuous covariate related to the censoring mechanism is stratified for use with the WS estimate, there will be some residual bias in the estimation procedure. However, in practice this bias is largely removed even when the number of covariate strata used to capture the continuous covariate is small. We shall show by simulation that the residual bias becomes negligible using only three strata to capture informative censoring related to a continuous covariate. Murray and Tsiatis described a similar phenomenon in Table 3 of their 1996 paper.

To simplify our formulas we use notation such as $S_{i_1\ldots i_p.}(t)$ for the expression $\sum_{i_{p+1}=0}^{k} S_{i_1\ldots i_{p+1}}(T^*_{p+1})\theta_{i_1\ldots i_{p+1}} \cdots \sum_{i_m=0}^{k} S_{i_1\ldots i_m}(t)\theta_{i_1\ldots i_m}$ for $T^*_{m-1} < t \leq T^*_m$. This term, which is only a function of $i_1, \ldots, i_p$, represents the conditional probability of survival past time t given $(Z_1, \ldots, Z_p) = (i_1, \ldots, i_p)$ and $T > T^*_{p-1}$. Similarly in referring to the probability of surviving past time t given only the strata information $(Z_1, \ldots, Z_p) = (i_1, \ldots, i_p)$ we shall use the simplified notation $S_{i_1\ldots i_p}(t)$ while its conditional probability equivalent is $S_{i_1\ldots i_p}(T^*_p)S_{i_1\ldots i_p.}(t)$ for $t > T^*_p$ and $S_{i_1\ldots i_p}(t)$ for $T^*_{p-1} < t \leq T^*_p$. Similar notational simplifications will be used for estimates. For instance

$$WS_{i_1\ldots i_p.}(t) = \sum_{i_{p+1}=0}^{k} \hat{S}_{i_1\ldots i_{p+1}}(T^*_{p+1}) \frac{n_{i_1\ldots i_{p+1}}}{n_{i_1\ldots i_p.}} \cdots \sum_{i_m=0}^{k} \hat{S}_{i_1\ldots i_m}(t) \frac{n_{i_1\ldots i_m}}{n_{i_1\ldots i_{m-1}.}}$$

is the WS estimate for $S_{i_1\ldots i_p.}(t)$ where $T^*_{m-1} < t \leq T^*_m$.

In constructing a test statistic we require $\mathrm{Cov}(WS(t_1), WS(t_2))$. The covariance derivation uses techniques similar to that of Murray and Tsiatis and can be found in Murray (1994, pp. 68–71). For $T^*_{m_1-1} < t_1 \leq T^*_{m_1}$, $m_1 = 1, \ldots, s$, and $T^*_{m_2-1} < t_2 \leq T^*_{m_2}$, $m_2 = 1, \ldots, s$,

$$\begin{aligned}
&\mathrm{cov}\left\{n^{\frac{1}{2}}\, WS(t_1), n^{\frac{1}{2}}\, WS(t_2)\right\} \\
&\quad = \sum_{\zeta=1}^{\min(m_1,m_2)} \sum_{i_1=0}^{k} \frac{S_{i_1}(T^*_1)}{H_{i_1}(T^*_1)}\theta_{i_1} \sum_{i_2=0}^{k} \frac{S_{i_1 i_2}(T^*_2)}{H_{i_1 i_2}(T^*_2)}\theta_{i_1 i_2} \cdots \sum_{i_{\zeta-1}=0}^{k} \frac{S_{i_1\ldots i_{\zeta-1}}(T^*_{\zeta-1})}{H_{i_1\ldots i_{\zeta-1}}(T^*_{\zeta-1})}\theta_{i_1\ldots i_{\zeta-1}} \\
&\qquad \times \left[\sum_{i_\zeta=0}^{k} \theta_{i_1\ldots i_\zeta} S_{i_1\ldots i_\zeta}(t_1) S_{i_1\ldots i_\zeta}(t_2) \int_{T^*_{\zeta-1}}^{\min(T^*_\zeta, t_1, t_2)} \frac{\lambda_{i_1\ldots i_\zeta}(u)du}{H_{i_1\ldots i_\zeta}(u) S_{i_1\ldots i_\zeta}(u)}\right. \\
&\qquad\quad \left. + \sum_{i_\zeta=0}^{k} \theta_{i_1\ldots i_\zeta} \left\{S_{i_1\ldots i_\zeta}(t_1) - S_{i_1\ldots i_{\zeta-1}.}(t_1)\right\}\left\{S_{i_1\ldots i_\zeta}(t_2) - S_{i_1\ldots i_{\zeta-1}.}(t_2)\right\}\right].
\end{aligned}$$

A little algebra in the case $t_1 = t_2$ gives the variance described by Murray and Tsiatis, which has been shown to be smaller than the variance of the KM estimate when the longitudinal covariate is prognostic and censoring is uninformative.

2.1. Inverse Probability Complete Case Survival Estimate

In addition to previous subscripts this section will denote individual values using i, $i = 1, \ldots, n$. Define $\Delta_i = I(T_i < C_i)$. The inverse probability complete case survival estimate is

$$\hat{S}_{RR}(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Delta_i I(T_i > t)}{\hat{H}_i(T_i)},$$

where $\hat{H}_i(T_i)$ estimates the censoring survival function for individual i at failure time T_i, usually using semiparametric models such as the proportional hazards model. The quality of $\hat{S}_{RR}(t)$ as an estimate is directly related to the appropriateness of the tech-

nique used to fit $\hat{H}_i(T_i)$. For instance, if the KM censoring survival estimate is used to fit $\hat{H}_i(T_i)$, then $\hat{S}_{RR}(t)$ reduces to the KM survival estimate, provided that the event rates are sufficient to give stable weights. Hence, when the KM estimate suffers from informative censoring, the KM censoring survival estimate would not be appropriate for $\hat{H}_i(T_i)$ and would result in biased $\hat{S}_{RR}(t)$. When incorporating covariates, a proportional hazards model is generally advocated to model $\hat{H}_i(T_i)$. In each case the inverse weights favor certain types of individuals, i, depending on the value of $\hat{H}_i(T_i)$. Favoring individuals in the wrong proportion through mismodeling of $\hat{H}_i(T_i)$ gives biased results. On the other hand, if the modeling assumptions are correct, then the semiparametrically constructed $\hat{H}_i(T_i)$ will result in more efficient, less-biased survival estimates $\hat{S}_{RR}(t)$, as opposed to nonparametrically constructed survival estimates using categorical representations of the same covariates. When appropriate, semiparametric censoring survival estimation also avoids curse of dimensionality problems that can occur when using too many covariate strata to capture the covariate information nonparametrically.

The following calculations demonstrate that $WS(t)$ reduces to $\hat{S}_{RR}(t)$ for $T^*_{m-1} < t \leq T^*_m$ when $\hat{H}_i(T_i)$ is constructed nonparametrically using categorical time-dependent covariate strata information on individual i of $i_1, i_2, \ldots i_m$ at $T^*_0, T^*_1, \ldots T^*_{m-1}$. This information is always available on relevant individuals with $T_i > t$. In this case the most efficient $\hat{H}_i(T_i)$ estimate is $\hat{H}_{i_1}(T^*_1)\hat{H}_{i_1 i_2}(T^*_2)\cdots\hat{H}_{i_1\ldots i_{m-1}}(T^*_{m-1})\hat{H}_{i_1\ldots i_m}(T_i)$. For convenience, group individual data according to covariate path so that, for example, T_i such that $T_i > t$ may be represented using $T_{i_1\ldots i_m j}$ for covariate strata values in $(0,\ldots,k)$ and j in $(1,\ldots,n_{i_1\ldots i_m})$. So

$$
\begin{aligned}
\hat{S}_{RR}(t) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta_i I(T_i > t)}{\hat{H}_i(T_i)} \\
&= \frac{1}{n}\sum_{i_1=0}^{k}\sum_{i_2=0}^{k}\cdots\sum_{i_m=0}^{k}\sum_{j=1}^{n_{i_1\ldots i_m}}\frac{\Delta_{i_1\ldots i_m j} I(T_{i_1\ldots i_m j} > t)}{\hat{H}_{i_1}(T^*_1)\cdots\hat{H}_{i_1\ldots i_{m-1}}(T^*_{m-1})\hat{H}_{i_1\ldots i_m}(T_{i_1\ldots i_m j})} \\
&= \frac{1}{n}\sum_{i_1=0}^{k}\sum_{i_2=0}^{k}\cdots\sum_{i_m=0}^{k}\frac{n_{i_1\ldots i_m}}{\hat{H}_{i_1}(T^*_1)\cdots\hat{H}_{i_1\ldots i_{m-1}}(T^*_{m-1})} \\
&\quad\times\left\{\frac{1}{n_{i_1\ldots i_m}}\sum_{j=1}^{n_{i_1\ldots i_m}}\frac{\Delta_{i_1\ldots i_m j} I(T_{i_1\ldots i_m j} > t)}{\hat{H}_{i_1\ldots i_m}(T_{i_1\ldots i_m j})}\right\} \\
&= \frac{1}{n}\sum_{i_1=0}^{k}\sum_{i_2=0}^{k}\cdots\sum_{i_m=0}^{k}\frac{n_{i_1\ldots i_m}}{\hat{H}_{i_1}(T^*_1)\cdots\hat{H}_{i_1\ldots i_{m-1}}(T^*_{m-1})}\hat{S}_{i_1\ldots i_m}(t).
\end{aligned}
$$

The connection to $WS(t)$ follows upon noting $\hat{H}_{i_1\ldots i_p}(T^*_p) = (n_{i_1\ldots i_p .}/n_{i_1\ldots i_p})\{\hat{S}_{i_1\ldots i_p}(T^*_p)\}^{-1}$, for p in $1,\ldots,m-1$.

From a purely operational point of view, by relating survival to prognostic covariates as with the WS estimator, we have the ability to immediately define covariates that are relevant to both bias reduction and efficiency gain. In modeling the censoring distribution

during construction of $\hat{S}_{RR}(t)$, it is straightforward to identify covariates that are prognostic for censoring. However, if these covariates are not related also to survival, the adjusted survival estimate although still unbiased will be less efficient than the unadjusted survival estimate based on results from Murray and Tsiatis. Hence, some knowledge of prognosis for survival is required in constructing $\hat{S}_{RR}(t)$ efficiently. Information from covariates that are prognostic for survival and not censoring also do not naturally enter into the modeling of the censoring distribution.

3. Nonparametric Covariate-Augmented Test Statistic

Now using the *WS* estimate we define a statistic for testing differences in survival based on Pepe and Fleming's 1989 $\mathcal{T}_1$ statistic,

$$\mathcal{T}_1 = \left(\frac{n_1 n_2}{n}\right)^{\frac{1}{2}} \int_0^{\tau} \hat{w}(t) \left\{\hat{S}_1(t) - \hat{S}_2(t)\right\} dt,$$

where n_g is the sample size in group g for $g = 1, 2, n = n_1 + n_2$ and $\hat{S}_g(t)$ is the KM survival estimate in group g. The upper limit of integration, τ, must be chosen so that the survival estimates are consistent within the domain of integration. If the largest event time from a particular group is censored, then there is no consistent survival estimate beyond this last event time and the test statistic requires τ less than this last censored event. If both groups have censored events as their largest event times, τ can be at most the smallest of these two times. These restrictions on τ assure a stable test statistic for any reasonable choice of weights, $\hat{w}(t)$. In the original presentation by Pepe and Fleming, stability was ensured by requiring $\hat{w}(t) = 0$ for values of $t > \tau$. To simplify notation, let $A(t) = \int_t^{\tau} w(u)S(u)\,du$. Pepe and Fleming showed that $\mathcal{T}_1$ converges in distribution to a normal random variable with mean zero and variance

$$\sigma^2 = \sum_{g=1}^{2} \frac{\pi_1 \pi_2}{\pi_g} \int_0^{\tau} \frac{A^2(t)\lambda(t)}{S(t)H_g(t)}\, dt,$$

where π_g is the probability of falling in group g and $H_g(t)$ is the censoring survival function for group g. The augmented statistic replaces the KM estimate with the *WS* estimate giving

$$\mathcal{T}_2 = \left(\frac{n_1 n_2}{n}\right)^{\frac{1}{2}} \int_0^{\tau^*} \hat{w}(t) \{WS_1(t) - WS_2(t)\}\, dt,$$

where $WS_g(t)$ is the covariate modified survival estimate for group g introduced in section 2. The upper limit of integration, τ^*, must be constrained so that all survival estimates used in constructing the *WS* estimates are consistent for each treatment group within the area of integration. Asymptotically, τ^* is equivalent to τ. The number and choice of covariate strata may be different for each treatment, but for notational simplicity we take the number of strata to be the same.

In the testing framework, we indicate treatment group by the "g" subscript prior to the covariate related subscripts, such as $i_1 \ldots i_\zeta$. Define $A_{gi}(x) = \int_x^{\tau^*} w(y)S_{gi}(y)dy$, let

$\bar{A}_g(x) = \sum \theta_{gi} A_{gi}(x)$ and $D_{gi}(x) = A_{gi}(x) - \bar{A}_g(x)$. We first derive the variance of $\mathcal{T}_2$ when the covariates are only measured at baseline. In this case

$$\begin{aligned}
\text{var}(\mathcal{T}_2) &= \text{var}\left[\left(\frac{n_1 n_2}{n}\right)^{\frac{1}{2}} \int_0^{\tau^*} \hat{w}(t)\{WS_1(t) - WS_2(t)\}\, dt\right] \\
&\approx \sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \int_0^{\tau^*}\int_0^{\tau^*} w(x)\, w(y)\, \text{cov}\left\{n_g^{\frac{1}{2}}\, WS_g(x), n_g^{\frac{1}{2}}\, WS_g(y)\right\} dx\, dy \\
&= \sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \int_0^{\tau^*}\int_0^{\tau^*} w(x)\, w(y) \left[\sum_{i=0}^{k} \theta_{gi} S_{gi}(x) S_{gi}(y) \int_0^{\min(x,y)} \frac{\lambda_{gi}(u)\, du}{H_{gi}(u) S_{gi}(u)}\right. \\
&\quad \left. + \sum_{i=0}^{k} \theta_{gi} \left\{S_{gi}(x) - \bar{S}_g(x)\right\}\left\{S_{gi}(y) - \bar{S}_g(y)\right\}\right] dx\, dy \\
&= \sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \int_{x=0}^{\tau^*}\int_{y=x}^{\tau^*} w(x)\, w(y) \left\{\sum_{i=0}^{k} \theta_{gi} S_{gi}(x) S_{gi}(y) \int_{u=0}^{x} \frac{\lambda_{gi}(u)\, du}{H_{gi}(u) S_{gi}(u)}\right\} dy\, dx \\
&\quad + \sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \int_{x=0}^{\tau^*}\int_{y=0}^{x} w(x)\, w(y) \left\{\sum_{i=0}^{k} \theta_{gi} S_{gi}(x) S_{gi}(y) \int_{u=0}^{y} \frac{\lambda_{gi}(u)\, du}{H_{gi}(u) S_{gi}(u)}\right\} dy\, dx \\
&\quad + \sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \int_0^{\tau^*}\int_0^{\tau^*} w(x)\, w(y) \\
&\quad \times \left[\sum_{i=0}^{k} \theta_{gi} \left\{S_{gi}(x) - \bar{S}_g(x)\right\}\left\{S_{gi}(y) - \bar{S}_g(y)\right\}\right] dy\, dx.
\end{aligned}$$

After some additional calculation this asymptotic variance becomes

$$\sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \sum_{i=0}^{k} \theta_{gi} \left\{\int_{u=0}^{\tau^*} \frac{A_{gi}^2(u)\lambda_{gi}(u)\, du}{H_{gi}(u) S_{gi}(u)} + D_{gi}^2(0)\right\},$$

which can be estimated by substituting maximum likelihood estimates for the unknown quantities, including those found in $A_{gi}(u)$ and $D_{gi}(0)$. Specifically, let $\hat{\pi}_g = \frac{n_g}{n}$,$\hat{\theta}_{gi} = \frac{n_{gi}}{n_g}$, $\tilde{N}(t)$ be the observed number of deaths at time t and $\tilde{Y}(t)$ be the observed number of individuals still at risk at time t. This notation will be subscripted similarly to past notation depending on the covariate path under consideration. Estimate $A_{gi}(u)$ with $\hat{A}_{gi}(u) = \int_u^{\tau^*} w(y)\hat{S}_{gi}(y)\, dy$ and $D_{gi}(0)$ with $\hat{D}_{gi}(0) = \hat{A}_{gi}(0) - \sum \theta_{gi}\hat{A}_{gi}(0)$. Then

$$\widehat{\text{var}}(\mathcal{T}_2) = \sum_{g=1}^{2} \frac{\hat{\pi}_1\hat{\pi}_2}{\hat{\pi}_g} \sum_{i=0}^{k} \hat{\theta}_{gi} \left[\int_{u=0}^{\tau^*} \frac{\hat{A}_{gi}^2(u) d\tilde{N}_{gi}(u) n_{gi}}{\tilde{Y}_{gi}(u)\{\tilde{Y}_{gi}(u) - \Delta\tilde{N}_{gi}(u)\}} + \hat{D}_{gi}^2(0)\right].$$

Variance calculations for time-dependent covariates become more complicated. The time axis can be described in reference to the intervals created by $T_0^*, \ldots, T_s^*$. Define the first

time interval for $t \in [0, T_1^*]$, the second time interval for $t \in (T_1^*, T_2^*], \ldots,$ the m^{th} time interval for $t \in (T_{m-1}^*, T_m^*], \ldots,$ the s^{th} time interval for $t \in (T_{s-1}^*, T_s^*]$ and the $s+1^{st}$ interval for $t \in (T_s^*, \tau^*]$. Then

$$\begin{aligned}
\mathrm{var}(\mathcal{T}_2) &= \mathrm{var}\left[\left(\frac{n_1 n_2}{n}\right)^{\frac{1}{2}} \int_0^{\tau^*} \hat{w}(t)\,\{WS_1(t) - WS_2(t)\}\,dt\right] \\
&\approx \sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \int_0^{\tau^*}\int_0^{\tau^*} w(x)\,w(y)\,\mathrm{cov}\left\{n_g^{\frac{1}{2}}\,WS_g(x), n_g^{\frac{1}{2}}\,WS_g(y)\right\} dx\,dy \\
&= \sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \sum_{m_1=1}^{s+1}\sum_{m_2=1}^{s+1} \int_{T_{m_1-1}^*}^{\min(T_{m_1}^*,\tau^*)} \int_{T_{m_2-1}^*}^{\min(T_{m_2}^*,\tau^*)} w(x)\,w(y) \\
&\quad \times \mathrm{cov}\left\{n_g^{\frac{1}{2}}\,WS_g(x), n_g^{\frac{1}{2}}\,WS_g(y)\right\} dx\,dy \\
&= \sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \sum_{m_1=1}^{s+1}\sum_{m_2=1}^{s+1} \int_{T_{m_1-1}^*}^{\min(T_{m_1}^*,\tau^*)} \int_{T_{m_2-1}^*}^{\min(T_{m_2}^*,\tau^*)} w(x)\,w(y) \qquad (1) \\
&\quad \times \sum_{\zeta=1}^{\min(m_1,m_2)} \sum_{i_1=0}^{k} \frac{S_{gi_1}(T_1^*)}{H_{gi_1}(T_1^*)}\theta_{gi_1} \sum_{i_2=0}^{k}\cdots\sum_{i_{\zeta-1}=0}^{k} \frac{S_{gi_1\ldots i_{\zeta-1}}(T_{\zeta-1}^*)}{H_{gi_1\ldots i_{\zeta-1}}(T_{\zeta-1}^*)}\theta_{gi_1\ldots i_{\zeta-1}} \qquad (2) \\
&\quad \times \left[\sum_{i_\zeta=0}^{k} \theta_{gi_1\ldots i_\zeta} S_{gi_1\ldots i_\zeta}(x) S_{gi_1\ldots i_\zeta}(y) \int_{T_{\zeta-1}^*}^{\min(T_\zeta^*,x,y)} \frac{\lambda_{gi_1\ldots i_\zeta}(u)\,du}{H_{gi_1\ldots i_\zeta}(u) S_{gi_1\ldots i_\zeta}(u)} \right. \qquad (3) \\
&\quad \left. + \sum_{i_\zeta=0}^{k} \theta_{gi_1\ldots i_\zeta}\{S_{gi_1\ldots i_\zeta}(x) - S_{gi_1\ldots i_{\zeta-1}.}(x)\}\{S_{gi_1\ldots i_\zeta}(y) - S_{gi_1\ldots i_{\zeta-1}.}(y)\}\right] dx\,dy. \qquad (4)
\end{aligned}$$

To reduce notation we define $A_{i_1\ldots i_\zeta}(x) = \int_x^{\tau^*} w(y) S_{i_1\ldots i_\zeta}(y)dy$, where $S_{i_1\ldots i_\zeta}(y)$ takes on the appropriate values in the appropriate areas of integration. Similarly, define $\bar{A}_{i_1\ldots i_{\zeta-1}}(x) = \sum_{i_\zeta=0}^{k} \theta_{i_1\ldots i_\zeta} A_{i_1\ldots i_\zeta}(x)$ and $D_{i_1\ldots i_\zeta}(x) = A_{i_1\ldots i_\zeta}(x) - \bar{A}_{i_1\ldots i_{\zeta-1}}(x)$. After further calculations available in Murray (1994, pp. 71–72), we find

$$\begin{aligned}
\mathrm{var}(\mathcal{T}_2) &= \mathrm{var}\left[\left(\frac{n_1 n_2}{n}\right)^{\frac{1}{2}} \int_0^{\tau^*} w(t)\,\{WS_1(t) - WS_2(t)\}\,dt\right] \\
&= \sum_{g=1}^{2} \frac{\pi_1\pi_2}{\pi_g} \sum_{\zeta=1}^{s+1}\sum_{i_1=0}^{k} \frac{S_{gi_1}(T_1^*)}{H_{gi_1}(T_1^*)}\theta_{gi_1}\cdots\sum_{i_\zeta=0}^{k}\theta_{gi_1\ldots i_\zeta} \\
&\quad \times \left\{\int_{u=T_{\zeta-1}^*}^{\min(T_\zeta^*,\tau^*)} \frac{A_{gi_1\ldots i_\zeta}^2(u)\lambda_{gi_1\ldots i_\zeta}(u)\,du}{H_{gi_1\ldots i_\zeta}(u) S_{gi_1\ldots i_\zeta}(u)} + D_{gi_1\ldots i_\zeta}^2(T_{\zeta-1}^*)\right\}.
\end{aligned}$$

This variance can be estimated similarly to the time independent covariate case. We substitute maximum likelihood estimates for the unknown quantities, including those found in

$A_{gi_1\ldots i_\zeta}(u)$ and $D_{gi_1\ldots i_\zeta}(T^*_{\zeta-1})$. Then

$$\hat{\text{var}}(\mathcal{T}_2) = \sum_{g=1}^{2} \frac{\hat{\pi}_1\hat{\pi}_2}{\hat{\pi}_g} \sum_{\zeta=1}^{s+1} \sum_{i_1=0}^{k} \frac{\hat{S}_{gi_1}(T^*_1)}{\hat{H}_{gi_1}(T^*_1)} \hat{\theta}_{gi_1} \cdots \sum_{i_\zeta=0}^{k} \hat{\theta}_{gi_1\ldots i_\zeta}$$
$$\times \left\{ \int_{u=T^*_{\zeta-1}}^{\min(T^*_\zeta,\tau^*)} \frac{\hat{A}^2_{gi_1\ldots i_\zeta}(u) d\tilde{N}_{gi_1\ldots i_\zeta}(u) n_{gi_1\ldots i_\zeta}}{\tilde{Y}_{gi_1\ldots i_\zeta}(u)(\tilde{Y}_{gi_1\ldots i_\zeta}(u) - \Delta\tilde{N}_{gi_1\ldots i_\zeta}(u))} + \hat{D}^2_{gi_1\ldots i_\zeta}(T^*_{\zeta-1}) \right\}.$$

It can be shown that the variance of $\mathcal{T}_2$ is asymptotically smaller than $\mathcal{T}_1$, when the incorporated covariates are prognostic and censoring is uninformative. The variance of $\mathcal{T}_2$ also decreases asymptotically with each additional prognostic covariate measurement included in the estimate. If the covariate is related to neither the outcome nor the censoring mechanism, the augmented test statistic is asymptotically equivalent to the original PF test. Details regarding these results can be found in Murray's 1994 thesis. Additional remarks pertaining to asymptotic normality of the adjusted test statistic appear in the appendix.

4. Simulations

Simulations were constructed to investigate power and size properties for increasing levels of censoring (40%, 50%, 60%) using the identity weight function. This weight function allows the test statistic to be interpreted in relation to the average years of life saved during the integrated time region while on treatment 1. To verify size in the baseline covariate case, 1000 simulations were run under the null hypothesis using a sample size of 200 per treatment group. A dichotomous covariate with $\theta_{gi_1} = .5$ for $i_1 = 1, 2$ was generated for each treatment group $g = 1, 2$. Exponential failure times with failure hazards changing fivefold depending on covariate strata were generated using parameters $\lambda_{g1} = .1$ and $\lambda_{g2} = .5$, for $g = 1, 2$. Non-informative censoring was generated using the Uniform distribution. Comparable size results of (0.047, 0.053, 0.050) and (0.049, 0.053, 0.047) were observed for the $\mathcal{T}_2$ statistic and the PF statistic, respectively, in increasing order of censoring percentages.

For 200 subjects per treatment with differing marginal survival functions, properties of the test statistics incorporating time-independent covariates were also studied using the exponential distribution with hazards increasing fivefold depending on covariate strata membership. Within each treatment, the proportions falling into two different covariate strata were 0.5 with survival hazards within strata 1 and 2 of (0.15, 0.75) and (0.1, 0.5) for treatments 1 and 2, respectively. To study asymptotic relative efficiencies (AREs) and power of the tests under informative censoring, simulations were generated with uniform censoring. In this case both augmented and unaugmented tests are valid. For increasing censoring percentages, power of the original PF statistic was (0.60, 0.57, 0.55) as opposed to (0.66, 0.63, 0.60) incorporating the covariate information in $\mathcal{T}_2$ after 5000 Monte Carlo simulations. Corresponding AREs calculated using closed form variances with Maple software were (1.08, 1.08, 1.10) for these increasing censoring percentages, verifying that modest gains in power and efficiency can be had by incorporating the covariates nonparametrically.

Table 1. Relative MSE and bias under informative censoring.

Percent Censoring	Pepe-Fleming Bias	Relative MSE	
		$N = 200$	$N = 500$
40%	0.43	1.61	2.44
50%	0.67	2.09	3.68
60%	1.02	2.17	4.17

1. Calculations based on closed form results, 2 fold increase in censoring hazards between covariate strata

Similar unreported results were seen in studying a second alternative where strata-specific survival hazards matched across treatment groups with $\lambda_{g1} = .1$ and $\lambda_{g2} = .5$ for $g = 1, 2$, but the proportion of patients falling into the two strata shifted slightly between the two treatments with $(\theta_{11}, \theta_{12}) = (0.65, 0.35)$ and $(\theta_{21}, \theta_{22}) = (0.35, 0.65)$. This scenario could occur if the time of the first prognostic biomarker measurement, T_0^*, occurs soon after the treatment is given and before anyone has dropped out of the study. The treatment benefit is then caused by the differing presence of the biomarker between the two treatments, rather than differences between survival hazards in corresponding covariate strata.

Another issue worth investigation is the assessment of treatment effect when censoring depends on the covariate. For the time independent covariate case we studied an instance where informative censoring was present, but at a level perhaps too subtle to detect, with parameters $(\lambda_{g1}, \lambda_{g2}) = (0.1, 0.5)$ for $g = 1, 2$, $(\theta_{11}, \theta_{12}) = (0.65, 0.35)$ and $(\theta_{21}, \theta_{22}) = (0.35, 0.65)$ and an exponential censoring distribution with censoring hazard for one strata 2.0 times greater than the censoring hazard for the other strata.

The estimate of the integrated difference in survival probabilities is unbiased for the statistic augmented by covariates. However, the PF test will be affected through the mismeasurement of the treatment effect. To compare the bias and variance of the augmented test with those of the PF test, we looked at ratios of Mean Squared Error (MSE) of the test statistics estimating $\int_0^{10} \{S_1(t) - S_2(t)\}\, dt$. We also display the bias of the PF version of this difference. All calculations are done in closed form using Maple software. In each case the true difference in integrated survival probabilities is 1.30, or an average of 1.30 years of life saved during the first 10 years on study. Table 1 shows results. The bias using the PF statistic increases with the percentage of censoring in the data. The PF integrated difference in survival is inflated by 33, 52 and 78 percent for the increasing percentages of censoring. MSE ratios become higher when censoring is more informative and when there is a higher percentage of censoring. Simulations were also conducted under the null hypothesis when censoring was informative through the covariate as described above. For increasing levels of censoring, the type I errors of the $\mathcal{T}_2$ statistic were (0.049, 0.062, 0.059) while the observed PF statistic type I errors (0.172, 0.297, 0.425) increased in tandem with the levels of censoring present in the data.

In the case a continuous covariate is related to the censoring mechanism a stratified form of the covariate may not eliminate all bias when incorporated into the test statistic. However, the following simulation suggests that any remaining bias will be small. We

Table 2. Informative censoring with a continuous covariate.

	Pepe-Fleming	$\mathcal{T}_2$			
		2 strata	3 strata	4 strata	5 strata
Bias of mean difference	0.574	0.194	0.099	0.060	0.039
Bias of standardized statistic	1.31	0.454	0.233	0.144	0.090
Size of test	0.253	0.072	0.056	0.052	0.051

1. 5000 replications, sample size of 200 each

simulated a continuous covariate along with dependent failure and censoring distributions where only 50% of deaths were observed. For the failure distribution a fivefold difference in hazard existed for subjects at the 25th percentile of the the covariate compared to subjects at the 75th percentile of the covariate. Similarly for the censoring distribution a twofold difference in hazard existed for subjects at these different percentiles. Bias of the difference in integrated survival curves, bias of the standardized test statistic and size calculations were calculated under the null hypothesis. Results are located in Table 2. Note that the size of the test based on the augmented statistic is very close to 0.05 using merely 3 prognostic covariate strata. But the PF test remains strongly affected by the informative censoring. Murray and Tsiatis (1996) showed that most gains in survival estimation efficiency are made after the continuous covariate is broken into 2 to 3 categories per covariate look across time. This gain translates to the tests as well.

Further unreported simulations were conducted to study the behavior of the various test statistics incorporating a time-dependent covariate. The reported baseline covariate simulations versus the unaugmented test are indicative of what occurs as additional covariate measurements are included in the test. Additional efficiency gain occurs with each additional prognostic covariate look, with the degree of gain related to the prognostic value added from measuring the covariate more than once across time. If the time-dependent covariate is connected to the censoring mechanism, bias will occur unless all relevant covariate information is incorporated in the test statistic.

5. Example

We now apply the $\mathcal{T}_2$ statistic to an AIDS example where patients were assigned either low dose ($n = 262$) or high dose ($n = 262$) zidovudine regimens, Fischl et al. (1990). For illustration we analyze the data at an earlier calendar date, focusing on the first 22 months on study, in order to provide our testing methods with an increased level of censoring (56%).

We use the PF and $\mathcal{T}_2$ statistics with the identity weight to detect whether one treatment is superior to the other in terms of days of life saved up to $\tau = 22$ months. In this trial CD4 count and hemoglobin level were known to be modestly predictive of survival, so we incorporate these covariates into the $\mathcal{T}_2$ test statistic in categorical form. Because of the restrictive size in each of the two treatment groups, the number of strata constructed will be conservative in order to maintain $\tau = 22$ as a proper upper limit of integration. After mild exploratory analysis, binary covariates were constructed at baseline and 200 days based on

CD4 and hemoglobin measurements taken at these times. At the first analysis time, patients with a better prognosis had either higher CD4 counts and moderate to high hemoglobin counts or smaller CD4 counts compensated with high hemoglobin counts. Within each of these two baseline strata, similar definitions for better prognosis patients were used based on the CD4 and hemoglobin counts available at 200 days (T_1^*). So $Z_1 = 1, 2$ and $Z_2 = 1, 2$ for a total of four possible covariate paths.

Among patients in the low dose treatment group who continued to be at risk at time T_1^* and were changing to a poorer prognosis path at that time, there was a slightly greater tendency to become censored after participating in the study for 15 months as can be seen in Figure 1. This difference in censoring patterns was not statistically significant, however it is interesting to compare whether this late term censoring effect has any bearing on the treatment difference detected. The binary covariate, Z_1, had virtually no predictive value until day 225 or so in this study. After day 225, both Z_1 and Z_2 are prognostic for survival. Following the recommendations of Murray and Tsiatis (1996), we shall use the flexibility allowed us in defining the time dependent covariates to gain the efficiency that is possible to gain after day 225 from prognostic covariate information without the penalty of defining unprognostic covariates early on. That is we shall define the new time dependent covariates $Z_1^* = 1$ and

$$Z_2^* = \begin{cases} 1 & \text{if } (Z_1 = 1 \text{ and } Z_2 = 1) \\ 2 & \text{if } (Z_1 = 1 \text{ and } Z_2 = 2) \\ 3 & \text{if } (Z_1 = 2 \text{ and } Z_2 = 1) \\ 4 & \text{if } (Z_1 = 2 \text{ and } Z_2 = 2) \end{cases}$$

to incorporate into our analysis. At this analysis time during the first 22 months on study, the PF statistic estimated the average number of days saved on the low dose treatment arm to be 31.50 ± 30.85 (p-value 0.0453). The $\mathcal{T}_2$ test statistic incorporating information only from the baseline CD4 and hemoglobin predictor, Z_1, estimated 31.24 ± 30.76 days of life saved (p-value 0.0465). Using the more flexible time-dependent covariate definitions Z_1^* and Z_2^*, the $\mathcal{T}_2$ test statistic estimated 31.98 ± 30.72 days of life saved (p-value 0.0413), where the previous estimated average days of life saved are displayed along with their surrounding 95% confidence intervals. All inferences in this example were essentially similar using these three test statistics, assuring us that the potential informative censoring observed in Figure 1 had no discernible effect on the original unaugmented analysis. As a comparison, the estimated days of life saved during the first 22 months on study was 39.28 ± 30.15 (p-value 0.0107) incorporating information from Z_1^* and Z_2^* at a later calendar time when censoring during the first 22 months was reduced to 11.4%. Hence, all previous estimates subject to 56% censoring were 7 or 8 days different from the estimate that used more complete survival information.

6. Discussion

By incorporating additional information through prognostic covariates, the $\mathcal{T}_2$ statistic asymptotically increases the efficiency of survival estimation and hence increases power

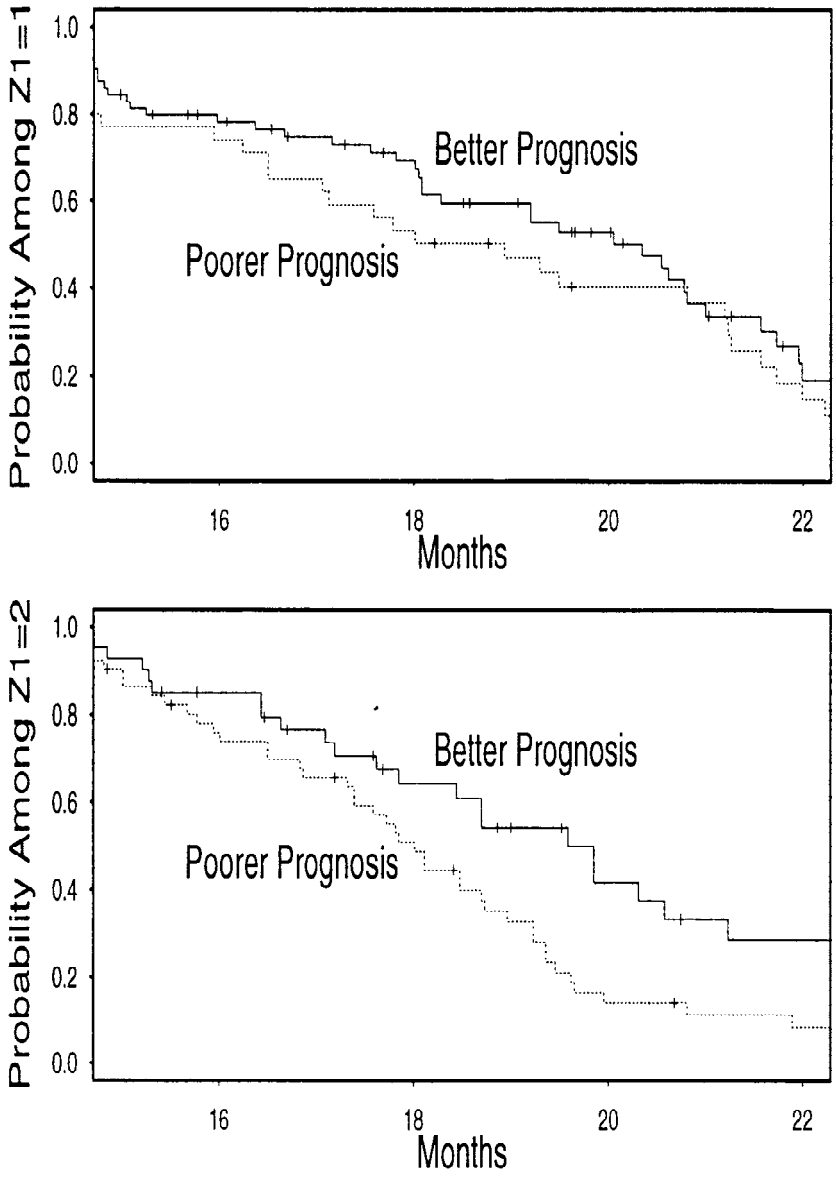

Figure 1. Censoring survival probabilities conditional on Z_1 and $T > T_1^*$ in AIDS data example.

slightly above that of the PF statistic. A comparison of the closed-form asymptotic variances of the statistics, found in Murray's thesis, confirms this. The proposed test statistic also eliminates informative censoring when the informativeness is captured by the incor-

porated covariates. Hence, the augmented test statistic provides a testing procedure for use in cases where traditional nonparametric tests are subject to bias. All these results are achieved without parametric assumptions. We have found that gains are dependent on the prognostic value of the covariates incorporated as well as the level of censoring in the data. Although the relationship between the variances of the augmented statistic and the PF statistic is complicated, we have observed higher gains in efficiency for higher percentages of censoring that reach a plateau and then decrease gradually as data become too sparse to retrieve information regarding survival behavior in the various covariate strata.

Even more striking are results from the bias and MSE ratio calculations described in the simulation section. With subtle informative censoring present in the data, large biases were identified for the PF test. The level of information present in the censoring mechanism was probably too low to detect in any small to moderately sized data set. This is a strong argument for using our testing methods since informative censoring is handled more correctly in questionable situations and there is potential to gain efficiency as well.

As demonstrated in the simulation section, these methods can be applied with continuous covariate data fairly successfully by using categorical versions of these predictors. Most bias from a continuous covariate source is eliminated using only 2-3 covariate strata as seen in simulation. It has been our experience that most efficiency gain attainable from the continuous covariates is accomplished with categorically formed covariates with few strata. In point of fact, defining too many covariate strata for use with the test statistic can be detrimental since the upper limit of integration, τ, is constrained to the study times when all strata can produce consistent survival estimates. In defining covariate strata when sample size is limited, we recommend incorporating additional longitudinal covariate measurements rather than finely stratifying covariate information early on as a general rule. If very many covariate strata are required at multiple time-points with limited data, one may need to consider the appropriateness of making semi-parametric assumptions. Instead of WS estimates one could use Robins and Rotnitzky survival estimates with semiparametric inverse probability weights as described in section 2.1 in adjusting the PF statistic.

7. Acknowledgments

This work was supported by the National Cancer Institute and the National Institute of Allergy and Infectious Diseases. This work was also supported in part by grant CA39640 from the National Institutes of Health. The authors thank Andrea Rotnitzky for helpful conversations.

8. Appendix

Remarks on Asymptotic Normality of the Adjusted Test Statistic

A formal proof of asymptotic normality is very complicated in the general case with time dependent covariates. We give some flavor of the proof by looking at the case with baseline covariates and considering the one-sample problem. In such a case we estimate $S(t) =$

$\sum_{i=0}^{k} S_i(t)\theta_i$ using $WS(t) = \sum_{i=0}^{k} \hat{S}_i(t)\hat{\theta}_i$. The normalized estimator minus estimand can be written as

$$n^{1/2}\int_0^{\tau^*} \hat{w}(t)\{WS(t) - S(t)\}\,dt = \sum_{i=0}^{k}\int_0^{\tau^*} \hat{w}(t)n^{1/2}\{\hat{S}_i(t) - S_i(t)\}\hat{\theta}_i\,dt + \sum_{i=0}^{k} n^{1/2}(\hat{\theta}_i - \theta_i)\int_0^{\tau^*} \hat{w}(t)S_i(t)\,dt. \quad (5)$$

Using standard counting process representation for the Kaplan-Meier estimator, the first term on the right hand side of equation (4) is equivalent in distribution to

$$n^{-1/2}\left[\sum_{i=0}^{k}\sum_{j=1}^{n_i}\int_0^{\tau^*} \frac{A_i(t)dM_{ij}(t)}{S_i(t)H_i(t)}\right] + o_p(1),$$

where $A_i(t) = \int_t^{\tau^*} w(u)S_i(u)\,du$, and $o_p(1)$ corresponds to a term that converges in probability to zero. With respect to the usual filtration, $\mathcal{F}(t)$, which includes the covariates and all the failure and censoring information up to time t, the above statistic is a realization of a martingale process that by standard application of the martingale central limit theorem is asymptotically normal with mean zero. The second term on the right hand side of (4) is a linear combination of $n^{1/2}(\hat{\theta}_i - \theta_i)$, which is also asymptotically normal with mean zero, and another $o_p(1)$ term. Since the estimates $\hat{\theta}_i$ are functions of the covariates they are $\mathcal{F}(0)$ measurable and hence uncorrelated with the first term on the rhs of (4). Consequently, $n^{1/2}\int_0^{\tau^*} \hat{w}(t)\{WS(t) - S(t)\}\,dt$ is asymptotically normal with mean zero. This argument can be extended to the time-dependent covariate case using an induction argument together with appropriate conditioning arguments.

References

D. R. Cox, "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society, B* vol. 34 pp. 1038–1041, 1972.

D. M. Finkelstein and D. A. Schoenfeld, "Analyzing survival in the presence of an auxiliary variable," *Statistics in Medicine* vol. 13 pp. 1747–1754, 1994.

R. J. Gray, "A kernel method for incorporating information on disease progression in the analysis of survival," *Biometrika* vol. 81 pp. 527–539, 1994.

E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Am. Statist. Assoc.* vol. 53 pp. 457–81, 1958.

M. R. Kosorok and T. R. Fleming, "Using surrogate failure time data to increase cost effectiveness in clinical trials," *Biometrika* vol. 80 pp. 823–33, 1993.

H. M. Malani, "A modification of the redistribution to the right algorithm using disease markers," *Biometrika* vol. 82 pp. 515–526, 1995.

S. Murray, "Nonparametric estimation and testing for survival data in the two sample censored data problem incorporating longitudinal covariates," Sc.D. dissertation, Department of Biostatistics, Harvard University, 1994.

S. Murray and A. A. Tsiatis, "A nonparametric approach to incorporating prognostic longitudinal covariate information in survival estimation," *Biometrics* vol. 52 pp. 137–151, 1996.

M. S. Pepe and T. R. Fleming, "Weighted Kaplan-Meier statistics-a class of distance tests for censored survival data," *Biometrics* vol. 45 pp. 497–507, 1989.

M. S. Pepe and T. R. Fleming, "Weighted Kaplan-Meier statistics: large sample theory and optimality considerations," *J. R. Statist. Soc. B* vol. 53 pp. 341–52, 1991.

J. M. Robins and A. Rotnitzky, "Recovery of information and adjustment for dependent censoring using surrogate markers," in *AIDS Epidemiology: Methodological Issues*, (N. Jewell, K. Dietz and V. Farewell, eds.), Boston: Birkhäuser-Boston, 1992.