Harvard University

Harvard University Biostatistics Working Paper Series

Year 2009   *Paper* 99

A Class of Semiparametric Mixture Cure Survival Models with Dependent Censoring

Megan Othus[*]   Yi Li[†]

Ram C. Tiwari[‡]

[*]Harvard University and Dana Farber Cancer Institute, mothus@fas.harvard.edu

[†]Harvard University and Dana Farber Cancer Institute, yili@jimmy.harvard.edu

[‡]Federal Drug Administration, ram.tiwari@fda.hhs.gov

A Class of Semiparametric Mixture Cure Survival Models with Dependent Censoring

Megan Othus

Harvard University and Dana Farber Cancer Institute

Boston, MA, 02115

Yi Li

Harvard University and Dana Farber Cancer Institute

Boston, MA, 02115

Ram C. Tiwari

Federal Drug Adminstration

Silver Springs, MD, 20993

Abstract

Modern cancer treatments have substantially improved cure rates and have generated a great interest in and need for proper statistical tools to analyze survival data with non-negligible cure fractions. Data with cure fractions are often complicated by dependent censoring, and the analysis of this type of data typically involves untestable parametric assumptions on the dependence of the censoring mechanism and the true survival times. Motivated by the analysis of prostate cancer survival trends, we propose a class of semiparametric transformation cure models that allows for dependent censoring without making parametric assumptions on the dependence relationship. The proposed class of models encompasses a number of common models for the latency survival function, including the proportional hazards model and the proportional odds model, and also allows for time-dependent covariates. An inverse censoring probability reweighting scheme is used to derive unbiased estimating equations. We validate small sample properties with simulations and demonstrate the method with a data application.

1. INTRODUCTION

Survival models incorporating a cure fraction are an important statistical tool for analyzing cancer survival data. Research for many cancer types, including prostate and breast cancer, non-Hodgkin's lymphoma, leukemia, melanoma, and head and neck cancer, has shown that a significant proportion of patients with these cancers are cured after therapy; that is, an individual will have little or no risk of experiencing the event of interest, e.g. breast cancer, after treatment. Cure survival models are

1

often developed assuming the study population is a mixture of cured and not cured subjects (Berkson and Gage 1952; Farewell 1982; Kuk and Chen 1992; Peng and Dear 2000; Sy and Taylor 2000). Recently, Lu and Ying (2004) used the mixture formulation to extend a class of semiparametric transformation models proposed by Cheng, Wie, and Ying (1995) to incorporate cure fractions. An alternative to the mixture cure formulation is the promotion cure model, which is popular for Bayesian cure survival analysis and has an attractive biological interpretation (Chen, Ibrahim, and Sinha 1999; Tsodikov, Ibrahim, and Yakovlev 2003). The promotion cure model has one covariate function that describes both the survival and cure components. In contrast, the two components of the mixture cure model can depend on different parameters, allowing for separate covariate interpretations for the cure function and for the survival function for those who are not cured.

A key assumption made in the literature is independence between the survival time and the censoring time, but such an assumption is often violated in observational studies. For example, in our motivating study a non-negligible number of patients diagnosed with prostate cancer were found to have died from cardiovascular disease. Research has shown that prostate cancer and heart disease share high fat diet as a common risk factor (Wilson et al. 1998; Strom et al. 2008). Therefore, assuming independence between the main end point (death from prostate cancer) and censoring causes (e.g. death from heart disease) is problematic. Dependent censoring will often hamper statistical analysis since classical models will not be valid, and ignoring dependent censoring will typically introduce bias (Li, Tiwari, and Guha 2007). To account for this, Li et al. (2007) developed a cure model in the presence of dependent censoring, but their model requires *a priori* assumptions on the dependence structure and does not allow for covariates.

In this project, we propose a semiparametric transformation model which allows

2

for covariates as well as dependent censoring. The key idea is to use an inverse censoring probability reweighting scheme to derive unbiased estimating equations that account for dependent censoring. In this manner, we are able to avoid making parametric assumptions about the dependence structure between the survival time and the censoring time. It is also worth noting that our proposed model, which accommodates time-dependent covariates, is more general than that proposed by Lu and Ying (2004), which only allows for time-independent covariates.

Broadly speaking, this project advances the field in three distinct ways. First, the proposed methods can be used to investigate trends in Surveillance Epidemiology and End Results (SEER) cancer survival data (www.seer.cancer.gov); a number of cancer sites in the SEER database may indicate cure patterns, including prostate, colon, and ovarian cancer (Tai et al. 2005). Second, survival studies with cure fractions are subject to heavy censoring that is likely to violate the traditional independent censoring assumption. This project addresses the problem by developing methods that can accommodate dependent censoring, and by providing procedures for model diagnostics and techniques for statistical inference. The proposed estimation procedure successfully avoids parametric assumptions on the joint distribution of the censoring and true failure times, which have caused difficulties in the study of dependent censoring in the existing literature. Finally, the theoretical proofs developed in this project provide an original application of empirical process theory and functional analysis techniques, which will be valuable for semiparametric estimation.

The rest of the paper is stuctured as follows: in Section 2 we define notation and set up the survival model. In Section 3 we define the weight function. Section 4 uses the results of Sections 2 and 3 to define estimating equations and provides theoretical results. Section 5 summarizes a computational algorithm and a weighted bootstrap method for inference procedures on the estimating equations. Section 6 provides

3

a model checking procedure. In Section 7 we present simulation results and data analysis findings. Section 8 contains a short discussion. Sketches of proofs of theorems are contained in the Appendix.

2.   A CLASS OF SEMIPARAMETRIC MODELS

Denote the failure time for a patient by T, which is assumed to have the decomposition $T = \eta T^* + (1 - \eta)\infty$, where $\eta \in \{0, 1\}$ indicates whether the sampled subject is cured ($\eta = 0$) or not ($\eta = 1$), and T^*, with support on $[0, \infty)$, denotes the failure time if the patient is not cured. Let C denote the censoring time, $\mathbf{X}$ a κ length vector of covariates related to the cure indicator η, which includes 1, and $\mathbf{Z}(\cdot)$ a ξ length vector of external time-dependent covariates related to T^* that may share time-independent components with $\mathbf{X}$. Define $\tilde{T} = \min(T, C)$, $\Delta = I(T \leq C)$, a counting process related to the survival progress $N(t) = I(\tilde{T} \leq t, \Delta = 1)$, and the at-risk process $Y(t) = I(\tilde{T} \geq t)$. In practice, we observe $(\tilde{T}_i, \Delta_i, \mathbf{Z}_i(\cdot), \mathbf{X}_i)$, for $i = 1, \ldots, n$, which are independent and identically distributed copies of $(\tilde{T}, \Delta, \mathbf{Z}(\cdot), \mathbf{X})$. To ensure that we have unbiased estimates for our parameters we assume that the support of the censoring time is not shorter than the support of the failure time (Fine, Ying, and Wei 1998; Fine 1999).

Using the above notation, we can define a model for T. Following Farewell (1982), we use a logistic model for the cure indicator η:

$$P(\eta = 1 | \mathbf{X}, \mathbf{Z}) = G(\boldsymbol{\gamma}' \mathbf{X}), \tag{1}$$

where $G(x) = \exp(x)/(1 + \exp(x))$. Among patients who are not cured, we assume that given the covariate path $\bar{\mathbf{Z}}(t) = \{\mathbf{Z}(u), u \leq t\}$, T^* possesses the following survival

4
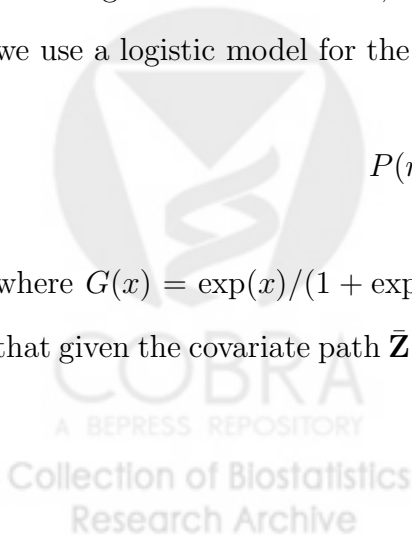
function:

$$S_g(t; \bar{\mathbf{Z}}(t)) = g\Big\{ \int_0^t \exp(\boldsymbol{\beta}'\mathbf{Z}(u)) \, dH(u) \Big\}, \tag{2}$$

where H is a fixed but unspecified nondecreasing function with $H(0) = 0$, and g is a known continuous and strictly decreasing function such that $g(0) = 1$ and $\lim_{t\to\infty} g(t) = 0$. This model has been considered in Kosorok and Song (2007) and Zeng and Lin (2007). Zeng and Lin (2007) suggested extending their model to cure applications in future research; in contrast to our proposal, their work assumed conditional independence between T and C. The survival function for the noncured subjects, $S_g(\cdot)$, is commonly referred to as the *latency survival function*. The formulation of (2) is comprehensive: when $g(x) = 1/(1+x)$ the resulting model is a time-dependent construction of the proportional odds model (Bennett 1983), and when $g(x) = \exp(-x)$ the model reduces to the time-dependent proportional hazards model (Cox 1972).

Combining (1) and (2), we can describe the survival function for T as

$$P(T \geq t | \bar{\mathbf{Z}}(t), \mathbf{X}) = g\Big\{ \int_0^t \exp(\boldsymbol{\beta}'\mathbf{Z}(u)) \, dH(u) \Big\} G(\boldsymbol{\gamma}'\mathbf{X}) + \bar{G}(\boldsymbol{\gamma}'\mathbf{X}), \tag{3}$$

where $\bar{G}(x) = 1 - G(x)$. Let $\mathbf{W}(\cdot)$ denote the external time-dependent covariate process associated with censoring and let $\bar{\mathbf{W}}(t) = \{\mathbf{W}(u), u \leq t\}$ denote the covariate path associated with $\mathbf{W}(\cdot)$. We assume that $P(T \geq t | \bar{\mathbf{Z}}(t), \mathbf{X}) = P(T \geq t | \bar{\mathbf{Z}}(t), \mathbf{X}, \bar{\mathbf{W}}(t))$, i.e. that $\bar{\mathbf{W}}(t)$ does not contain any extra information for T beyond what is already in $\bar{\mathbf{Z}}(t)$ and $\mathbf{X}$.

3.   A REWEIGHTING SCHEME FOR DEPENDENT CENSORING

We wish to make minimal assumptions on the dependence between T and C, and so it will not be feasible to construct likelihood-based inference procedures. Instead, we

5

will derive unbiased estimating equations for $\boldsymbol{\beta}$, the parameter associated with the latency survival function, and $\boldsymbol{\gamma}$, the parameter associated with the cure function. To this end, we first define some notation related to the censoring process. Define the crude hazard for censoring as

$$\lambda_c(t|\bar{\mathbf{W}}(t)) = \lim_{dt \to 0} (dt)^{-1} P(C \in (t, t+dt), \Delta = 0 | \tilde{T} \geq t, \bar{\mathbf{W}}(t)) \qquad (4)$$

which can also be regarded as a cause specific hazard. Conveniently, (4) can be estimated with observed data even when T and C are dependent (Fleming and Harrington 1991, Theorem 1.3.1).

We make the following assumption on the crude hazard for censoring:

$$\lambda_c(t|\bar{\mathbf{W}}(t), T, T > t) = \lambda_c(t|\bar{\mathbf{W}}(t), T > t).$$

This assumption implies that, conditional on the covariate process $\bar{\mathbf{W}}(t)$, the crude hazard for censoring at time t does not further depend on the possibly unobserved failure time. This assumption has been described as "no unmeasured confounders for censoring" and the assumption would fail if a covariate related to both T and C were not included in $\bar{\mathbf{W}}(t)$ (Robins and Finkelstein 2000). This assumption has been well discussed for semiparametric survival problems in the literature with examples and information on how to select covariates for $\bar{\mathbf{W}}(t)$ (Robins 1992; Robins and Rotnitzky 1992; Robins 1993; Rotnitzky and Robins 1995; Robins and Finkelstein 2000; Rotnitzky and Robins 2003; Miloslavsky, Keles, Laan, and Butler 2004).

Flexible models on $\lambda_c(t|\bar{\mathbf{W}}(t))$ can be considered. For example, if $\boldsymbol{\alpha}_E$ is the

6

parameter vector associated with $\mathbf{W}(\cdot)$, we could write the crude hazard function as

$$\lambda_c(t|\bar{\mathbf{W}}(t)) = \lambda_0(t)\exp(\boldsymbol{\alpha}'_E\mathbf{W}(t)), \tag{5}$$

where $\lambda_0(t)$ is left unspecified.

Let $\Lambda_c(t|\bar{\mathbf{W}}(t)) = \int_0^t \lambda_c(u|\bar{\mathbf{W}}(u))\,du$ and $K(t|\bar{\mathbf{W}}(t)) = \prod_{u\leq t}(1 - d\Lambda_c(u|\bar{\mathbf{W}}(u)))$. Throughout we will write $K(t|\bar{\mathbf{W}}(t)) = K(t)$. Note that if all the components of $\mathbf{W}(\cdot)$ are time-independent and T and C are independent, then $K(t) = P(C > t|\mathbf{W})$ and $K(t)$ can be interpreted as the probability that a person with covariate $\mathbf{W}$ is not censored up to time t. This interpretation does *not* hold in general as the crude hazard $\lambda_c(t)$ is not equal to the net hazard of censoring when T and C are dependent.

Similar arguments as in Robins (1993) lead to

$$E[I(\tilde{T} \in (t, t+dt), \Delta = 1)/K(\tilde{T}-)] = P(T \in (t, t+dt)) \tag{6}$$

$$\text{and} \qquad E[I(\tilde{T} \geq t)/K(t-)] = P(T \geq t). \tag{7}$$

These equations imply that the reweighted sample of patients observed to fail at time t is representative of all patients who fail at time t and that the reweighted sample of patients who have not failed by time t is representative of the general survival probability of all patients at time t.

4. ESTIMATING EQUATIONS AND THEORETICAL RESULTS

Equations (6) and (7) are important as they hold even when T and C are dependent, and we can use them to derive estimating equations for the unknown regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, and the function H. Specifically, (6) implies that $E[dN(t)/K(t-)+ dP(T \geq t|\bar{\mathbf{Z}}(t), \mathbf{X})] = 0$, or, equivalently, $E_{\mathbf{W}}[E[dN(t)+K(t-)\,dP(T \geq t|\bar{\mathbf{Z}}(t), \mathbf{X})|\mathbf{W}]] =$

7

0; while (7) implies that $E_{\mathbf{W}}[E[Y(t) - K(t-)P(T \geq t | \bar{\mathbf{Z}}(t), \mathbf{X}) | \mathbf{W}]] = 0$. Together these constitute a basis for estimation. Here, $P(T \geq t | \bar{\mathbf{Z}}(t), \mathbf{X})$ is specified in Equation (3) and $dP(\cdot)$ is the derivative of Equation (3) with respect to t.

Let $\boldsymbol{\alpha}$ denote the parameters associated with $K(\cdot)$; define $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, H)$ and write the true value of $\boldsymbol{\theta}$ as $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, H_0)$. Denote the estimating equations associated with $\boldsymbol{\alpha}$ as $\mathbf{U}_{n\boldsymbol{\alpha}}(\boldsymbol{\alpha})$. In the following, the subscript i marks individuals. Define

$$
\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{\beta},i}(\boldsymbol{\theta}) &= \int_0^\nu \mathbf{Z}_i(u)[dN_i(u) + K_i(u-)G(\boldsymbol{\gamma}'\mathbf{X}_i)\, dS_g(u; \bar{\mathbf{Z}}_i(u))] \\
\boldsymbol{\mu}_{\boldsymbol{\gamma},i}(\boldsymbol{\theta}) &= \int_0^\nu \mathbf{X}_i[Y_i(u) - K_i(u-)[G(\boldsymbol{\gamma}'\mathbf{X}_i)S_g(u; \bar{\mathbf{Z}}_i(u)) + \bar{G}(\boldsymbol{\gamma}'\mathbf{X}_i)]]\, du \\
\mu_{H,i}(\boldsymbol{\theta})(t) &= \int_0^t dN_i(u) + K_i(u-)G(\boldsymbol{\gamma}'\mathbf{X}_i)\, dS_g(u; \bar{\mathbf{Z}}_i(u)).
\end{aligned}
$$

A natural set of estimating equations for $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and H using (3) and our results from (6), and (7) is:

$$
\mathbf{U}_{n\boldsymbol{\beta}}(\boldsymbol{\theta}) \stackrel{def}{=} \sum_{i=1}^n \boldsymbol{\mu}_{\boldsymbol{\beta},i}(\boldsymbol{\theta}) \tag{8}
$$

$$
\mathbf{U}_{n\boldsymbol{\gamma}}(\boldsymbol{\theta}) \stackrel{def}{=} \sum_{i=1}^n \boldsymbol{\mu}_{\boldsymbol{\gamma},i}(\boldsymbol{\theta}) \tag{9}
$$

$$
U_{nH}(\boldsymbol{\theta})(t) \stackrel{def}{=} \sum_{i=1}^n \mu_{H,i}(\boldsymbol{\theta})(t), \ \forall\, t \in [0, \nu]. \tag{10}
$$

In the above, choose ν to be a constant such that $P(\tilde{T} > \nu) > \check{c} > 0$ for some fixed $\check{c}$. The constraint on ν is imposed to avoid tail instability associated with large values of t. In practice the weight function $K(\cdot)$ is unknown and can be replaced with an empirical estimator. This is easily done by replacing $\Lambda_c(u | \bar{\mathbf{W}}(u))$ in the definition of $K(t)$ with a consistent estimate $\hat{\Lambda}_c(u | \bar{\mathbf{W}}(u))$. Since the left term of U_{nH}

8

only jumps at observed failure times, the estimate for H will be a step-function that jumps at the observed failure times. Let $\mathbf{U}_n(\boldsymbol{\theta}) = (\mathbf{U}_{n\boldsymbol{\alpha}}, \mathbf{U}_{n\boldsymbol{\beta}}, \mathbf{U}_{n\boldsymbol{\gamma}}, U_{nH})(\boldsymbol{\theta})$ and define $\mathbf{U}(\boldsymbol{\theta}) = E\mathbf{U}_n(\boldsymbol{\theta})$.

To make the definition of $\mathbf{U}_{n\boldsymbol{\alpha}}$ more concrete, we provide an example where the crude hazard for censoring (Equation (4)), follows a proportional hazards model (Equation (5)). In this case we can decompose $\boldsymbol{\alpha}$ into two parts $(\boldsymbol{\alpha}_E, \boldsymbol{\alpha}_\Lambda)$ where $\boldsymbol{\alpha}_E$ represents the Euclidean parameter of the model and $\boldsymbol{\alpha}_\Lambda$ denotes the baseline cumulative hazard function for the model. The estimating equations associated with $(\boldsymbol{\alpha}_E, \boldsymbol{\alpha}_\Lambda)$ are the usual partial likelihood score equations and the Breslow formula for the cumulative hazard:

$$
\begin{aligned}
\mathbf{U}_{n\boldsymbol{\alpha}_E}(\boldsymbol{\alpha}) &= \sum_{i=1}^n \int_0^\nu \mathbf{W}_i(u)(dN_{ci}(u) - Y_i(u)\exp(\boldsymbol{\alpha}'_E \mathbf{W}_i(u))\, d\Lambda_0(u)) \\
U_{n\boldsymbol{\alpha}_\Lambda}(\boldsymbol{\alpha})(t) &= \sum_{i=1}^n \int_0^t dN_{ci}(u) - Y_i(u)\exp(\boldsymbol{\alpha}'_E \mathbf{W}_i(u))\, d\Lambda_0(u), \quad \forall\, t \in [0, \nu]
\end{aligned}
$$

where $N_{ci}(t) = I(\tilde{T}_i \leq t, \Delta_i = 0)$ for $i = 1, \ldots, n$, and $\Lambda_0(\cdot)$ is the baseline cumulative hazard function. For this proportional hazards case we would write $\mathbf{U}_{n\boldsymbol{\alpha}} = (\mathbf{U}_{n\boldsymbol{\alpha}_E}, U_{n\boldsymbol{\alpha}_\Lambda})'$.

Define $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{H})$ as zeroes of $\mathbf{U}_n(\boldsymbol{\theta}) = (\mathbf{U}_{n\boldsymbol{\alpha}}, \mathbf{U}_{n\boldsymbol{\beta}}, \mathbf{U}_{n\boldsymbol{\gamma}}, U_{nH})(\boldsymbol{\theta})$. In general one is interested in the zeroes of (8) and (9), which we denote $\hat{\boldsymbol{\phi}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$. We note that the asymptotic variance formula for $n^{1/2}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi})$ is analytically so complicated that it will be of limited computational utility. Instead, we propose an application of a weighted bootstrap, which is introduced in Section 5 and is shown to yield a consistent estimate for the variance. We will write the weighted bootstrap version of $\mathbf{U}_n(\boldsymbol{\theta})$ by $\mathbf{U}_n(\boldsymbol{\theta})^*$ and denote the weighted bootstrap estimate of $\boldsymbol{\theta}$ from $\mathbf{U}_n(\boldsymbol{\theta})^*$ as $\hat{\boldsymbol{\theta}}^*$.

9

Conditions for and sketches of proofs for the following asymptotic results are contained in the Appendix.

Theorem 1. *Under regularity conditions (c.1)–(c.4), the survival model in Equation (3) is identifiable.*

This result is more general than that for cure survival models provided by Li, Taylor, and Sy (2001) since time-dependent cure proportional odds and proportional hazards models are included in our result as special cases. Further, in contrast to Li et al. (2001), we do not require the value zero to be in the support of the covariates, which may not be a realistic assumption in many situations. As pointed out by a reviewer, a key component of the identifiability result is the presence of covariates in both components of the mixture model. In fact, Li et al. (2001) show that if the latency survival function is nonparameteric and independent of covariates and if the cure fraction is taken to be a constant (and so does not depend on covariates) the mixture cure survival model is not identifiable. The dependence of our result on covariates is analogous to identifiability results in the competing risks literature. A number of authors have put forward identifiable competing risks models, and all of the results depend on the presence of covariates in the models (Heckman and Honore 1989; Abbring and van den Berg 2003; Lee 2006).

Theorem 2. *Under regularity conditions (c.1)–(c.7), $\hat{\boldsymbol{\theta}}$ is a consistent estimator for $\boldsymbol{\theta}_0$.*

Theorem 3. *Under regularity conditions (c.1) – (c.7) and (b.1)–(b.5), $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges weakly to a zero-mean Gaussian process. Conditional on the observed data, $n^{1/2}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})$ has the same limiting distribution as $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$.*

10

5. Inference Procedures and Weighted Bootstrap Scheme

5.1. Algorithm

With the large sample theory established, we propose an iterative algorithm to solve estimating equations (8)–(10). To start, let $t_0 < t_1 < \ldots < t_k$ be the ordered observed failure times such that $t_k \leq \nu$ and $t_0 = 0$. In practice, ν can be set equal to the 90th or 95th percentile of the observed failure times.

Step 1: Choose initial values for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and H, denoted $\hat{\boldsymbol{\beta}}^{(0)}$, $\hat{\boldsymbol{\gamma}}^{(0)}$, $\hat{H}^{(0)}$, respectively. In general, the initial values for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ can be obtained from estimates that ignore dependent censoring or potential cure fractions and one can let $\hat{H}^{(0)}$ be the constant value 0.

Step 2: Compute estimates for $K(\cdot)$. For example, if the crude censoring distribution follows a time-dependent proportional hazards model, the calculation can be done following Robins (1993).

Suppose that we have estimates for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and H from the $(m-1)^{th}$ iteration; denote these estimates by $\hat{\boldsymbol{\beta}}^{(m-1)}$, $\hat{\boldsymbol{\gamma}}^{(m-1)}$, and $\hat{H}^{(m-1)}$.

Step 3: Recall that $\hat{H}(\cdot)$ is a step function with jumps at t_j, $j = 1, \ldots, k$. Denote the jump size at t_j as $\hat{H}\{t_j\}$. For $j = 1, \ldots, k$ obtain estimates $\hat{H}^{(m)}\{t_j\}$, in order, as solutions to $U_{nH}(t_j) = 0$ with $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ set equal to $\hat{\boldsymbol{\beta}}^{(m-1)}$ and $\hat{\boldsymbol{\gamma}}^{(m-1)}$, respectively. As $\hat{H}^{(m)}(t)$ is a step function, $\int_0^{t_j} e^{\mathbf{Z}_i(u)'\hat{\boldsymbol{\beta}}^{(m-1)}} d\hat{H}^{(m)}(u) = \sum_{l=1}^{j} e^{\mathbf{Z}_i(t_l)'\hat{\boldsymbol{\beta}}^{(m-1)}} \hat{H}^{(m)}\{t_l\}$. Hence $\hat{H}^{(m)}\{t_j\}$ can be obtained recursively given the estimates of $\hat{H}^{(m)}\{t_l\}$, $l \leq j - 1$. A unique solution is guaranteed to exist due to the monotonicity of g.

11

Step 4: To estimate $\hat{\boldsymbol{\beta}}^{(m)}$, solve for $\boldsymbol{\beta}$ such that $\mathbf{U}_{n\boldsymbol{\beta}} = 0$ with $\boldsymbol{\gamma}$ set equal to $\hat{\boldsymbol{\gamma}}^{(m-1)}$ and H set equal to $\hat{H}^{(m)}$. The proof of Theorem 2 establishes that the derivative of $\mathbf{U}_{n\boldsymbol{\beta}}$ is negative definite, therefore the Newton-Raphson solution will be unique.

Step 5: To find $\hat{\boldsymbol{\gamma}}^{(m)}$, solve for $\boldsymbol{\gamma}$ such that $\mathbf{U}_{n\boldsymbol{\gamma}} = 0$, with $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(m)}$ and $H = \hat{H}^{(m)}$. As with the solution $\hat{\boldsymbol{\beta}}^{(m)}$, the derivative of $\mathbf{U}_{n\boldsymbol{\gamma}}$ is negative definite (see the proof of Theorem 2), so the Newton-Raphson method can be used to find a unique solution.

Step 6: Repeat *Steps 3 - 5* until predetermined convergence criteria are met.

5.2. Weighted Bootstrap

To conduct inference on the parameter estimates we use a weighted bootstrap (Wellner and Zhan 1996), which is applicable even in this case with an infinite dimensional nuisance parameter. Define $\boldsymbol{\mu}_{\boldsymbol{\alpha} i}$ to be such that $\sum_{i=1}^{n} \boldsymbol{\mu}_{\boldsymbol{\alpha} i}$ are the estimating equations for $\boldsymbol{\alpha}$. Recall that the estimates for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, H)$ are approximate zeros of $\mathbf{U}_{nj} = \sum_{i=1}^{n} \boldsymbol{\mu}_{ji}$ for $j = \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, H$. Choose a positive distribution B and generate $(b_{n1}, \ldots, b_{nn})$ such that B and $(b_{n1}, \ldots, b_{nn})$ meet conditions (b.1)–(b.5) in the Appendix. Denote the zeroes of $\sum_{i=1}^{n} \boldsymbol{\mu}_{ji} b_{ni}$, $j = \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, H$ as $\hat{\boldsymbol{\theta}}^{*} = (\hat{\boldsymbol{\alpha}}^{*}, \hat{\boldsymbol{\beta}}^{*}, \hat{\boldsymbol{\gamma}}^{*}, \hat{H}^{*})$. As stated in Theorem 3, the conditional distribution of $n^{1/2}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^{*})$ is asymptotically equivalent to the unconditional distribution of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. We can approximate the distribution of $\hat{\boldsymbol{\theta}}^{*}$ by generating a large number of samples, $(b_{n1}, \ldots, b_{nn})$, from B and for each sample calculating a realization of $\hat{\boldsymbol{\theta}}^{*}$. The empirical distribution can be used to compute statistics for inference.

6. MODEL DIAGNOSTICS

In practice, it is of substantial interest to verify whether the functional forms imposed on the regression models are plausible. To this end we propose a two-step

12
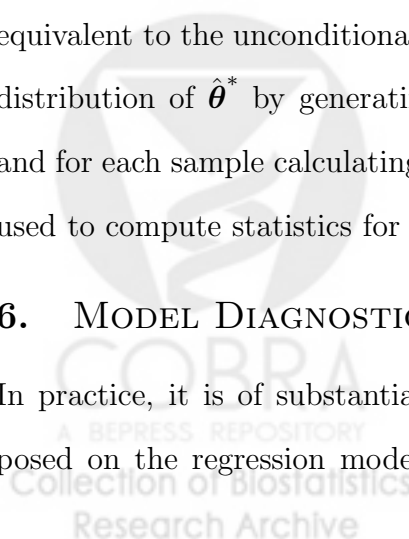
model checking procedure. First, assumptions on the form of K need to be verified. If the assumed model for K is correct, the asymptotic distribution of $\Xi(t) = n^{-1/2} \sum_{i=1}^n \tilde{\mathbf{q}}(\mathbf{W}_i, t)' \boldsymbol{\mu}_{\boldsymbol{\alpha}i}(\hat{\boldsymbol{\alpha}})$ will be a zero-mean Gaussian process, where $\tilde{\mathbf{q}}(\cdot, \cdot)$ is a weight function that is assumed to be known and bounded. Furthermore, if we let $(b_{n1}, \ldots, b_{nn})$ and $\hat{\boldsymbol{\alpha}}^*$ be defined as in the previous section, we can approximate the distribution of $\Xi(t)$ with $\tilde{\Xi}(t) = n^{-1/2} \sum_{i=1}^n \tilde{\mathbf{q}}(\mathbf{W}_i, t)'(\boldsymbol{\mu}_{\boldsymbol{\alpha}i}(\hat{\boldsymbol{\alpha}})(1 - b_{ni}) - \boldsymbol{\mu}_{\boldsymbol{\alpha}i}(\hat{\boldsymbol{\alpha}}) + \boldsymbol{\mu}_{\boldsymbol{\alpha}i}(\hat{\boldsymbol{\alpha}}^*))$. That the distribution $\tilde{\Xi}(t)$ approximates $\Xi(t)$ follows from the argument in Appendix C of Peng and Huang (2008).

A formal lack-of-fit test can be based upon the statistic $\mathcal{T}_K = ||\Xi(t)||_\infty$, where $|| \cdot ||_\infty$ denotes the supremum norm on $[0, \nu]$. We can generate a large number of sample paths from $\tilde{\Xi}(\cdot)$ by repeatedly sampling new $(b_{n1}, \ldots, b_{nn})$. A p-value for $\mathcal{T}_K$ can be estimated by the empirical proportion of $\mathcal{T}_k \leq ||\tilde{\Xi}(t)||_\infty$. Similar arguments as in Lin, Wei, and Ying (1993) show that this test is consistent against the alternative that the model specification of K is false. A corresponding visual diagnostic method is found by plotting a number of realizations of $\tilde{\Xi}(\cdot)$ and observing how extreme the pattern of $\Xi(\cdot)$ looks in comparison to the simulated paths.

Once we are satisfied with the fit of K we can check the form of (3). Define $M_i(t, \boldsymbol{\theta}) = Y_i(t)/K_i(t-) - G(\boldsymbol{\gamma}'\mathbf{X}_i)S_g(t, \bar{\mathbf{Z}}(t)) - \bar{G}(\boldsymbol{\gamma}'\mathbf{X}_i)$. Consider the function $\boldsymbol{\Gamma}(t) = n^{-1/2} \sum_{i=1}^n w(\mathbf{Z}_i, \mathbf{X}_i, t)M_i(t, \hat{\boldsymbol{\theta}})$, where the weight function $w(\cdot, \cdot, \cdot)$ is assumed to be known and bounded. If (3) is correctly specified and $K(\cdot)$ is consistently estimated, $\boldsymbol{\Gamma}(t)$ converges to a zero-mean Gaussian process. We can approximate the distribution of $\boldsymbol{\Gamma}(t)$ with the distribution of

$$\tilde{\boldsymbol{\Gamma}}(t) = n^{-1/2} \sum_{i=1}^n w(\mathbf{Z}_i, \mathbf{X}_i, t)(M_i(t, \hat{\boldsymbol{\theta}})(1 - b_{ni}) - M_i(t, \hat{\boldsymbol{\theta}}) + M_i(t, \hat{\boldsymbol{\theta}}^*)), \quad (11)$$

where b_{ni} and $\hat{\boldsymbol{\theta}}^*$ are defined as in the previous section. If $M(\cdot, \boldsymbol{\theta}_0)$ is specified

13

correctly, $\boldsymbol{\Gamma}(t)$ should have mean zero for any t. Following the arguments for Ξ we can establish that the distribution of $\tilde{\boldsymbol{\Gamma}}$ is asymptotically the same as the distribution of $\boldsymbol{\Gamma}$.

As in the model-checking procedure for $K(\cdot)$, a formal test for lack-of-fit can be made based upon $\mathcal{T}_S = \|\boldsymbol{\Gamma}(t)\|_\infty$. As with the diagnostic method for $K(\cdot)$, we can calculate a p-value for $\mathcal{T}_S$ (the corresponding test can be shown to be consistent against the general alternative that (3) is not correctly specified) and develop a goodness-of-fit plot.

7. NUMERICAL STUDIES

7.1. Simulations

We evaluated the finite sample performance of the proposed methods with simulations. Two such studies are summarized in Table 1. We considered proportional odds and proportional hazards models. Dependent censoring can occur when a covariate that affects both the survival and censoring times is not included in the model. To mimic this situation we generated two covariates for each subject: a Bernoulli(0.5) random variable and a Uniform(0,1) random variable. The censoring times followed a proportional hazards model with both the Bernoulli and Uniform random variables as covariates. The latency survival times followed either the proportional hazards or proportional odds models with both the Bernoulli and Uniform random variables as covariates. The cure functions only depended on the Bernoulli random variable. For the proportional hazards model we took $H(t) = t$ and for the proportional odds model we took $H(t) = 2t$.

In all of the analyses we only included one covariate in the cure function and in the latency survival function, the Bernoulli random variable. To model the crude hazard for censoring we used a proportional hazards model with the Bernoulli random

14
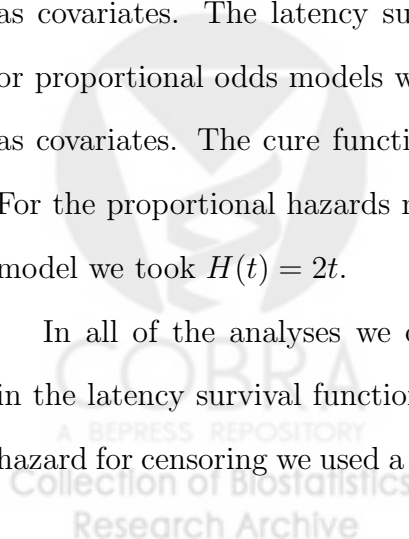
variable as the only covariate. Bias, monte carlo standard error estimates, bootstrap based standard error estimates, and estimates of the root mean squared error (RMSE) for several different parameter configurations are presented in Table 1. We found the "true" values of the misspecified latency survival parameters by generating a very large dataset (sample size = 1,000,000) with the two covariates, but fitting the model on the dataset with just one covariate.

To investigate the biases that occur when dependent censoring is ignored we compared our method to the method of Lu and Ying (2004) (we will refer to this method as LY hereafter), which does not compensate for dependent censoring. Summaries for bias, monte carlo standard errors, and estimates of RMSE can also be found in Table 1. The proposed method has the most significant gains over LY in estimating the parameter associated with the cure proportion. The proposed method and LY have comparable bias in estimating the latency survival parameter. In practice, if correct estimates of the cure proportion are of primary interest, the proposed method can reduce bias even when accounting for only some of the dependent censoring.

7.2. SEER Prostate Cancer Data

We applied the proposed method to prostate cancer data from the SEER database, using data released in 2004. Specifically, we looked at males diagnosed with prostate cancer from the Detroit, Michigan area whose cancer stage was classified as local or regional. The Detroit registry was one of the original nine SEER registries and so has longer follow-up than the more recent registries. To avoid potential confounding related to prior cancer diagnoses, we restricted our sample to those men whose prostate cancer was their first cancer diagnosis. The Federal Drug Administration (FDA) approved the prostate-specific antigen (PSA) test for use in monitoring patients diagnosed with prostate cancer in 1986. The use of the PSA test for screening

15

of prostate cancer increased dramatically starting in 1990, and so to avoid potential biases associated with the new screening method we only considered cases diagnosed up to 1989. Our random variable of interest was time from diagnosis of prostate cancer to death from prostate cancer.

One question of interest was whether survival or cure fractions differed in this dataset by race. Due to the small number of subjects in some racial categories in this dataset we restricted our attention to the two largest racial groups, black and white. There were 15,524 cases in the SEER database that met the criteria; 3,837 were classified as black and 11,687 were classified as white.

In this cohort, 3,717 people died of prostate cancer while 4,349 people died of heart disease. These numbers are summarized in Table 2. Heart disease and prostate cancer share the common risk factor of high fat diet (Wilson et al. 1998; Strom et al. 2008). High fat diets have been shown to increase androgen levels in the body, increase cholesterol levels, and are associated with increased consumption of Omega-6 fatty acids. Increased levels of androgens have been shown to be associated with prostate cancer cell proliferation (Fleshner et al. 2004; Pandini et al. 2005; Xu et al. 2006; Asirvatham et al. 2006). High cholesterol and Omega-6 fatty acids have been shown to be associated with prostate carcinogenesis (Homma et al. 2004; Hughes-Fulford et al. 2005; Zhuang et al. 2005; Bravi et al. 2006; Ritch et al. 2007). Further complicating the relationship between heart disease and prostate cancer is the fact that many men diagnosed with prostate cancer are treated with androgen therapy to reduce hormone levels. This treatment has been shown to be associated with an increased risk of death from heart disease (Keating et al. 2006). Given these connections and the high rate of death from heart disease in this population, there is some evidence that there may exist dependent censoring in this dataset.

Figure 1 shows the naïve (assuming independent censoring) Kaplan-Meier survival

16

plots for the two subgroups. We can observe that both curves plateau above 0.5, indicating that a large proportion of the population may not die of prostate cancer. While for the first several years the survival curves are indistinguishable, the white subgroup's survival tail is noticeably higher than the tail of the black group.

There are several situations when physicians consider patients to be cured of prostate cancer. First, surgery (radical prostatectomy) and radiotherapy (external-beam radiotherapy, brachytherapy, or both) in early stage prostate cancer patients can entirely eradicate the disease, inducing a cure among a significant proportion of patients (Hanlon and Hanks 2000; Jani and Hellman 2003). Also, improvements in prostate cancer treatment have made it so that many prostate cancer patients who are not cured are treated for chronic disease rather than acute disease, and their probability of dying of prostate cancer is not elevated above that of the general population (American Cancer Society 2007). Given these considerations, a model that explicitly accounts for a cure proportion and allows for inference on the cure proportion can be of use for this prostate cancer application.

We considered both proportional hazards and proportional odds models. Race was a covariate for the survival and cure functions, denoted R_S and R_C, respectively. The models also contain a time-dependent covariate written PSA_S, which takes the value one after the year 1985 and the value zero otherwise. This covariate can help to account for changes in survival induced by the policy change in 1986 when the FDA approved the PSA test to help monitor progress of prostate cancer. We modeled the crude hazard for censoring with a proportional hazards model containing PSA_S, R_S, and covariates for grade, year of diagnosis, and age at diagnosis. We took $B \sim$Exponential(1) and standard error estimates are based on 1000 bootstrap replicates.

Figure 2 provides examples of model-checking plots for the time-dependent pro-

17
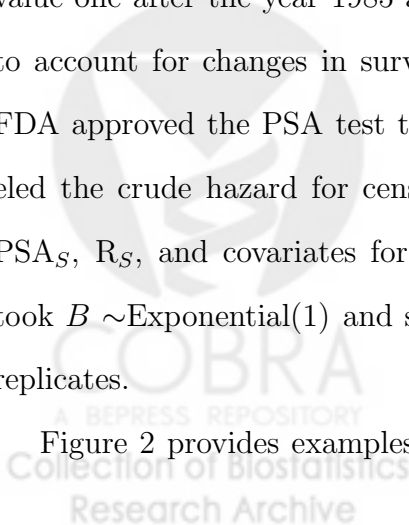
portional hazards and proportional odds cure models. The plot of Equation (11) against time for the proportional odds cure model does not appear to have mean zero and the p-value from 1000 bootstrap replicates with $B \sim \text{Exponential}(1)$ is 0.008, which indicates that the proportional odds cure model may not be appropriate. The corresponding plot for the proportional hazards cure model shows an approximately mean zero trend and the p-value from 1000 bootstrap replicates with $B \sim \text{Exponential}(1)$ is 0.416. These data suggest the proportional hazards cure model fits the data better than the proportional odds cure model, so we will present results for the proportional hazards cure model. Results for the time-dependent model and a univariate model containing only R_S and R_C can be found in Table 3.

The race covariate is significant in the univariate model for both the survival and cure functions. The signs of the estimates indicate that black race is associated with worse survival and a smaller probability of being cured. These covariates remain significant in the time-dependent model, though the time-dependent covariate indicating PSA approval is not significant.

We contrast our univariate results with alternative methods in Tables 4 and 5. All models contain one covariate denoting race. The standard proportional hazards model (ignoring possible cure fractions and dependent censoring) estimate is almost identical to the competing risks proportional hazards model estimate with deaths from other causes treated as censored events (Fine and Gray 1999). The three parameter values presented in Table 4 have different interpretations. The parameter estimate from our model has the interpretation of the log-hazard ratio *for subjects who are not cured.* The standard proportional hazards model parameter has the interpretation of the log-hazard ratio assuming the population is composed entirely of persons who are not cured. The log-hazard ratio for the subdistribution for death from prostate cancer are not easily interpretable, as noted in Fine and Gray (1999).

18

Under the assumption of independent censoring, the tail end of the Kaplan-Meier curve can estimate the proportion of cured subjects in a sample (Maller and Zhou 1992). Table 5 compares estimates of the tail of the Kaplan-Meier survival function stratified by race to estimates of the cure proportion found from transforming the regression results from the our proportional hazards cure model. The proportional hazards cure model has a smaller cure fraction than the tail end of the Kaplan-Meier plot indicates, and the confidence intervals for the estimates do not overlap.

As a referee noted, competing risks models (Prentice et al. 1978; Fine and Gray 1999) are a useful way to model survival data with the potential for dependent censoring. However, this type of analysis may not be ideal for this specific application. First, competing risks models can characterize the crude risk of failure for a specific failure type, but do so without removing dependent risks (Pepe and Mori 1993). We wish to evaluate the effect of race on prostate cancer survival, but these competing risks models will not allow for the direct covariate interpretation we are interested in. Second, a subject who is cured of prostate cancer could die of heart disease (a competing risk). Therefore the probability of being cured is not readily modeled using the existing competing risk framework, in contrast to the model we proposed.

8. DISCUSSION

This paper provides a unified framework for a class of transformation cure models that allow for dependent censoring. We utilize an inverse censoring probability formula to develop unbiased estimating equations. Our method allows one to estimate cure fractions more accurately in cases where there is dependence between censoring and survival times. We also develop a weighted bootstrap for inference procedures and supply a method to check whether the model specification is adequate.

Extending our estimation procedure to be able to incorporate internal time-

19

dependent covariates is a subject of future research. This extension would allow for a greater range of application in dependent censoring situations. A common cause of dependent censoring is excluding covariates related to both the survival and censoring times. This is an issue in any model, including models with only time-independent covariates. As shown in our simulation study, our method can decrease bias in situations with only time-independent covariates.

Rigorous testing for whether there is sufficient follow-up to identify cure proportions is difficult even without dependent censoring. A test for sufficient follow-up in the presence of dependent censoring may deserve a separate paper. A number of tests for sufficient follow-up in cure survival situations have been put forward (Maller and Zhou 1994, 1995; Klebanov and Yakovlev 2007). These tests, combined with subject matter specific knowledge on whether a mixture cure formulation is suitable, can provide information to allow an investigator to decide whether it is appropriate to use this model.

Appendix: Technical Details

Before we present the asymptotic results we will review and define some notation. All expectations are with respect to the true distributions of all random variables involved. Denote the Euclidean norm as $|| \cdot ||$, the supremum norm on $[0, \nu]$ as $|| \cdot ||_\infty$, and the norm corresponding to the parameters of $\boldsymbol{\alpha}$ as $|| \cdot ||_{\boldsymbol{\alpha}}$. For example, if we assumed the crude hazard for censoring followed a proportional hazards form, our model would have two parameters: $\boldsymbol{\alpha}_E$, a Euclidean parameter associated with $\mathbf{W}(\cdot)$, and $\boldsymbol{\alpha}_\Lambda$, associated with the cumulative baseline hazard. In this case a sensible choice for $||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0||_{\boldsymbol{\alpha}}$ would be $||\boldsymbol{\alpha}_E - \boldsymbol{\alpha}_{0E}|| + ||\boldsymbol{\alpha}_\Lambda - \boldsymbol{\alpha}_{0\Lambda}||_\infty$. Define $||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| = ||\boldsymbol{\alpha} - \boldsymbol{\alpha}_0||_{\boldsymbol{\alpha}} + ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| + ||\boldsymbol{\gamma} - \boldsymbol{\gamma}_0|| + ||H - H_0||_\infty$. Let a single superscript dot denote derivatives and a double superscript dot denote second derivatives.

20

Consider the following regularity conditions:

(c.1) the covariates $\mathbf{X}_i$ and the covariate processes $\mathbf{Z}_i(\cdot)$ and $\mathbf{W}_i(\cdot)$ are uniformly bounded, have nondegenerate variance-covariance matrices, and are cadlag for $i = 1, \ldots, n$;

(c.2) $\boldsymbol{\phi}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0)$ is a point in a bounded, open, convex subset of $\mathbb{R}^{\xi+\kappa}$;

(c.3) the transformation function g is twice differentiable. The function g, its derivative $\dot{g}$, and its second derivative $\ddot{g}$ are all Lipschitz;

(c.4) $H_0(t)$ is differentiable and is of bounded variation over $[0, \nu]$;

(c.5) $E(n^{-1}\dot{\mathbf{U}}_{n\boldsymbol{\phi}}(\boldsymbol{\theta})|_{\boldsymbol{\phi}=\boldsymbol{\phi}_0})$ is nondegenerate with probability 1;

(c.6) $\max N_i(\nu) > 0, \; i = 1, \ldots, n$;

(c.7) $K(\cdot)$ is uniformly bounded away from 0 on $[0, \nu]$. The estimating equation $\mathbf{U}_{n\boldsymbol{\alpha}}$ meets conditions 1, 3, and 4 of Theorem 2. $\mathbf{U}_{n\boldsymbol{\alpha}}$ is Fréchet differentiable at $\boldsymbol{\alpha}_0$ and the derivative is continuously invertible;

(b.1) the vectors $(b_{n1}, \ldots, b_{nn})'$ are exchangeable for all $n = 1, 2, \ldots$;

(b.2) $b_{ni} \geq 0$ for all n and all i and $\sum_{i=1}^n b_{ni} = n$ for all n;

(b.3) the $L_{2,1}$ norm of b_{n1} is bounded, i.e. $\int_0^\infty \sqrt{P_B(b_{n1} \geq u)}\, du \leq M^* < \infty$;

(b.4) $\lim_{\psi \to \infty} \limsup_{n \to \infty} \sum_{t \geq \psi} t^2 P(b_{n1} \geq t) = 0$;

(b.5) $n^{-1} \sum_{i=1}^n (b_{ni} - 1)^2 \to c^2 > 0$ in P_B-probability.

Conditions (c.1)-(c.5) are standard assumptions in survival analysis literature. (c.6) is used in the proofs of Theorems 1 and 2. (c.7) requires that the estimates

21

for the weight function, K, meet the same regularity conditions that we impose on $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and H. These conditions have been previously shown for a variety of survival functions (e.g. van der Vaart 1998, Section 25.12 for the proportional hazards model).

Conditions (b.1) - (b.5) are met by most positive distributions by defining $b_{ni} = b_i'/\bar{b}_n'$, where b_i' is an observed realization from the bootstrap distribution, B, and $\bar{b}'$ denotes the average of n observed realizations. If $B \sim \text{Exponential}(1)$, B meets all the regularity conditions.

The following are sketches of the proofs. More complete details can be found in a technical report available from the first author.

Proof of Theorem 1

We want to show that

$$
\bar{G}(\boldsymbol{\gamma}'\mathbf{X}) + G(\boldsymbol{\gamma}'\mathbf{X})g\left\{\int_0^t \exp(\boldsymbol{\beta}'\mathbf{Z}(u))\,dH(u)\right\}
$$
$$
= \bar{G}(\tilde{\boldsymbol{\gamma}}'\mathbf{X}) + G(\tilde{\boldsymbol{\gamma}}'\mathbf{X})g\left\{\int_0^t \exp(\tilde{\boldsymbol{\beta}}'\mathbf{Z}(u))\,d\tilde{H}(u)\right\} \tag{12}
$$

for all $t \in [0, \nu]$ with probability 1 implies that $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$, $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, $H = \tilde{H}$. Equation (12) implies that there exists a time-independent constant v^* (possibly random) such that

$$
\frac{g\left\{\int_0^t \exp(\tilde{\boldsymbol{\beta}}'\mathbf{Z}(u))\,d\tilde{H}(u)\right\} - 1}{g\left\{\int_0^t \exp(\boldsymbol{\beta}'\mathbf{Z}(u))\,dH(u)\right\} - 1} = v^*
$$

for all $t \in [0, \nu]$. Rewrite $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \boldsymbol{\varpi}$ and $\tilde{H}(t) = H(t) + a\int_0^t \omega(u)\,dH(u)$ for some nonrandom function $\omega(u)$ and $a \in \mathbb{R}$. Express v^* as

$$
\frac{g\{\int_0^t \exp(\tilde{\boldsymbol{\beta}}'\mathbf{Z}(u))\,d\tilde{H}(u)\} - 1}{g\{\int_0^t \exp(\boldsymbol{\beta}'\mathbf{Z}(u))\,dH(u)\} - 1}
$$

22

$$= \frac{g\{\int_0^t (1 + a\omega(u)) \exp(\mathbf{Z}(u)'(\boldsymbol{\beta} + \boldsymbol{\varpi})) \, dH(u)\} - 1}{g\{\int_0^t \exp(\boldsymbol{\beta}'\mathbf{Z}(u)) \, dH(u)\} - 1} \tag{13}$$

for all $t \in [0, \nu]$. It follows that (13) is constant with respect to t if and only if $\boldsymbol{\varpi} = 0$ and $a = 0$. To see that this is the case, use the mean value theorem and write the right hand side of (13) as

$$1 + \bar{g}(t) \int_0^t \exp(\boldsymbol{\beta}'\mathbf{Z}(u))[\mathbf{Z}(u)'\boldsymbol{\varpi} + a] \, dH(u), \tag{14}$$

where

$$\bar{g}(t) = \frac{\dot{g}\{\int_0^t \{1 + \bar{a}\omega(u)\} \exp(\bar{\boldsymbol{\beta}}'\mathbf{Z}(u)) \, dH(u)\}}{g\{\int_0^t \exp(\boldsymbol{\beta}'\mathbf{Z}(u)) \, dH(u)\} - 1},$$

$\bar{\boldsymbol{\beta}} = \boldsymbol{\beta} + \bar{s}\tilde{\boldsymbol{\beta}}$ and $\bar{a} = \bar{s}a$ for some constant $\bar{s} \in [0, 1]$. Equation (14) is independent of t if and only if $\boldsymbol{\varpi} = 0$ and $a = 0$. Hence $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}$ and $\tilde{H} = H$. Therefore, a reëxamination of (12) yields that $G(\boldsymbol{\gamma}'\mathbf{X}) = G(\tilde{\boldsymbol{\gamma}}'\mathbf{X})$ with probability 1. Since the logistic function $G(\cdot)$ is invertible, it follows that $\boldsymbol{\gamma}'\mathbf{X} = \tilde{\boldsymbol{\gamma}}'\mathbf{X}$ or $(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}})'\mathbf{X} = 0$ with probability 1. As $\mathbf{X}$ has a nondegenerate variance-covariance matrix, $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$.

Proof of Theorem 2

First consider the convergence of $\hat{H}(t, \boldsymbol{\phi}_0, \boldsymbol{\alpha}_0)$, which is solved recursively from (10) with $\boldsymbol{\phi}$ and $K(\cdot)$ fixed at the true values, $\boldsymbol{\phi}_0$ and $K_0(\cdot)$. Let $\mathbb{H}$ be the set of all nondecreasing, nonnegative, finite step functions with $H(\cdot) \in \mathbb{H}$ such that $H(0) = 0$ and $H(\cdot)$ jumps only at the observed failure times from the data, $t_1 < \ldots < t_k \leq \nu$. We want to show that $||\hat{H}(t, \boldsymbol{\phi}_0, \boldsymbol{\alpha}_0) - H_0(t)||_\infty \to 0$ almost surely as $n \to \infty$. Define the mapping Q on $\mathbb{H}$ as

$$Q(H)(t) = \sum_{i=1}^n \int_0^t dN_i(u)/K_i(u-) + G(\boldsymbol{\gamma}_0'\mathbf{X}_i) \, dS_g(\bar{\mathbf{Z}}_i(t), t, \boldsymbol{\beta}_0).$$

23

For an arbitrary but fixed $\varepsilon' > 0$ consider $H_1 \in \mathbb{H}$ and $H_2 \in \mathbb{H}$ such that $||H_1 - H_2||_\infty \geq \varepsilon'$. Denote the jump size of H_l at t_j as $H_l\{t_j\}$, $l = 1, 2$. Write $Q(H_1)(t) - Q(H_2)(t) = \sum_{t_j \leq t} w_j a_j$ where $w_j = \sum_{i=1}^n G(\boldsymbol{\gamma}_0' \mathbf{X}_i) \dot{g}(s_i) \exp(\boldsymbol{\beta}_0' \mathbf{Z}_i(t_j))$, $0 \leq s_i \leq |\int_0^t \exp(\boldsymbol{\beta}_0' \mathbf{Z}_i(u)) d\, H_1(u) - \int_0^t \exp(\boldsymbol{\beta}_0' \mathbf{Z}_i(u)) d\, H_2(u)|$, and $a_j = H_1\{t_j\} - H_2\{t_j\}$. Since $||H_1(t) - H_2(t)||_\infty \geq \varepsilon'/2$, there must exist t_{j_0} such that $a_{j_0} = |H_1\{t_{j_0}\} - H_2\{t_{j_0}\}| \geq \varepsilon'/2k$. Note that since $\mathbf{Z}(\cdot)$ and $\mathbf{X}$ are uniformly bounded there must exist some finite $\bar{c}$ such that $n/\bar{c} \leq w_j \leq n\bar{c}$ for $j = 1, \ldots, k$.

Suppose that $|\sum_{t_j \leq t_{j_0}} w_j a_j| \leq \varepsilon' n/4k\bar{c}$. It follows that $|\sum_{t_j \leq t_{j_0-1}} w_j a_j| \geq |w_{j_0} a_{j_0}| - \varepsilon' n/4k\bar{c} \geq \varepsilon' n/2k\bar{c} - \varepsilon' n/4k\bar{c} = \varepsilon' n/4k\bar{c}$. Therefore, either $|\sum_{t_j \leq t_{j_0}} w_j a_j| \geq \varepsilon' n/4k\bar{c}$ or $|\sum_{t_j \leq t_{j_0-1}} w_j a_j| \geq \varepsilon' n/4k\bar{c}$. Hence

$$\|\sum_{i=1}^n G(\boldsymbol{\gamma}_0' \mathbf{X}_i)[S_g(t, \bar{\mathbf{Z}}_i(t), \boldsymbol{\beta}_0, H_1) - S_g(t, \bar{\mathbf{Z}}_i(t), \boldsymbol{\beta}_0, H_2)]\|_\infty \geq \varepsilon' n/4k\bar{c} \geq \varepsilon'/4\bar{c}.$$

Select $H' \in \mathbb{H}$ such that $H'(t_j) = H_0(t_j)$ for $i = 1, \ldots, k$. The law of large numbers and the continuity of H_0 imply that $\|Q(H')(t)\|_\infty \to 0$ almost surely as $n \to \infty$. We defined $\hat{H}(t, \boldsymbol{\phi}_0, \boldsymbol{\alpha}_0)$ such that $Q(\hat{H}(t, \boldsymbol{\phi}_0, \boldsymbol{\alpha}_0)) = o_p(1)$ for all $t \in [0, \nu]$, which implies that $\|Q(H')(t) - Q\{\hat{H}(\boldsymbol{\phi}_0, \boldsymbol{\alpha}_0)\}(t)\|_\infty \to 0$ almost surely as $n \to \infty$. Therefore $\hat{H}(\cdot, \boldsymbol{\phi}_0, \boldsymbol{\alpha}_0)$ is in an ε' neighborhood of H' in $[0, \nu]$ under $|| \cdot ||_\infty$ with probability 1. Using the facts that $\hat{H}(\cdot, \boldsymbol{\phi}_0, \boldsymbol{\alpha}_0)$ and H_0 are monotone, and that $\varepsilon' > 0$ can be made arbitrarily small, we conclude that $||\hat{H}(\cdot, \boldsymbol{\phi}_0, \boldsymbol{\alpha}_0) - H_0(\cdot)||_\infty \to 0$ in $[0, \nu]$ almost surely.

We apply the inverse function theorem to prove consistency for $\boldsymbol{\phi}$ (Foutz 1977). That the estimating equations for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are asymptotically unbiased follows from the consistency of $\hat{K}(\cdot)$ and the convergence of $\hat{H}(\cdot, \boldsymbol{\phi}_0, \boldsymbol{\alpha}_0)$. Tedious but straight-forward calculations can show that $En^{-1}\dot{\mathbf{U}}_{n\boldsymbol{\phi}}(\boldsymbol{\theta})|_{\boldsymbol{\phi}=\boldsymbol{\phi}_0}$ converges uniformly in a neighborhood of the true parameters and is negative definite. Hence $\boldsymbol{\phi}$ is consis-

tent. Finally, the consistency of $\hat{H}(\cdot, \boldsymbol{\theta})$ follows immediately from the convergence of $H(\cdot, \boldsymbol{\phi}_0, \boldsymbol{\alpha}_0)$, the consistency of $\boldsymbol{\phi}$ and K (consistency of K is assumed in (c.7)), and the continuity of $\hat{H}(\cdot, \boldsymbol{\phi})$ with respect to $\boldsymbol{\phi}$ and $K(\cdot)$.

Proof of Theorem 3

We use Theorem 3.3.5 of van der Vaart and Wellner (1996) and Theorem 3.1 from Wellner and Zhan (1996). In the following, define $\mathbf{U}_j = E\boldsymbol{\mu}_j$ for $j = \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, H$ and $\mathbf{U} = (\mathbf{U}_{\boldsymbol{\alpha}}, \ldots, U_H)$, where the $\boldsymbol{\mu}_j$ are defined as in Section 4. Let $\mathcal{H}_{\boldsymbol{\beta}}$ and $\mathcal{H}_{\boldsymbol{\gamma}}$ to be finite sets with same cardinality as the parameters and let $\mathcal{H}_H = \{h_t = 1_{[0,t]} : t \in [0, \nu]\}$. The following four conditions need to be verified:

1. For $j = \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, H$ and for any $\delta_n \to 0$: $\sup\{||\sqrt{n}(\mathbf{U}_{nj} - \mathbf{U}_j)(\boldsymbol{\theta}) - \sqrt{n}(\mathbf{U}_{nj} - \mathbf{U}_j)(\boldsymbol{\theta}_0)||_j / 1 + \sqrt{n}||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| : ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| \le \delta_n\} = o_p(1)$, where $|| \cdot ||_j$ denotes the norm associated with parameter j.

2. U is Fréchet differentiable at $\boldsymbol{\theta}_0$ with derivative denoted $\dot{\mathbf{U}}_0$, and $\dot{\mathbf{U}}_0$ has a continuous inverse.

3. $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}_0$ and $U_n(\hat{\boldsymbol{\theta}}) = o_p(n^{-1/2})$.

4. If $\tilde{D}_{nj}$ is the envelope function of the class $\mathbb{D}_{nj} \equiv \{\frac{\mu_j(\boldsymbol{\theta})(h) - \mu_j(\boldsymbol{\theta}_0)(h)}{1 + \sqrt{n}||\boldsymbol{\theta} - \boldsymbol{\theta}_0||} : h \in \mathcal{H}_j, ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < \delta_n\}$, for every $\delta_n \to 0$ and $j = \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, H$,

$$\lim_{\psi \to \infty} \limsup_{n \to \infty} \sup_{t \ge \psi} t^2 P(\tilde{D}_{nj}(\mathbf{X}, \mathbf{W}, \mathbf{Z}, \Delta) > t) = 0.$$

Conditions *3* and *4* clearly hold in this situation. Theorem 2 and (c.5) establish the consistency of $\hat{\boldsymbol{\theta}}$, and the second requirement of Condition *3* holds by definition of $\hat{\boldsymbol{\theta}}$. Let $\tilde{M} = \max(||\mathbf{Z}(t)||_{\infty}, ||\mathbf{X}||, 2)$ and let $\tilde{D}_{n\boldsymbol{\beta}} = \tilde{D}_{n\boldsymbol{\gamma}} = \tilde{D}_{nH} = 2\tilde{M}$. It follows that Condition *4* is met for $j = \boldsymbol{\beta}, \boldsymbol{\gamma}, H$. Condition *4* is met for $j = \boldsymbol{\alpha}$ by (c.7).

25

For $j = \boldsymbol{\beta}, \boldsymbol{\gamma}, H$, Condition *1* can be shown using standard arguments by building up Donsker classes of terms (Parner 1998; van der Vaart and Wellner 1996) and by using the continuity of K, G, and g and the pointwise convergence of $\boldsymbol{\theta}$ to $\boldsymbol{\theta}_0$. Condition *1* holds for $j = \boldsymbol{\alpha}$ by (c.7).

The last step is to verify Condition *2*. Write the Fréchet derivative, $\dot{\mathbf{U}}_0$, of $\mathbf{U}$ at $\boldsymbol{\theta}_0$ as a partitioned matrix: $\begin{pmatrix} \mathtt{B} & \mathtt{C} \\ \mathtt{D} & \mathtt{E} \end{pmatrix}$, where $\mathtt{B}$ corresponds to the the Fréchet derivative of $\mathbf{U}_{\boldsymbol{\alpha}}$ with respect to $\boldsymbol{\alpha}$ evaluated at $\boldsymbol{\theta}_0$; $\mathtt{C}$ to the Fréchet derivative of $\mathbf{U}_{\boldsymbol{\alpha}}$ with respect to $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and H evaluated at $\boldsymbol{\theta}_0$; $\mathtt{D}$ to the Fréchet derivative of $\mathbf{U}_{\boldsymbol{\beta}}, \mathbf{U}_{\boldsymbol{\gamma}}, U_H$ with respect to $\boldsymbol{\alpha}$ evaluated at $\boldsymbol{\theta}_0$; and $\mathtt{E}$ to the Fréchet derivative of $\mathbf{U}_{\boldsymbol{\beta}}, \mathbf{U}_{\boldsymbol{\gamma}}, U_H$ with respect to $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and H evaluated at $\boldsymbol{\theta}_0$. The inverse of this partitioned matrix can be written as $\begin{pmatrix} \mathtt{B}^{-1} + \mathtt{B}^{-1}\mathtt{C}\mathtt{F}^{-1}\mathtt{D}\mathtt{B}^{-1} & -\mathtt{B}^{-1}\mathtt{C}\mathtt{F}^{-1} \\ -\mathtt{F}^{-1}\mathtt{D}\mathtt{B}^{-1} & \mathtt{F}^{-1} \end{pmatrix}$, where $\mathtt{F} = \mathtt{E} - \mathtt{D}\mathtt{B}^{-1}\mathtt{C}$ (Stapleton 1995). Therefore, to prove that the partitioned matrix is continuously invertible, one only needs to show that $\mathtt{B}$ and $\mathtt{F}$ are continuously invertible. Since $\mathbf{U}_{\boldsymbol{\alpha}}$ is not a function of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, or H, $\mathtt{C}$ is the zero matrix and we only need to show that $\mathtt{B}$ and $\mathtt{E}$ are continuously invertible. $\mathtt{B}$ is continuously invertible by (c.5).

To verify that $\mathtt{E}$ is continuously invertible, write $\mathtt{E}$ as a partitioned matrix and use the same argument as above. Specifically, write: $\mathtt{E} = \begin{pmatrix} \dot{\mathbf{U}}_p & \dot{\mathbf{U}}_q \\ \dot{\mathbf{U}}_r & \dot{\mathbf{U}}_s \end{pmatrix}$, where $\dot{\mathbf{U}}_p$ is the derivative of $(\mathbf{U}_{\boldsymbol{\beta}}, \mathbf{U}_{\boldsymbol{\gamma}})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ evaluated at $\boldsymbol{\theta}_0$; $\dot{\mathbf{U}}_q$ is the derivative of $(\mathbf{U}_{\boldsymbol{\beta}}, \mathbf{U}_{\boldsymbol{\gamma}})$ with respect to H evaluated at $\boldsymbol{\theta}_0$; $\dot{\mathbf{U}}_r$ is the derivative of U_H with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ evaluated at $\boldsymbol{\theta}_0$; and $\dot{\mathbf{U}}_s$ is the derivative of U_H with respect to H evaluated at $\boldsymbol{\theta}_0$. We need to show that $\dot{\mathbf{U}}_p$ and $\dot{\mathbf{U}}_s - \dot{\mathbf{U}}_r \dot{\mathbf{U}}_p^{-1} \dot{\mathbf{U}}_q$ are continuously invertible. $\dot{\mathbf{U}}_p$ is finite dimensional and so continuous invertibility is a consequence of it being one-to-one. $\dot{\mathbf{U}}_s$ can be shown to be continuously invertible by direct computation, and $\dot{\mathbf{U}}_r \dot{\mathbf{U}}_p^{-1} \dot{\mathbf{U}}_q$ can be shown to be compact by invoking Helly's

26

selection theorem and the dominated convergence theorem. Direct calculation can show that $(\dot{\mathbf{U}}_s - \dot{\mathbf{U}}_r\dot{\mathbf{U}}_p^{-1}\dot{\mathbf{U}}_q)[A] = 0$ if and only if $||A||_\infty = 0$. We can therefore conclude that $\dot{\mathbf{U}}_s - \dot{\mathbf{U}}_r\dot{\mathbf{U}}_p^{-1}\dot{\mathbf{U}}_q$ is one-to-one. The sum of a continuously invertible operator and a compact operator is continuously invertible if the sum is one-to-one (Rudin 1987). Therefore $\dot{\mathbf{U}}_s - \dot{\mathbf{U}}_r\dot{\mathbf{U}}_p^{-1}\dot{\mathbf{U}}_q$ is continuously invertible and the result follows.

27

References

Abbring, J. and van den Berg, G. (2003), "The Identifiability of the Mixed Proportional Hazards Competing Risks Model," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 701–710.

American Cancer Society (2007), "Prostate Cancer: Detailed Guide," .

Asirvatham, A., Schmidt, M., Gao, B., and Chaudhary, J. (2006), "Androgens Regulate the Immune/Inflammatory Response and Cell Survival Pathways in Rat Ventral Prostate Epithelial Cells," *Endocrinology*, 147, 257–271.

Bennett, S. (1983), "Analysis of Survival Data by the Proportional Odds Model," *Statistics in Medicine*, 2, 273–277.

Berkson, J. and Gage, R. (1952), "Survival Curve for Cancer Patients Following Treatment," *Journal of the American Statistical Association*, 47, 501–515.

Bravi, F., Scotti, L., Bosetti, C., Talamini, R., Negri, E., Montella, M., Franceschi, S., and La Vecchia, C. (2006), "Self-reported History of Hypercholesterolaemia and Gallstones and the Risk of Prostate Cancer," *Annals of Oncology*, 17, 1014–1017.

Chen, M., Ibrahim, J., and Sinha, D. (1999), "A New Bayesian Model for Survival Data with a Surviving Fraction." *Journal of the American Statistical Association*, 94, 909–919.

Cheng, S., Wie, L., and Ying, Z. (1995), "Analysis of Transformation Models with Censored Data," *Biometrika*, 82, 835–845.

Cox, D. (1972), "Regression Models and Life-tables (with Discussion)," *Journal of the Royal Statistical Society, Series B (Methodological)*, 34, 187–220.

Farewell, V. (1982), "The Use of Mixture Models for the Analysis of Survival Data with Long-term Survivors," *Biometrics*, 38, 1041–1046.

Fine, J. (1999), "Analysing Competing Risks Data with Transformation Models," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61, 817–830.

Fine, J. and Gray, R. (1999), "A Proportional Hazards Model for the Subdistribution of a Competing Risk," *Journal of the American Statistical Association*, 94, 496–509.

Fine, J., Ying, Z., and Wei, L. (1998), "On the Linear Transformation Model for Censored Data," *Biometrika*, 85, 980–986.

Fleming, T. and Harrington, D. (1991), *Counting Processes and Survival Analysis*, Wiley.

Fleshner, N., Bagnell, P., Kltoz, L., and Venkateswaran, V. (2004), "Dietary Fat and Prostate Cancer," *The Journal of Urology*, 171, 19–24.

Foutz, R. (1977), "On the Unique Consistent Solution to the Likelihood Equations," *Journal of the American Statistical Association*, 72, 147–148.

Hanlon, A. and Hanks, G. (2000), "Failure Patterns and Hazard Rates for Failure Suggest the Cure of Prostate Cancer by External Beam Radiation," *Urology*, 55, 725–729.

Heckman, J. and Honore, B. (1989), "The Identifiability of the Competing Risks Model," *Biometrika*, 76, 325–330.

Homma, Y., Kondo, Y., Kaneko, M., Kitamura, T., Nyou, W., Yanagisawa, M., Ya-

29

mamoto, Y., and Kakizoe, T. (2004), "Promotion of Carcinogenesis and Oxidative Stress by Dietary Cholesterol in Rat Prostate," *Carcinogenesis*, 25, 1011–1014.

Hughes-Fulford, M., Tjandrawinata, R., Li, C., and Sayyah, S. (2005), "Arachidonic Acid, an Omega-6 Fatty Acid, Induces Cytoplasmic Phospholipase A_2 in Prostate Carcinoma Cells," *Carcinogenesis*, 26, 1520–1526.

Jani, A. and Hellman, S. (2003), "Early Prostate Cancer: Clinical Decision-Making," *The Lancet*, 361, 1045–1053.

Keating, N., O'Malley, A., and Smith, M. (2006), "Diabetes and Cardiovascular Disease During Androgen Deprivation Therapy for Prostate Cancer," *Journal of Clinical Oncology*, 24, 4448–4455.

Klebanov, L. and Yakovlev, A. (2007), "A New Approach to Testing for Sufficient Follow-up in Cure-rate Analysis," *Journal of Statistical Planning and Inference*, 137, 3557–3569.

Kosorok, M. and Song, R. (2007), "Inference Under Right Censoring for Transformation Models with a Change-point Based on a Covariate Threshold," *The Annals of Statistics*, 35, 957–989.

Kuk, A. and Chen, C. (1992), "A Mixture Model Combining Logistic Regression with Proportional Hazards Regression," *Biometrika*, 79, 531–541.

Lee, S. (2006), "Identification of a Competing Risks Model with Unknown Transformations of Latent Failure Times," *Biometrika*, 93, 996.

Li, C., Taylor, J., and Sy, J. (2001), "Identifiability of Cure Models," *Statistics & Probability Letters*, 54, 389–395.

30

Li, Y., Tiwari, R., and Guha, S. (2007), "Mixture Cure Survival Models with Dependent Censoring," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 285–306.

Lin, D., Wei, L., and Ying, Z. (1993), "Checking the Cox Model with Cumulative Sums of Martingale-based Residuals," *Biometrika*, 80, 557–572.

Lu, W. and Ying, Z. (2004), "On Semiparametric Transformation Cure Models," *Biometrika*, 91, 331–343.

Maller, R. and Zhou, S. (1992), "Estimating the Proportion of Immunes in a Censored Sample," *Biometrika*, 79, 731–739.

— (1994), "Testing for Sufficient Follow-Up and Outliers in Survival Data." *Journal of the American Statistical Association*, 89, 1499–1506.

— (1995), "Testing for the Presence of Immune or Cured Individuals in Censored Survival Data," *Biometrics*, 51, 1197–1205.

Miloslavsky, M., Keles, S., Laan, M., and Butler, S. (2004), "Recurrent Events Analysis in the Presence of Time-dependent Covariates and Dependent Censoring," *Journal of the Royal Statistical Society Series B*, 66, 239–257.

Pandini, G., Mineo, R., Frasca, F., Roberts, C., Marcelli, M., Vigneri, R., and Belfiore, A. (2005), "Androgens Up-regulate the Insulin-like Growth Factor-I Receptor in Prostate Cancer Cells," *Cancer Research*, 65, 1849–1857.

Parner, E. (1998), "Asymptotic Theory for the Correlated Gamma-frailty Model," *Annals of Statistics*, 26, 183–214.

31

Peng, L. and Huang, Y. (2008), "Survival Analysis with Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637–649.

Peng, Y. and Dear, K. (2000), "A Nonparametric Mixture Model for Cure Rate Estimation," *Biometrics*, 56, 237–243.

Pepe, M. and Mori, M. (1993), "Kaplan-Meier, Marginal or Conditional Probability Curves in Summarizing Competing Risks Failure Time Data?" *Statistics in Medicine*, 12, 737–751.

Prentice, R., Kalbfleisch, J., Peterson Jr, A., Flournoy, N., Farewell, V., and Breslow, N. (1978), "The Analysis of Failure Times in the Presence of Competing Risks," *Biometrics*, 34, 541–554.

Ritch, C., Wan, R., Stephens, L., Taxy, J., Huo, D., Gong, E., Zagaja, G., and Brendler, C. (2007), "Dietary Fatty Acids Correlate With Prostate Cancer Biopsy Grade and Volume in Jamaican Men," *The Journal of Urology*, 177, 97–101.

Robins, J. (1992), "Estimation of the Time-dependent Accelerated Failure Time Model in the Presence of Confounding Factors," *Biometrika*, 79, 321–334.

— (1993), "Information Recovery and Bias Adjustment in Proportional Hazards Regression Analysis of Randomized Trials Using Surrogate Markers," *Proceedings of the Biopharmaceutical Section, American Statistical Association*, 24–33.

Robins, J. and Finkelstein, D. (2000), "Correcting for Noncompliance and Dependent Censoring in an AIDS Clinical Trial with Inverse Probability of Censoring Weighted (IPCW) Log-Rank Tests," *Biometrics*, 56, 779–788.

Robins, J. and Rotnitzky, A. (1992), "Recovery of Information and Adjustment for

Dependent Censoring Using Surrogate Markers," in *AIDS Epidemiology: Methodological Issues*, eds. Jewell, N., Dietz, K., and Farewell, V., Birkhauser, pp. 297–331.

Rotnitzky, A. and Robins, J. (1995), "Semiparametric Regression Estimation in the Presence of Dependent Censoring," *Biometrika*, 82, 805–820.

— (2003), "Inverse Probability Weighted Estimation in Survival Analysis," in *The Encyclopedia of Biostatistics, Second Edition*, eds. Armitage, P. and Colton, T., Wiley, pp. 2619–2625.

Rudin, W. (1987), *Functional Analysis*, McGraw-Hill.

Stapleton, J. H. (1995), *Linear Statistical Models*, Wiley.

Strom, S., Yamamura, Y., Forman, M., Pettaway, C., Barrera, S., and Digiovanni, J. (2008), "Saturated Fat Intake Predicts Biochemical Failure After Prostatectomy." *International Journal of Cancer*, 122, 2581–2585.

Sy, J. and Taylor, J. (2000), "Estimation in a Cox Proportional Hazards Cure Model," *Biometrics*, 56, 227–236.

Tai, P., Yu, E., Cserni, G., Vlastos, G., Royce, M., Kunkler, I., and Vinh-Hung, V. (2005), "Minimum Follow-up Time Required for the Estimation of Statistical Cure of Cancer Patients: Verification Using Data from 42 Cancer Sites in the SEER Database," *BMC Cancer*, 5, 48.

Tsodikov, A., Ibrahim, J., and Yakovlev, A. (2003), "Estimating Cure Rates from Survival Data: An Alternative to Two-component Mixture Models," *Journal of the American Statistical Association*, 98, 1063–1079.

van der Vaart, A. (1998), *Asymptotic Statistics*, Cambridge University Press.

33

van der Vaart, A. and Wellner, J. (1996), *Weak Convergence and Empirical Processes*, Springer.

Wellner, J. and Zhan, Y. (1996), "Bootstrapping Z-estimators," *University of Washington Department of Statistics Technical Report*, 308.

Wilson, P., D'Agostino, R., Levy, D., Belanger, A., Silbershatz, H., and Kannel, W. (1998), "Prediction of Coronary Heart Disease Using Risk Factor Categories," *Circulation*, 97, 1837–1847.

www.seer.cancer.gov (2004), "National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, Surveillance, Epidemiology, and End Results (SEER) Program, Public-Use Data (1973-2001)," Released April 2004, based on November 2003 submission.

Xu, Y., Chen, S., Ross, K., and Balk, S. (2006), "Androgens Induce Prostate Cancer Cell Proliferation Through Mammalian Target of Rapamycin Activation and Post-transcriptional Increases in Cyclin D Proteins," *Cancer Research*, 66, 7783–7792.

Zeng, D. and Lin, D. (2007), "Maximum Likelihood Estimation in Semiparametric Regression Models with Censored Data," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 507–564.

Zhuang, L., Kim, J., Adam, R., Solomon, K., and Freeman, M. (2005), "Cholesterol Targeting Alters Lipid Raft Composition and Cell Survival in Prostate Cancer Cells and Xenografts," *Journal of Clinical Investigation*, 115, 959–968.

34

Table 1: Simulation results

Proportional Hazards Model

			Our Method				Lu and Ying (2004)	
Parameter	True Values	Bias	SE_{MC}	SE_B	RMSE	Bias	SE_{MC}	RMSE
β	-0.507	0.136	0.162	0.175	0.212	0.040	0.126	0.132
γ_0	1	0.133	0.147	0.151	0.198	-0.136	0.148	0.201
γ_1	-1	0.038	0.164	0.209	0.168	0.078	0.203	0.217
β	-0.507	-0.014	0.164	0.236	0.165	0.017	0.117	0.118
γ_0	0.8	-0.074	0.147	0.160	0.165	-0.103	0.145	0.179
γ_1	-0.75	-0.017	0.191	0.212	0.192	0.054	0.199	0.206

Proportional Odds Model

			Our Method				Lu and Ying (2004)	
Parameter	True Values	Bias	SE_{MC}	SE_B	RMSE	Bias	SE_{MC}	RMSE
β	-0.608	0.017	0.173	0.232	0.174	0.243	0.133	0.277
γ_0	1	-0.133	0.219	0.209	0.256	0.455	0.255	0.525
γ_1	-1	0.101	0.299	0.274	0.316	-0.353	0.303	0.465
β	-0.608	-0.241	0.317	0.370	0.398	0.197	0.130	0.236
γ_0	0.8	-0.137	0.214	0.223	0.254	0.389	0.234	0.454
γ_1	-0.75	0.171	0.295	0.300	0.341	-0.213	0.278	0.350

Bias = average of estimate - true value
SE_{MC} = standard deviation of estimates
SE_B = mean of bootstrap standard errors
RMSE = $(Bias^2 + SE_{MC}^2)^{0.5}$

Table 2: Summary of causes of death among SEER data

Cause of Death	N (total=15,524)	%	SEER Code*
Prostate Cancer	3,717	23.9	28010
Diseases of the Heart	4,349	28.0	50060
Alive	2,208	14.2	00000
Other	5,250	33.8	all others

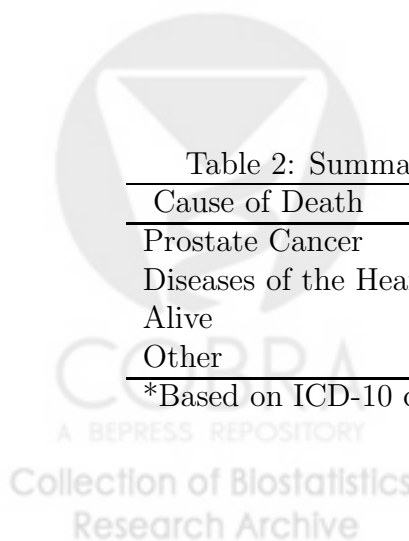*Based on ICD-10 codes, using SEER COD Recode 9/17/2004

Table 3: SEER data results for a proportional hazards cure model. Parameter estimates and associated standard errors from the regression models are provided. Estimates with the subscript S can be interpreted as the log-hazard ratio for non-cured subjects. Estimates with the subscript C can be interpreted as the log-odds for being not cured. Black race is the reference category for the race parameters.

Univariate

Parameter	Estimate	Standard error*
Race_S	-0.129	0.033
Intercept_C	0.245	0.008
Race_C	-0.169	0.011

Time-Dependent

Parameter	Estimate	Standard error*
Race_S	-0.123	0.031
PSA_S	-0.083	0.042
Intercept_C	0.245	0.008
Race_C	-0.170	0.011

*All standard error estimates are based on 1000 bootstrap replications with $B \sim$ Exponential(1).

Table 4: Parameter estimates for proportional hazards functions for three different models. Black race is the reference category for each of the models.

Model	Parameter estimate	Standard Error
Our proportional hazards cure model	-0.129	0.033
Standard proportional hazards model	-0.139	0.037
Competing risks proportional hazards model	-0.138	0.037

Table 5: Estimates of the proportion of men cured of prostate cancer: (1) from the tail of the Kaplan-Meier survival curve stratified by race and (2) from transforming the regression results from our proportional hazards cure model.

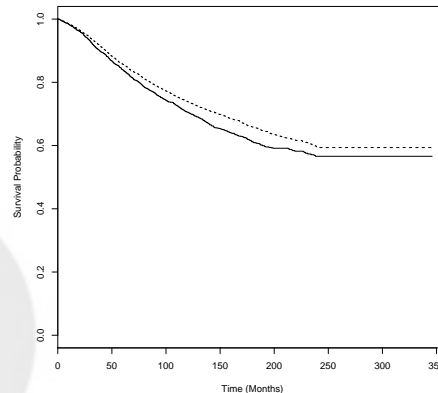|	Cure estimate	Confidence Interval
Our proportional hazards cure model		
Cure_B	0.439	(0.435, 0.443)
Cure_W	0.481	(0.474, 0.481)
Naïve Kaplan-Meier tail estimates		
Cure_B	0.566	(0.540, 0.593)
Cure_W	0.594	(0.578, 0.610)

Figure 1: Naïve Kaplan-Meier curves (ignoring potential dependent censoring) for survival comparing white subgroup (dashed line) to the black subgroup (solid line).
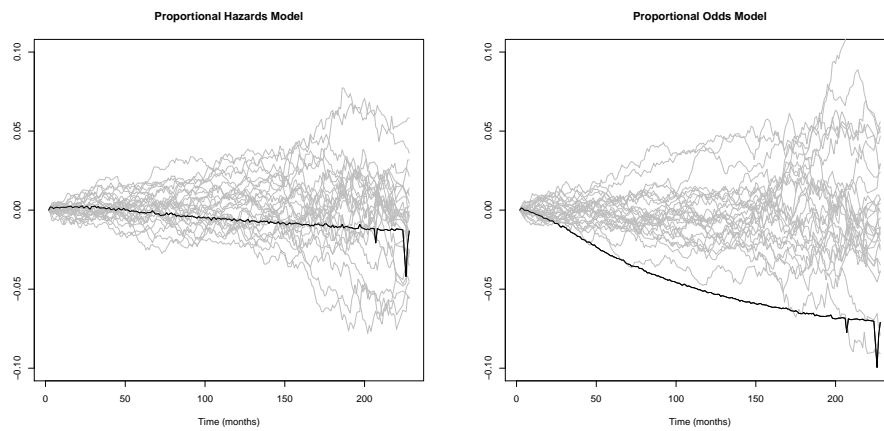
37

Figure 2: Model checking plots. Twenty-five random sample paths are plotted in grey with the observed paths marked in black.