

Semiparametric estimation of an accelerated failure time model with time-dependent covariates

BY JAMES ROBINS AND ANASTASIOS A. TSIATIS

Department of Biostatistics, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

SUMMARY

A class of semiparametric accelerated failure time models is introduced that are useful for modelling the relationship of survival distributions to time dependent covariates. A class of semiparametric rank estimators is derived for the parameters in this model when the survival data are right censored. These estimates are shown to be consistent, and asymptotically normal with variances that can be consistently estimated. It is also shown that estimators within this class achieve the semiparametric efficiency bound.

Some key words: Accelerated failure time model; Semiparametric efficiency; Survival analysis; Time dependent covariates.

1. INTRODUCTION

We consider an accelerated failure time model for time-to-event data in which the survival distribution is scaled by a factor which may be a function of time-dependent covariates. These models directly model the effect of covariates on the length of survival. This is in contrast to the more commonly-used proportional hazards model which models the hazard rate itself. In some applications it may be easier to visualize the concept that a treatment intervention or exposure to an environmental contaminant increases or decreases the length of survival by a certain proportion, as compared to the concept that the hazard rate is changed.

For time independent covariates, say, $Z = (Z_1, \dots, Z_p)'$, the class of accelerated failure time models assumes that the survival distribution at time t , given a set of covariates $Z = z$, is given by $\bar{F}(t|z) = \bar{F}_0\{h(z)t\}$. The proportionality constant $h(z)$ can be interpreted as the scale factor by which lifetime is decreased as a function of the covariates z . Often $h(z)$ is taken to be log linear or $h(z) = \exp(\beta'z)$. In such cases, if $z = 0$, we can interpret $\bar{F}_0(t)$ as the 'baseline' survival distribution or the survival distribution for individuals with covariates all equal to zero. A useful way of envisioning this model is to consider a hypothetical random variable, U , that corresponds to an individual's survival time if that individual had covariate values all equal to zero. This baseline failure time, U , would be modified for values of the covariate different than zero. For example, if we denote the actual survival time by T , then for an individual with covariate value Z we have $T = e^{-\beta'Z}U$. The parameters, β , in this model have a direct interpretation in terms of the increase or decrease in lifespan as a function of the covariates.

The assumption that U is independent of Z in this hypothetical construct, induces the probabilistic model

$$\begin{aligned}\bar{F}(t|Z) &= \text{pr}(T \geq t | Z) = \text{pr}\{e^{-\beta'Z}U \geq t | Z\} \\ &= \text{pr}(U \geq e^{\beta'Z}t | Z) = \bar{F}_0(e^{\beta'Z}t).\end{aligned}$$

This interpretation of the parameters β is similar to that in a linear regression model which assumes that $T_i = Z_i' \beta + \varepsilon_i$ and that the error, ε_i , for the i th individual is independent of the covariates Z_i .

This model is discussed by Kalbfleisch & Prentice (1980, Ch. 6) and Cox & Oakes (1984, § 5.2). Cox & Oakes describe a method of extending the accelerated failure time model to time-dependent covariates through a particular example of a transformation function $h(\cdot)$ which will be discussed later. However, they restrict their discussion to fully parametric models for which estimates can be obtained using standard maximum likelihood methods. In this paper, we propose a class of semiparametric models related to the parametric models of Cox & Oakes. These models are semiparametric in the sense that the relationship of the scale change of the survival distribution to the time-dependent covariates is parametric but the underlying baseline survival distribution is left arbitrary. We shall define a class of semiparametric estimates using estimating equations based on weighted log rank tests that can be applied to right censored data. We also consider a more general class of transformation functions. We shall outline some of the statistical properties of these estimates including asymptotic normality and also describe methods for estimating the asymptotic variance.

2. ACCELERATED FAILURE TIME MODELS WITH TIME-DEPENDENT COVARIATES

Let T_i be a continuous random variable recording time to failure for the i th subject, $i = 1, \dots, n$. Let $Z_i(t)$ be a random vector recording the value at time t of a vector of time-dependent covariates, and let $\bar{Z}_i(t) = \{Z_i(s); 0 \leq s \leq t\}$ be the history of the covariate process through time t . For the moment we shall assume that there is no censoring. For each subject, i , the observable random variables are $\{T_i, \bar{Z}_i(T_i)\}$.

For purposes of illustration, we will consider the use of an accelerated failure time model in the modelling of survival as a function of exposure to some environmental contaminant. For example, T_i might be survival and $Z_i(t)$ might represent the level of radon gas in the i th subject's home at time t . Then $\bar{Z}_i(t)$ would be the history of radon-exposure up to time t . We are interested in modelling the effect of radon exposure on an individual's survival time.

In the previous section, we suggested a causal interpretation in which a baseline failure time, U_i , is defined for each individual corresponding to the survival time of that individual in the hypothetical case that he/she was never exposed; that is, $Z_i(t) \equiv 0$. With time-dependent covariates it is useful to consider two time scales. One time scale will be the baseline time scale, u , for an individual if he/she was never exposed. On this time scale the i th individual would live U_i years. However, we shall also consider the observed time scale, t . On this time scale the i th individual who experiences a specific history of exposure $\bar{Z}_i(T_i)$ is observed to live T_i years. The model we consider assumes that there is a monotone transformation from the baseline time scale, u , to the observed time scale, t , as a function of the covariate history.

For ease of interpretation it is most convenient to consider the derivative of the transformation function, du/dt , as a function of the covariate history up to time t . Therefore, we take $du/dt = h\{\bar{Z}(t), \beta\}$ to correspond to the relative rate at which baseline time is being used up compared to actual time as a function of the history of exposure, $\bar{Z}(t)$, up to that time. For example, suppose $Z(t)$ is a time-dependent indicator of exposure; that is, $Z(t) = 1$ if the individual is exposed at time, t , and 0, otherwise. If we let $du/dt = e^{\beta Z(t)}$, then during times of exposure, the relative rate is e^β . For this model,

constant exposure, that is, $Z(t) = 1$ for all t , would shorten life by a constant factor e^β , that is, $T = e^{-\beta}U$, leading to the accelerated failure time model for independent covariates.

We shall require only that $du/dt = h\{\bar{Z}(t), \beta\}$ be a positive smooth known function with $h\{\bar{Z}(t), \beta\} = 1$ when $\bar{Z}(t) \equiv 0$ or $\beta = 0$, and $\beta \in R^p$ is an unknown vector of parameters to be estimated. The transformation function from u to t is

$$u = \int_0^t h\{\bar{Z}(s), \beta\} ds,$$

which we shall denote as $u = \psi\{\bar{Z}(t), \beta\}$. Therefore, an individual who lived T_i years with a specific history of exposure, would have lived U_i years if unexposed where $U_i = \psi\{\bar{Z}_i(T_i), \beta\}$. For the specific case when $h\{Z(t), \beta\} = e^{\beta Z(t)}$ this leads to

$$u = \int_0^t e^{\beta'Z(s)} ds, \quad U_i = \int_0^{T_i} e^{\beta'Z_i(s)} ds. \quad (2.1)$$

The transformation given by (2.1) was suggested by Cox & Oakes (1984, § 5.2).

In order to estimate the parameters, β , of our transformation function from a sample of data, we must first consider the probability model which generates the data. We assume that the distribution of baseline failure times, U_i , in a population has a distribution function, $\bar{F}_0(u)$, that is, $\text{pr}(U \geq u) = \bar{F}_0(u)$ which corresponds to the population lifetime distribution if no one in the population was exposed. If an individual is exposed to a specific history, $\bar{Z} = \{Z(s), 0 \leq s < \infty\}$, which is assumed to be independent of the baseline failure time, then the distribution of the observed failure time, T , would satisfy the relationship

$$\text{pr}[\psi\{\bar{Z}(T), \beta_0\} \geq u | \bar{Z}] = \bar{F}_0(u). \quad (2.2)$$

We shall denote by β_0 the true value of β that generates the data. We shall also denote by $U(\beta)$ the transformation of T given by $\psi\{\bar{Z}(T), \beta\}$, and let $U(\beta_0)$ be U , the true baseline failure time. The model (2.2) assumes the existence of the value of $Z(t)$ for values of $t \geq T$. However, the value of $Z(t)$ is generally not measured after the individual is no longer under observation. Therefore the definition (2.2) is not in terms of the observable random quantities, $\{T, \bar{Z}(T)\}$. For this reason, it is more convenient to express the model in terms of the hazard rates, namely $\lambda_0(u)$ is the limit as $h \rightarrow 0$ of

$$h^{-1} \text{pr}[u \leq U < u + h | U \geq u, \bar{Z}\{\psi^{-1}(u, \bar{Z}, \beta_0)\}], \quad (2.3)$$

where $U = \psi\{\bar{Z}(T), \beta_0\}$ and $\psi^{-1}(u, \bar{Z}, \beta_0)$ is defined as the value, t , such that $\psi\{\bar{Z}(t), \beta_0\} = u$. Since $U \geq u$ or $\psi\{\bar{Z}(T), \beta_0\} \geq u$ implies that $\psi^{-1}(u, \bar{Z}, \beta_0) < T$, it follows that (2.3) concerns only the joint distribution of the observable random variables, $\{T, \bar{Z}(T)\}$. In model (2.3), β_0 still indexes a transformation of the observed time scale to some common baseline scale. On this baseline scale the hazard rate is the same regardless of past covariate history. Model (2.3) will be referred to as the accelerated failure time model for time-dependent covariates. In this sense an accelerated failure time model is a direct competitor to the Cox proportional hazards model with time-dependent covariates in that they are both models for aspects of the joint distribution of the observable random variables $\{T, \bar{Z}(T)\}$.

In order to generalize (2.3) to allow for right censoring, let C be the random variable recording a subject's potential censoring time. We assume that the observable random variables are $\{X, \Delta, \bar{Z}(X)\}$, where X is the minimum of T and C , and $\Delta = 1$ if $X = T$ and zero otherwise. If we additionally assume that censoring is noninformative (Cox &

Oakes, 1984, § 5.2; Tsiatis, 1990), then the model can be described in terms of the cause-specific hazard function, namely $\lambda_0(u)$ is the limit as $h \rightarrow 0$ of

$$h^{-1} \text{pr}[u \leq V < u + h, \Delta = 1 | V \geq u, \bar{Z}\{\psi^{-1}(u, \bar{Z}, \beta_0)\}], \quad (2.4)$$

where $V(\beta) = \psi\{\bar{Z}(X), \beta\}$ and V is equal to $V(\beta_0)$. Since, as in the uncensored case, $V \geq u$ implies $\psi^{-1}(u, \bar{Z}, \beta_0) \leq X$, the model (2.4) refers only to the observable random variables $\{X, \Delta, \bar{Z}(X)\}$.

In the next section we propose semiparametric estimates, $\hat{\beta}$, for the model (2.4). These estimates are semiparametric in the sense that they will be shown to be consistent and asymptotically normal regardless of the underlying baseline distribution function, $\bar{F}_0(u)$. The estimates will be solutions to the estimating equation $S(\hat{\beta}) = 0$, where $S(\beta_0)$ is a nonparametric test of $\beta = \beta_0$, taking the form of a weighted log rank test. These estimates were originally proposed by Robins (1989, § 15c) and are similar to those proposed by Tsiatis (1990) for the censored linear regression model.

3. INTERVAL ESTIMATION AND TESTING OF β

For ease of exposition, we assume for now that β is a scalar parameter. In § 5 we discuss the generalization of these results to the multiparameter problem. The observed data can be represented as n independent vectors $\{X_i, \Delta_i, \bar{Z}_i(X_i)\}$, $i = 1, \dots, n$. The test statistic $S(\beta_0)$ is motivated by the definition of the model given in (2.4). Namely, if we choose the correct value of β_0 in the transformation $V_i(\beta) = \psi\{\bar{Z}_i(X_i), \beta\}$, then the cause-specific hazard $\lambda_0(u)$ of $V_i = V_i(\beta_0)$ at any u will be independent of the covariate history up to u . Therefore, for each subject, we define a scoring function

$$G_i(u, \beta) = g[\bar{Z}_i\{\psi^{-1}(u, \bar{Z}_i, \beta)\}].$$

This scoring function is a real-valued function of the baseline time u , and the covariate history defined up to the corresponding real time $t = \psi^{-1}(u, \bar{Z}_i, \beta)$. Typical examples of $g\{\bar{Z}(t)\}$ might include cumulative exposure, $\int Z(u) du$, where the integral is over the range $(0, t)$, or current dose, $Z(t)$.

The test statistic $S(\beta_0)$ is defined as follows. We first transform all the variables by computing $V_i(\beta_0) = \psi\{\bar{Z}_i(X_i), \beta_0\}$. Using the transformed variables we then compute a sum over the death times of the observed minus the expected scores, G_i at each death time. Specifically,

$$S(\beta_0) = \sum_{i=1}^n \Delta_i [G_i\{V_i(\beta_0), \beta_0\} - G^{\text{av}}\{V_i(\beta_0), \beta_0\}], \quad (3.1)$$

where $G^{\text{av}}\{V_i(\beta_0), \beta_0\}$ is the average of the scores evaluated at time $V_i(\beta_0)$ among the individuals who on the baseline scale are at risk at that time. Letting $Y_j(u, \beta_0)$ denote the indicator of whether the j th individual is at risk at time, u , that is $Y_j(u, \beta_0) = 1$ if $V_j(\beta_0) \geq u$ and 0 otherwise, then

$$G^{\text{av}}\{V_i(\beta_0), \beta_0\} = \sum_{j=1}^n G_j\{V_i(\beta_0), \beta_0\} Y_j\{V_i(\beta_0), \beta_0\} / \sum_{j=1}^n Y_j\{V_i(\beta_0), \beta_0\}.$$

The statistic (3.1) is similar to a weighted log rank test. According to model (2.4) the cause-specific hazard for $V_i(\beta_0)$ does not depend on the past covariate history. Further, the observed minus the expected scores

$$A_i(\beta_0) = [G_i\{V_i(\beta_0), \beta_0\} - G^{\text{av}}\{V_i(\beta_0), \beta_0\}]$$

are functions only of the past covariate history. Therefore, if we (i) condition on all the past history up to time $V_i(\beta_0)$, and (ii) condition on the fact that an individual died at time $V_i(\beta_0)$, then, by the assumption of model (2.4), each of the individuals at risk at that time are equally likely to have failed. Thus, the statistic, $A_i(\beta_0)$, has mean zero.

Similarly, we can show that $\text{cov}\{A_i(\beta_0), A_j(\beta_0)\} = 0$ if $V_j(\beta_0) < V_i(\beta_0)$, since $A_j(\beta_0)$ is fixed given the conditioning events (i) and (ii) from above. From these considerations we can show that $n^{-1}S(\beta_0)$ given by (3.1) has mean zero with variance that can be consistently estimated by $n^{-1}\Omega(\beta_0)$, where

$$\Omega(\beta_0) = \sum_{i=1}^n \Delta_i \left(\sum_{j=1}^n [G_j\{V_i(\beta_0), \beta_0\} - G^{\text{av}}\{V_i(\beta_0), \beta_0\}]^2 Y_j\{V_i(\beta_0), \beta_0\} \right. \\ \left. \times \left[\sum_{j=1}^n Y_j\{V_i(\beta_0), \beta_0\} \right]^{-1} \right) \quad (3.2)$$

is the sum over the failure times of the empirical variance of $G_j\{V_i(\beta_0), \beta_0\}$, over the subjects at risk at time $V_i(\beta_0)$.

The arguments above are identical to those used to show the properties of the weighted log rank test. It can also be shown that $S(\beta_0)/\{\Omega(\beta_0)\}^{1/2}$ converges to a standard normal distribution. This result can be proved formally (Tsiatis, 1990) by showing that, under (2.4), $S(u, \beta_0) = \sum_{\{i: V_i(\beta_0) \leq u\}} \Delta_i A_i(\beta_0)$ is a martingale process. Asymptotic normality follows by a martingale central limit theorem (Gill, 1980; Anderson et al., 1982). Therefore, the set of β 's such that $|S(\beta)/\{\Omega(\beta)\}^{1/2}|$ is less than 1.96 constitute an asymptotic 95% confidence set for β .

4. POINT ESTIMATION

Since the test statistic $S(\beta_0)$ has mean zero when $\beta = \beta_0$, we propose $S(\beta)$ as an estimating equation and define the point estimate, $\hat{\beta}$, to be the solution to $S(\hat{\beta}) = 0$. However, since $S(\beta)$ is a step function, we shall define $\hat{\beta}$ as the value for which $S(\beta)$ changes sign. Under regularity conditions similar to those of Tsiatis (1990), $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean zero and variance that can be consistently estimated.

If $S(\beta)$ were differentiable with respect to β , then asymptotic normality would hold by the general theory of differentiable estimating equations as given by Serfling (1980, § 7). In this case, by a Taylor series expansion, for values of β in an $n^{-1/2}$ neighbourhood of β_0 yields

$$n^{-1/2}S(\beta) = n^{-1/2}S(\beta_0) + n^{1/2}(\beta - \beta_0)K(\beta_0) + o_p(1), \quad (4.1)$$

where $K(\beta_0)$ is the limit of $E\{n^{-1}\partial S(\beta)/\partial\beta\}$ evaluated at $\beta = \beta_0$.

Tsiatis (1990) showed that, when $S(\beta)$ is not differentiable as in the current problem, result (4.1) still holds true with $K(\beta_0)$ being the limit of $\partial[n^{-1}E\{S(\beta)\}]/\partial\beta$ evaluated at $\beta = \beta_0$. Together (4.1) and the asymptotic normality of $S(\beta_0)$ imply that $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically normal with a variance that can be consistently estimated by $n^{-1}\{\Omega(\hat{\beta})\}/\hat{K}^2$, where \hat{K} is a consistent estimate of $K(\beta_0)$ given by a 'numerical derivative' of $n^{-1}S(\beta)$ with step size of order $n^{-1/2}$. Specifically,

$$\hat{K} = \frac{1}{2}n^{-1/2}\{S(\hat{\beta} + cn^{-1/2}) - S(\hat{\beta} - cn^{-1/2})\}/c.$$

Note that, by expanding $S(\hat{\beta} + cn^{-1/2})$ and $S(\hat{\beta} - cn^{-1/2})$ using (4.1), we calculate that \hat{K} converges in probability to $K(\beta_0)$.

Again, the proposed estimates are semiparametric as the arguments above do not involve the specific form of the underlying baseline hazard function, $\lambda_0(u)$, given by model (2.4). Subject to some regularity conditions we also have great flexibility in the choice of the weighting function $G_i(u, \beta)$.

The efficiency, however, of our estimate $\hat{\beta}$ will depend on the choice of scoring function $g[\bar{Z}_i\{\psi^{-1}(u, \bar{Z}_i, \beta)\}]$. Using the approach of Schoenfeld (1981), Gill (1980) and Tsiatis (1990), it can be shown that the optimal choice $g_{\text{opt}}[\bar{Z}_i\{\psi^{-1}(u, \bar{Z}_i, \beta)\}]$ is proportional to the derivative of the log hazard ratio under local alternatives. That is, it is proportional to

$$\partial \log \{\lambda_i(u, \beta) / \lambda_0(u)\} / \partial \beta,$$

evaluated at $\beta = \beta_0$, where $\lambda_i(u, \beta)$ is the limit as $h \rightarrow 0$ of

$$h^{-1} \text{pr} [u \leq V_i(\beta) < u + h, \Delta_i = 1 \mid V_i(\beta) \geq u, \bar{Z}_i\{\psi^{-1}(u, \bar{Z}_i, \beta)\}].$$

Note, by (2.4), $\lambda_i(u, \beta_0) = \lambda_0(u)$.

In Corollary A.1 of the Appendix, we derive g_{opt} and show that if $\lambda_0(u)$ is constant, then the optimal weighting function under the Cox-Oakes model (2.1) is $Z_i\{\psi^{-1}(u, \bar{Z}_i, \beta)\}$. That is, the optimal scoring function is to choose the value of the time dependent covariate at the real time corresponding to baseline time u .

The estimate that would result if we used the optimal weighting function can be shown to attain the semiparametric efficiency bound in the sense of Begun et al. (1983), by adopting essentially unchanged Ritov & Wellner's (1988) derivation in the context of censored linear regression. A semiparametric estimator is defined as an estimator which is asymptotically normal and unbiased for β_0 whatever be $\lambda_0(u)$. A consequence of semiparametric efficiency is that the estimator that uses the scoring function $Z_i\{\psi^{-1}(u, \bar{Z}_i, \beta)\}$ is the semiparametric estimator that has minimum asymptotic variance when the true baseline hazard $\lambda_0(u)$ is constant.

5. EXTENSION TO MULTIPLE PARAMETERS

Up to this point, we assumed that β is a single parameter. The above methods may be extended to the case where β is a $p \times 1$ vector of parameters as follows. If $\beta \in R^p$, we define

$$g\{\bar{Z}(t)\} = [g_1\{\bar{Z}(t)\}, \dots, g_p\{\bar{Z}(t)\}]'$$

to be a $p \times 1$ vector of weighting functions. For example, suppose we have multiple time-dependent covariates $Z_1(t), \dots, Z_p(t)$, and a transformation function

$$\psi\{\bar{Z}(t), \beta\} = \int_0^t \exp\{\beta_1 Z_1(x) + \dots + \beta_p Z_p(x)\} dx.$$

We might then choose $g_j\{\bar{Z}(t)\}$ to correspond to $Z_j(t)$, the value of the j th time dependent covariate evaluated at time t for $j = 1, \dots, p$.

In the multiparameter problem,

$$A_i(\beta) = G_i\{V_i(\beta), \beta\} - G^{\text{av}}\{V_i(\beta), \beta\},$$

where the $p \times 1$ vector $G_i\{V_i(\beta), \beta\}$ has the l th component equal to $g_l[\bar{Z}_i\{\psi^{-1}\{V_i(\beta), \bar{Z}_i, \beta\}\}]$ and $G^{\text{av}}\{V_i(\beta), \beta\}$ is as defined in § 3. Similarly, the covariance

matrix of $S(\beta_0)$ is estimated by the $p \times p$ matrix, $\Omega(\beta_0)$, where $\Omega(\beta)$ is defined as in (3.2) except with

$$[G_j\{V_i(\beta), \beta\} - G^{av}\{V_i(\beta), \beta\}][G_j\{V_i(\beta), \beta\} - G^{av}\{V_i(\beta), \beta\}]'$$

replacing $[G_j\{V_i(\beta), \beta\} - G^{av}\{V_i(\beta), \beta\}]^2$.

We define $\hat{\beta}$ to be the solution to $S(\hat{\beta}) = 0$. However, since each component $S_j(\beta)$ of $S(\beta)$ is a step function in β , we define $\hat{\beta}$ to be the value of β which minimizes the quadratic form $S(\hat{\beta})'S(\hat{\beta})$.

Let c_1, \dots, c_p be fixed constants and e_k be the $p \times 1$ vector with k th element 1 and the remaining elements zero. Also define $\hat{K}(\hat{\beta})$ to be the $p \times p$ matrix of numerical partial derivatives with (j, k) th element

$$\frac{1}{2}n^{-1}\{S_j(\hat{\beta} + c_k e_k n^{-1/2}) - S_j(\hat{\beta} - c_k e_k n^{-1/2})\} / c_k,$$

which is the analogue of $\hat{K}(\hat{\beta})$ in § 4. Then, following Tsiatis (1990), under the suitable regularity conditions $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean zero and covariance matrix that can be consistently estimated by $n^{-1}\{\hat{K}^{-1}(\hat{\beta})\Omega(\hat{\beta})\{\hat{K}^{-1}(\hat{\beta})\}'$.

ACKNOWLEDGEMENT

This work was supported in part by grants from the National Institute of Environmental Health Sciences, the National Institutes of Health, the National Cancer Institute, and the National Institute of Allergic and Infectious Diseases.

APPENDIX

Derivation of the optimal weighting function

In this Appendix, we prove that the optimal weighting function under the Cox-Oakes model (2.1) is $\bar{Z}_i\{\psi^{-1}(u, \bar{Z}_i, \beta)\}$ when $\lambda_0(u)$ is constant.

PROPOSITION A.1. *The evaluation of the derivative of the log hazard rate is given by*

$$\frac{\partial \log \{\lambda_i(u, \beta_0)\}}{\partial \beta} = - \left[\frac{\partial \log \{\lambda_0(u)\}}{\partial u} \frac{\partial \psi_i(W_i, \beta_0)}{\partial \beta} + \left\{ \frac{\partial^2 \psi_i(W_i, \beta_0)}{\partial W_i \partial \beta} \right\} / \left\{ \frac{\partial \psi_i(W_i, \beta_0)}{\partial W_i} \right\} \right], \quad (A.1)$$

where $W_i \equiv \psi^{-1}(u, \bar{Z}_i, \beta_0)$ and $\psi_i(u, \beta) \equiv \psi(\bar{Z}_i(u), \beta)$ and the dependence of W_i on u has been suppressed.

COROLLARY A.1. *If $\lambda_0(u)$ is constant, and $h\{\bar{Z}_i(u), \beta\} = e^{\beta Z(u)}$, then*

$$\partial \log \{\lambda_i(u, \beta_0)\} / \partial \beta = -Z_i(W_i).$$

Proof of Corollary A.1. Corollary (A.1) follows from Proposition (A.1) by noting

$$\partial \psi_i(W_i, \beta_0) / \partial W_i = \exp \{\beta_0 Z_i(W_i)\},$$

$$\frac{\partial \psi_i(W_i, \beta_0)}{\partial \beta} = \int_0^{W_i} Z_i(x) \exp \{\beta_0 Z_i(x)\} dx, \quad \frac{\partial^2 \psi_i(W_i, \beta_0)}{\partial W_i \partial \beta} = Z_i(W_i) \exp \{\beta_0 Z_i(W_i)\},$$

and thus,

$$\frac{\partial \log \{\lambda_i(u, \beta_0)\}}{\partial \beta} = - \frac{\lambda'_0(u)}{\lambda_0(u)} \int_0^{W_i} Z_i(x) \exp \{\beta_0 Z_i(x)\} dx - Z_i(W_i). \quad \square$$

Proof of Proposition A.1. Define

$$\psi_i^{-1}(u, \beta) \equiv \psi^{-1}(u, \bar{Z}_i, \beta), \quad b_i(u, \beta) = \psi_i\{\psi_i^{-1}(u, \beta), \beta_0\}.$$

Since $b_i(U_i(\beta), \beta) = U_i(\beta_0)$ and $\lambda_i(u, \beta)$ is a cause-specific hazard for the random-variable $U_i(\beta)$, we have by the transformation formula for hazard functions

$$\lambda_i(u, \beta) = \lambda_0\{b_i(u, \beta)\} \frac{\partial b_i(u, \beta)}{\partial u}.$$

Therefore, by the chain rule,

$$\frac{\partial \lambda_i(u, \beta_0)}{\partial \beta} = \lambda_0'\{b_i(u, \beta_0)\} \frac{\partial b_i(u, \beta_0)}{\partial \beta} \frac{\partial b_i(u, \beta_0)}{\partial u} + \lambda_0\{b_i(u, \beta_0)\} \frac{\partial^2 b_i(u, \beta_0)}{\partial u \partial \beta}.$$

But $b_i(u, \beta_0) = u$ and, thus, $\partial b_i(u, \beta_0)/\partial u = 1$. Therefore,

$$\frac{\partial \lambda_i(u, \beta_0)}{\partial \beta} = \lambda_0'(u) \frac{\partial b_i(u, \beta_0)}{\partial \beta} + \lambda_0(u) \frac{\partial^2 b_i(u, \beta_0)}{\partial u \partial \beta}. \quad (\text{A}\cdot 2)$$

Upon comparing equation (A·2) to the right-hand side of (A·1), we see that the following lemmas will complete the proof of Proposition (A·1).

LEMMA A·1. *The derivative $\partial b_i(u, \beta_0)/\partial \beta$ equals $-\partial \psi_i(W_i, \beta_0)/\partial \beta$.*

LEMMA A·2. *The second derivative $\partial^2 b_i(u, \beta_0)/\partial u \partial \beta$ equals*

$$-\left\{ \frac{\partial^2 \psi_i(W_i, \beta_0)}{\partial W_i \partial \beta} \right\} / \left\{ \frac{\partial \psi_i(W_i, \beta)}{\partial W_i} \right\}.$$

Proof of Lemma A·1. Differentiating the identity $\psi_i\{\psi_i^{-1}(u, \beta), \beta\} = u$ we obtain

$$\begin{aligned} 0 &= \left[\frac{\partial}{\partial \beta} \psi_i\{\psi_i^{-1}(u, \beta), \beta\} \right]_{\beta=\beta_0} = \left[\frac{\partial}{\partial \beta} \psi_i\{\psi_i^{-1}(u, \beta), \beta_0\} \right]_{\beta=\beta_0} + \left[\frac{\partial}{\partial \beta} \psi_i\{\psi_i^{-1}(u, \beta_0), \beta\} \right]_{\beta=\beta_0} \\ &= \frac{\partial b_i(u, \beta_0)}{\partial \beta} + \frac{\partial \psi_i(W_i, \beta_0)}{\partial \beta}, \end{aligned}$$

which completes the proof. □

Proof of Lemma A·2. By Lemma (A·1),

$$\frac{\partial^2 b_i(u, \beta_0)}{\partial u \partial \beta} = -\frac{\partial^2 \psi_i(W_i, \beta_0)}{\partial u \partial \beta} = -\frac{\partial^2 \psi_i(W_i, \beta_0)}{\partial W_i \partial \beta} \frac{\partial W_i}{\partial u}.$$

But since $W_i \equiv \psi_i^{-1}(u, \beta_0)$,

$$\frac{\partial W_i}{\partial u} = \left(\frac{\partial u}{\partial W_i} \right)^{-1} = \left\{ \frac{\partial \psi_i(W_i, \beta_0)}{\partial W_i} \right\}^{-1},$$

completing the proof. □

REFERENCES

- ANDERSON, P. K., BORGAN, O., GILL, R. & KEIDING, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data. *Int. Statist. Rev.* **50**, 219-58.
- BEGUN, J. M., HALL, W. J., HUANG, W. M. & WELLNER, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432-52.
- COX, D. R. & OAKES, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- GILL, E. (1980). *Censoring and Stochastic Integrals*, Mathematical Center Tract 124. Amsterdam: Mathematische Centrum.
- KALBFLEISCH, J. D. & PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- RITOV, Y. & WELLNER, J. A. (1988). Censoring martingales and the Cox model. *Contemporary Mathematics* **80**, 191-220.

- ROBINS, J. M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In *Health Service Research Methodology: A Focus on AIDS*, Ed. L. Sechrest, H. Freeman and A. Mulley, pp. 113-59. Washington, D.C.: U.S. Department of Health and Human Services, U.S. Government Printing Office.
- SCHOENFELD, D. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika* **68**, 316-9.
- SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
- TSIATIS, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354-72.

[Received April 1990. Revised July 1991]