The Annals of Statistics
1981, Vol. 9, No. 1, 93–108

A LARGE SAMPLE STUDY OF COX'S REGRESSION MODEL[1]

BY ANASTASIOS A. TSIATIS

St. Jude Children's Research Hospital

Strong consistency and asymptotic normality are established for the maximum partial likelihood estimate of the regression parameter in Cox's regression model.

Estimates are also derived for the underlying cumulative hazard function and survival distribution. We establish the asymptotic normality of these estimates and calculate the limiting variances.

1. Introduction. Regression models for survival analysis with censored observations have been used quite extensively in the past few years. One of the more widely used models is the one suggested by Cox (1972), which assumes

$$\lambda(t \mid z) = \lambda_0(t)\exp(\boldsymbol{\beta}'\mathbf{z}), \tag{1.1}$$

where $\lambda(t \mid z)$ is the hazard function for an individual with regressor variables $\mathbf{z}' = (z_1, \cdots, z_p)$, regression parameters $\boldsymbol{\beta}' = (\beta_1, \cdots, \beta_p)$, and $\lambda_0(t)$ the underlying hazard function.

Estimates for the regression parameters were derived by Cox (1972). A more detailed justification was given by Cox (1975) under the term partial likelihood.

In this paper, we establish the asymptotic consistency and normality of the maximum partial likelihood estimates by the use of weak convergence results.

Estimates for the underlying survival distribution

$$S_0(t) = \exp(-\Lambda_0(t)); \qquad \Lambda_0(t) = \int_0^t \lambda_0(x)\, dx,$$

are also considered. Survivor function estimators have been suggested by Cox (1972), Kalbfleisch and Prentice (1973), Breslow (1974) and Prentice and Gloeckler (1978) for grouped data. However, the large sample properties of these estimates have not been examined sufficiently. Using weak convergence results, we are able to show that the survivor function estimate converges to a normal process.

The assumptions of the model and some key relationships are given in Section 2. In Section 3 the large sample properties of Cox's maximum partial likelihood estimate, $\hat{\beta}$, are established including strong consistency of $\hat{\beta}$ and asymptotic normality of $\sqrt{n}(\hat{\beta} - \beta)$. An estimate for the cumulative hazard function, $\Lambda_0(t) = \int_0^t \lambda_0(x)\, dx$, is derived in Section 4. The large sample properties of the cumulative hazard estimates including weak convergence to a mean zero Gaussian process are established in Section 5. Finally, in Section 6 we extend the results of the previous sections to include the asymptotic distribution of the estimate for survival distribution.

2. Notation and formulae. Let the covariate Z, be a random variable with density $f(z)$. The variable Z will be single valued but in Section 6 we shall indicate how to extend the results to a vector valued set of covariates. Denote the true survival time and the time to censoring by the positive random variables Y_1, Y_2 respectively. It is assumed that Y_1, Y_2 are conditionally independent given the covariate Z. The observable time until death or censoring

Received May 1978; revised November 1979.

[1] This research was conducted while the author was at the University of Wisconsin, Madison. This investigation was supported in part by Grant Number 1 R01 CA 18332 awarded by the National Cancer Institute, DHEW.

AMS 1970 *subject classifications.* Primary 62E20; secondary 62P10.

Key words and phrases. Cox's regression model, censored survival study, weak convergence, cumulative hazard function, Gaussian process.

will be denoted by the random variable Δ. That is

$$
\begin{aligned}
T &= \min\,(Y_1, Y_2)\\
\Delta &= 1 \quad \text{if} \quad Y_1 \le Y_2 \text{ (death)}\\
&= 0 \qquad\quad Y_1 > Y_2 \text{ (censoring)}.
\end{aligned}
$$

The distribution of the survival time Y_1 is related to the covariate Z according to Cox's regression model. That is

$$\lambda(t \mid z) = \lambda_0(t)\exp(\beta z),$$

where $\lambda(t \mid z)$ denotes the hazard function given that $Z = z$. No specific relationship for the conditional distribution of Y_2 given $Z = z$ will be made, except that Y_2 is a positive random variable whose conditional hazard function, given that $Z = z$, exists and is denoted by $\mu(t, z)$. Since most survival studies are ended after some prespecified time, T_0, we also assume that Y_2 is bounded $Y_2 \le T_0 < \infty$.

Notation and relationships which are useful throughout the remainder of the paper will be established.

The conditional probability of surviving until time t without being censored, given that $Z = z$, is given by

$$H(t \mid z) = P(T \ge t \mid Z = z) = \exp - \int_0^t [\lambda_0(x)\exp(\beta z) + \mu(x, z)]\, dx. \tag{2.1}$$

The probability of surviving until time t without being censored and eventually dying before being censored is

$$Q(t) = P(T \ge t, \Delta = 1) = \int Q(t \mid z) f(z)\, dz, \tag{2.2}$$

where

$$Q(t \mid z) = P(T \ge t, \Delta = 1 \mid Z = z) = \int_t^{T_0} \lambda_0(x)\exp(\beta z) H(x \mid z)\, dx. \tag{2.3}$$

The derivative of $Q(t)$ is given by

$$dQ(t)/dt = -\lambda_0(t) \int \exp(\beta z) H(t \mid z) f(z)\, dz. \tag{2.4}$$

Letting I_A denote the indicator function of the event A, and letting $g(z)$ be a continuous function of z, define

$$E(g(z), t) = E\{g(Z) I_{[T \ge t]}\} = E[g(Z) P(T \ge t \mid Z)] = \int g(z) H(t \mid z) f(z)\, dz, \tag{2.5}$$

and

$$E_1(g(z), t) = E\{g(Z) I_{[T \ge t, \Delta = 1]}\} = E[g(Z) Q(t \mid Z)]. \tag{2.6}$$

The derivative of $E_1(g(z), t)$ is equal to

$$dE_1(g(z), t)/dt = -\lambda_0(t) \int g(z)\exp(\beta z) H(t \mid z) f(z)\, dz. \tag{2.7}$$

Using (2.4)–(2.6) we get

$$\lambda_0(t) = -(dQ(t)/dt)/E(\exp(\beta z), t), \tag{2.8}$$

and

$$dE_1(g(z), t)/dt = dQ(t)/dt E(g(z)\exp(\beta z), t)/E(\exp(\beta z), t). \tag{2.9}$$

Let n be the number of individuals in the study. Associated with each individual is the random vector (T_i, Δ_i, Z_i), $i = 1, \cdots, n$ which are assumed to be independent and identically distributed random vectors. The empirical estimates of $Q(x)$, and $E(\exp(\beta z), x)$ are denoted by $\hat{Q}(x)$, $\hat{E}(\exp(\beta z), x)$; $\hat{\beta}$ is the estimate for the regression variables suggested by Cox (1972). That is

$$\hat{Q}(x) = \sum_{i=1}^{n} I_{[T_i \geq x, \Delta_i = 1]}/n \tag{2.10}$$

$$\hat{E}(\exp(\beta z), x) = \sum_{i=1}^{n} \exp(\beta Z_i) I_{[T_i \geq x]}/n = \sum_{j \in R(x)} \exp(\beta Z_j)/n,$$

where $R(x)$ denotes the risk set at time x, or the set of indices $i = 1, \cdots, n$ corresponding to individuals who survived until time x.

3. Asymptotic properties of Cox's estimate for β. Cox (1972), suggested the estimation of the regression parameter by maximizing the partial likelihood

$$L(\beta) = \prod_{i \in D} \{\exp(\beta Z_i)/(\sum_{j \in R(t_i)} \exp(\beta Z_j)/n\},$$

or log partial likelihood

$$l(\beta) = \beta \sum_{i \in D} Z_i/n - \sum_{i \in D} 1/n \log(\sum_{j \in R(t_i)} \exp(\beta Z_j)/n), \tag{3.1}$$

where D denotes the set of indices $i = 1, \cdots, n$ corresponding to individuals who died. The estimate $\hat{\beta}$ is the solution to the likelihood equation

$$\sum_{i \in D} Z_i/n - \sum_{i \in D} 1/n(\sum_{j \in R(t_i)} Z_j \exp(\beta Z_j)/n)/(\sum_{j \in R(t_i)} \exp(\beta Z_j)/n). \tag{3.2}$$

In this section we shall prove the strong consistency of $\hat{\beta}$ and asymptotic normality of $\sqrt{n}(\hat{\beta} - \beta)$. It will be assumed that the covariate Z satisfies the following assumption:

ASSUMPTION 3.1. $E[Z \exp(\beta Z)]^2$ is bounded uniformly in a neighborhood of β.

We shall also assume that T_0, which denotes the time that the study is terminated is such that $P(T \geq T_0) > 0$. This assumption implies that at the end of the study there is a positive chance that an individual will have survived without being censored. Although this last assumption is not necessary to prove the desired results, we feel justified that such an assumption is true in most survival studies. Later we shall indicate how this assumption may be weakened.

In proving the large sample results, it will be convenient to characterize the log partial likelihood (3.1) as

$$\hat{H}(\beta) = \beta \hat{E}_1(z, 0) - \int_0^{T_0} - d\hat{Q}(x) \log \hat{E}(e^{\beta z}, x), \tag{3.3}$$

and the likelihood equation (3.2) as

$$F(\beta) = \frac{d\hat{H}(\beta)}{d\beta} = \hat{E}_1(z, 0) - \int_0^{T_0} - d\hat{Q}(x)\hat{E}(ze^{\beta z}, x)/\hat{E}(e^{\beta z}, x) = 0, \tag{3.4}$$

where $\hat{E}_1(z, 0)$ is the empirical estimate of $E_1(z, 0)$. The quantities $E_1(z, 0)$, $Q(x)$, $E(e^{\beta z}, x)$ and their estimates are defined in Section 2.

The consistency of $\hat{\beta}$ is shown in the following theorem.

THEOREM 3.1. *There exists a sequence of solutions $\hat{\beta}(n)$ of equation* (3.2) *such that $\hat{\beta}(n)$ converges almost surely to β.*

PROOF. Consider the function

$$H(\beta) = \beta E_1(z, 0) - \int_0^{T_0} - dQ(x) \log E(e^{\beta z}, x).$$

Using (2.9) we show that the first derivative

$$H'(\beta) = E_1(z, 0) - \int_0^{T_0} - dQ(x)E(ze^{\beta z}, x)/E(e^{\beta z}, x) = 0.$$

The second derivative

$$H''(\beta) = \int_0^{T_0} - dQ[E(z^2e^{\beta z}, x)/E(e^{\beta z}, x) - (E(ze^{\beta z}, x)/E(e^{\beta z}, x))^2]$$

$$= -\int_0^{T_0} - dQ[E((z- E(z\,|\,R(x))^2e^{\beta z}, x))/E(e^{\beta z}, x)] < 0$$

where

$$E(z\,|\,R(x)) = E(ze^{\beta z}, x)/E(e^{\beta z}, x).$$

Therefore, the function $H(x)$ has a local maximum at $x = \beta$. This implies for β^* in a δ neighborhood of $\beta(|\beta^* - \beta| \leq \delta)$ we get

$$H(\beta) - H(\beta^*) \geq 0,$$

with strict inequality when $|\beta^* - \beta| = \delta$.

The use of Lemmas A.1 and A.2 of Appendix 1, together with the strong law of large numbers implies that

$$\hat{H}(\beta) - \hat{H}(\beta^*) \rightarrow H(\beta) - H(\beta^*) \text{ a.s.}$$

(The notation $\rightarrow$ a.s. means converges almost surely). Therefore for almost all realizations there exists an n_0 (depending on the realization) such that for all $n \geq n_0$

$$\hat{H}(\beta^*) < \hat{H}(\beta) \quad \text{for} \quad |\beta^* - \beta| = \delta.$$

The function $\hat{H}(x)$ is continuous and differentiable, therefore on the set $|\beta^* - \beta| \leq \delta$, $\hat{H}(\beta^*)$ has a maximum which is not on the boundary. This implies that we have local maximum and the first derivative vanishes. That is

$$\frac{d\hat{H}(\beta)}{d\beta} = 0,$$

which is precisely the solution to the likelihood equation (3.4). We can repeat this argument for balls of size δ which get smaller and by this means find a consistent sequence $\hat{\beta}(n) \rightarrow \beta$ a.s.

To show that the statistic $\sqrt{n}\ (\hat{\beta} - \beta)$ converges to a normal distribution, we will approximate it by a sum of random variables which converge jointly to a multivariate normal. Using a Taylor series expansion we get

$$F(\hat{\beta}) = F(\beta) - (\hat{\beta} - \beta) \textstyle\sum_{i\in D} n^{-1}\hat{\mathrm{V}}\mathrm{ar}(z|\,R(t_i))\,|\beta$$
$$+ (\hat{\beta} - \beta)^2/2 \textstyle\sum_{i\in D} n^{-1}\hat{E}((z - \hat{E}(z|\,R(t_i)))^3|\,R(t_i))\,|\beta^*,$$

β^* lies between β and $\hat{\beta}$, where using the notation from Efron (1977),

$$\hat{E}(h(z)|\,R(t_i))\,|\beta^* = \textstyle\sum_{j\in R(t_i)} h(Z_j)\exp(\beta^* Z_j)/\sum_{R(t_i)} \exp(\beta^* Z_j),$$

and

$$\hat{\mathrm{V}}\mathrm{ar}(z|\,R(t_i))\,|\beta^* = \textstyle\sum_{R(t_i)} Z_j^2\exp(\beta^* Z_j)/\sum_{R(t_i)} \exp(\beta^* Z_j)$$
$$- (\textstyle\sum_{R(t_i)} Z_j\exp(\beta^* Z_j)/\sum_{R(t_i)} \exp(\beta^* Z_j))^2.$$

Therefore the statistic $\sqrt{n}(\hat{\beta} - \beta)$ can be expressed as

$$\sqrt{n}(\hat{\beta} - \beta) = \left[C_n \Big/ \int_0^{T_0} - dQ\mathrm{Var}(z \mid R(t)) \right] - F_n \tag{3.5}$$

where

$$C_n = \sqrt{n}\left[\hat{E}_1(z, 0) - \int_0^{T_0} - d\hat{Q}\hat{E}(z \mid R(t)) \right],$$

$$F_n = C_n F_{2_n}/F_{3_n},$$

$$F_{2_n} = \int_0^{T_0} - d\hat{Q}\hat{\mathrm{V}}\mathrm{ar}(z \mid R(t)) - \int_0^{T_0} - dQ\mathrm{Var}(z \mid R(t)) - R_n(\hat{\beta} - \beta)/2, \tag{3.6}$$

$$F_{3_n} = \left[\int_0^{T_0} - d\hat{Q}\hat{\mathrm{V}}\mathrm{ar}(z \mid R(t)) - (R_n)(\hat{\beta} - \beta)/2 \right] \int_0^{T_0} - dQ\mathrm{Var}(z \mid R(t)),$$

$$R_n = \textstyle\sum_{i \in D} n^{-1}\hat{E}((z - \hat{E}(z \mid R(t_i)))^3 \mid R(t_i)) \mid \beta^*.$$

THEOREM 3.2. *The statistic $\sqrt{n}(\hat{\beta} - \beta)$ converges in distribution to a normal random variable with mean* 0 *and variance equal to* $[\int_0^{T_0} - dQ\mathrm{Var}(z \mid R(t))]^{-1}$.

PROOF. The proof will consist of (i) showing that the statistic C_n of (3.6) converges in distribution to $N(0, \int_0^{T_0} - dQ\mathrm{Var}(z \mid R(t)))$, and (ii) that F_n converges in probability to zero.

(i) Using (2.9) the statistic C_n can be written as

$$C_n = \sqrt{n}[\hat{E}_1(z, 0) - E_1(z, 0)] - \sqrt{n}\left[\int_0^{T_0} - d\hat{Q}\hat{E}_z/\hat{E} - \int_0^{T_0} - dQE_z/E \right],$$

where $E_z(t) = E(z \exp(\beta z), t)$ and $E(t) = E(\exp(\beta z), t)$. Using techniques similar to Breslow-Crowley (1974) C_n can be expressed as

$$C_n = C_{1n} + \cdots + C_{6n} + R_{1n} + \cdots + R_{4n}$$

where

$$C_{1n} = \sqrt{n}[\hat{E}_1(z, 0) - E_1(z, 0)], \qquad C_{2n} = -\sqrt{n}[\hat{Q}(0) - Q(0)]E_z(0)/E(0),$$

$$C_{3n} = -\int_0^{T_0} [\sqrt{n}(\hat{Q} - Q)/E]\, dE_z, \qquad C_{4n} = \int_0^{T_0} [\sqrt{n}(\hat{Q} - Q)E_z/E^2]\, dE \tag{3.7}$$

$$C_{5n} = \int_0^{T_0} [\sqrt{n}(\hat{E}_z - E_z)/E]\, dQ, \qquad C_{6n} = -\int_0^{T_0} [\sqrt{n}(\hat{E} - E)E_z/E^2]\, dQ,$$

and

$$R_{1n} = \int_0^{T_0} [\sqrt{n}(\hat{E}_z - E_z)/\hat{E}]d(\hat{Q} - Q), \qquad R_{2n} = -\int_0^{T_0} [\sqrt{n}(\hat{E}_z - E_z)(\hat{E} - E)/\hat{E}E]\, dQ,$$

(3.8)

$$R_{3n} = -\int_0^{T_0} [\sqrt{n}(\hat{E} - E)E_z/\hat{E}E]\, d(\hat{Q} - Q), \quad R_{4n} = \int_0^{T_0} [\sqrt{n}(\hat{E} - E)^2E_z/\hat{E}E^2]\, dQ.$$

In Theorem 5.1 we shall show that $(C_{1n}, \cdots, C_{6n})$ will converge in distribution to a multivariate normal random variable which is denoted by $(C_1, \cdots, C_6)$. It will also be shown that $R_{1n}, \cdots, R_{4n}$ converge in probability to zero. Therefore the statistic $C_n = \sum_{i=1}^6 C_{in} + \sum_{i=1}^4 R_{in}$ will converge in distribution to the normal random variable $\sum_{i=1}^6 C_i$ whose mean is zero and variance (which is calculated in Appendix 2) is equal to $\int_0^{T_0} - dQ\mathrm{Var}(z \mid R(t))$.

(ii). As a consequence of Lemmas A.1, A.2, and Assumption 3.1, we can show

$$\int_0^{T_0} -d\hat{Q}\hat{\mathrm{V}}\mathrm{ar}(z \mid R(t)) - \int_0^{T_0} -dQ\mathrm{Var}(z \mid R(t)) \to 0 \text{ a.s.}$$

This, together with the consistency of $\hat{\beta}$, implies that $F_{2n} \to 0$ a.s. Similarly, F_{3n} converges almost surely to $[\int_0^{T_0} -dQ\mathrm{Var}(z \mid R(t))]^2$. Therefore by applying Slutsky's theorem we prove that F_n converges to 0 in probability.

REMARK. The asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta)$, denoted by $\sigma_{\hat{\beta}}^2$

$$\sigma_{\hat{\beta}}^2 = \left[\int_0^{T_0} -dQ(t)\mathrm{Var}(z \mid R(t))\right]^{-1} \tag{3.9}$$

can be estimated by substituting the empirical estimates into (3.9). That is

$$\hat{\sigma}_{\hat{\beta}}^2 = \left[\int_0^{T_0} -d\hat{Q}\hat{\mathrm{V}}\mathrm{ar}(z \mid R(t)) \mid \hat{\beta}\right]^{-1} = \left[\textstyle\sum_{i\in D} \frac{1}{n}\hat{\mathrm{V}}\mathrm{ar}(z \mid R(t_i)) \mid \hat{\beta}\right]^{-1}. \tag{3.10}$$

Formula (3.10) is minus the inverse of the second derivate of Cox's partial likelihood.

4. Estimate for the cumulative hazard function. The cumulative hazard function is defined as

$$\Lambda_0(t) = \int_0^t \lambda_0(x)\,dx.$$

By using equation (2.8) we can express the cumulative hazard function as

$$\Lambda_0(t) = \int_0^t -dQ(x)/E(e^{\beta z}, x). \tag{4.1}$$

An intuitive estimate of $\Lambda_0(t)$ would be

$$\hat{\Lambda}_0(t) = \int_0^t -d\hat{Q}(x)/\hat{E}(e^{\hat{\beta}z}, x).$$

That is

$$\hat{\Lambda}_0(t) = \textstyle\sum_{i\in D(t)} \frac{1}{n}\Big/ \sum_{j\in R(t_i)} \exp(\hat{\beta}Z_j)/n = \sum_{i\in D(t)} 1/\sum_{j\in R(t_i)} \exp(\hat{\beta}Z_j), \tag{4.2}$$

where $D(t)$ denotes the set of indices $i = 1, \cdots, n$ corresponding to individuals who died before time t. The estimate (4.2) is the nonparametric MLE under $\beta = \hat{\beta}$ as developed by Breslow (1974).

We shall examine the large sample properties of the statistic $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$ by approximating it with a sum of stochastic integrals which will be shown to converge to a joint normal process on the interval $[0, M]$. The value M, $M \le T_0$, is such that $P(T \ge M)$ is strictly positive.

Expanding $\hat{\Lambda}_0(t)$ about β in (4.2) using a Taylor series expansion we get

$$\begin{aligned}\Lambda_0(t) &= \textstyle\sum_{i\in D(t)} (1/\sum_{R(t_i)} \exp(\beta Z_j)) \\ &\quad - (\hat{\beta} - \beta) \textstyle\sum_{i\in D(t)} [\sum_{R(t_i)} Z_j\exp(\beta Z_j)/(\sum_{R(t_i)} \exp(\beta Z_j))^2] + (\hat{\beta} - \beta)^2 K(t)/2,\end{aligned} \tag{4.3}$$

where

$$\begin{aligned}K(t) = \textstyle\sum_{i\in D(t)} [&-\textstyle\sum_{R(t_i)} Z_j^2\exp(\beta^* Z_j)/(\sum_{R(t_i)} \exp(\beta^* Z_j))^2 \\ &+ 2(\textstyle\sum_{R(t_i)} Z_j\exp(\beta^* Z_j))^2/(\sum_{R(t_i)} \exp(\beta^* Z_j))^3],\end{aligned}$$

$$= \int_0^t d\hat{Q}[\hat{E}(z^2 e^{\beta^* z}, x)/\hat{E}^2(e^{\beta^* z}, x) - 2\hat{E}^2(ze^{\beta^* z}, x)/\hat{E}^3(e^{\beta^* z}, x)],$$

and β^* lies between β and $\hat{\beta}$. By the use of (4.1) and (4.3) the statistic $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$ can be expressed as

$$\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t)) = A_n(t) + B_n(t) + E_n(t), \tag{4.4}$$

where

$$A_n(t) = \sqrt{n}\left[\int_0^t - d\hat{Q}(x)/\hat{E}(\exp(\beta z), x) - \int_0^t - dQ(x)/E(\exp(\beta z), x)\right],$$

$$B_n(t) = -\sqrt{n}(\hat{\beta} - \beta)\int_0^t - d\hat{Q}(x)\hat{E}(z\exp(\beta z), x)/\hat{E}^2(\exp(\beta z), x),$$

and

$$E_n(t) = \sqrt{n}(\hat{\beta} - \beta)^2 K(t)/2.$$

LEMMA 4.1. *$\sup_{0\le t\le M} E_n(t)$ converges in probability to zero* ($\to_p$).

PROOF. See Appendix 1.

In order to determine the asymptotic distribution of $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$ it will suffice to look at the joint asymptotic distribution of $A_n(t)$ and $B_n(t)$. Using (4.4) and (3.5) through (3.8) we get

$$B_n(t) = \left\{\left[(C_{1n} + \cdots + C_{6n} + R_{1n} + \cdots + R_{4n}) \Big/ \int_0^{T_0} - dQ\,\mathrm{Var}(z \mid R(t))\right] - F_n\right\} \times \left[\int_0^t - dQE_z/E^2 + R_{0n}(t)\right], \tag{4.5}$$

where

$$R_{0n}(t) = \left[\int_0^t - d\hat{Q}\hat{E}_z/\hat{E}^2 - \int_0^t - dQE_z/E^2\right].$$

As a direct consequence of Lemmas A.1, A.2, and Assumption 3.1 we get $\sup_{0\le t\le M} R_{0n}(t)$ converging almost surely to zero.

Similarly, the statistic $A_n(t)$ of (4.4) can be expressed as

$$A_n(t) = A_{1n}(t) + \cdots + A_{3n}(t) + E_{1n}(t) + E_{2n}(t), \tag{4.6}$$

where

$$\begin{aligned} A_{1n}(t) &= \int_0^t [\sqrt{n}(\hat{E} - E)^2/E^2]\, dQ, \qquad A_{2n}(t) = -\int_0^t [\sqrt{n}(\hat{Q} - Q)/E^2]\, dE, \\ A_{3n}(t) &= \sqrt{n}(\hat{Q}(0) - Q(0))/E(0) - \sqrt{n}(\hat{Q}(t) - Q(t))/E(t), \end{aligned} \tag{4.7}$$

and

$$E_{1n}(t) = \int_0^t [\sqrt{n}(\hat{E} - E)^2/E^2\hat{E}]\, dQ, \qquad E_{2n}(t) = -\int_0^t [\sqrt{n}(\hat{E} - E)/\hat{E}E]\, d(\hat{Q} - Q). \tag{4.8}$$

The representation of $A_n(t)$ and $B_n(t)$ given in (4.5)–(4.8) will facilitate the proof of weak convergence of $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$ to a normal process, as will be shown in Section 5.

5. Weak Convergence. Let $D[0, M]$ denote the space of functions defined on $[0, M]$ that are left continuous and have right-hand limits. Let $(X_n(t), Y_n(t), Z_n(t), W_n(t))$ be random functions defined on the product space $D^4[0, M]$, where

$$X_n(t) = \sqrt{n}(\hat{Q}(t) - Q(t)), \qquad Y_n(t) = \sqrt{n}(\hat{E}(t) - E(t)), \tag{5.1}$$
$$Z_n(t) = \sqrt{n}(\hat{E}_z(t) - E_z(t)), \qquad W_n(t) = \sqrt{n}(\hat{E}_1(z, t) - E_1(z, t)).$$

LEMMA 5.1. *If the random variable Z satisfies Assumption 3.1 then the sequence of distributions induced by (X_n, Y_n, Z_n, W_n) converges weakly to a joint Gaussian process (X, Y, Z, W) which has mean zero and covariance structure given by*

$$\begin{aligned}
\operatorname{Cov}(X(s), X(t)) &= Q(u) - Q(s)Q(t)\\
\operatorname{Cov}(Y(s), Y(t)) &= E(\exp(2\beta z), u) - E(s)E(t)\\
\operatorname{Cov}(Z(s), Z(t)) &= E(z^2\exp(2\beta z), u) - E_z(s)E_z(t)\\
\operatorname{Cov}(W(s), W(t)) &= E_1(z^2, u) - E_1(z, s)E_1(z, t)\\
\operatorname{Cov}(X(s), Y(t)) &= E_1(\exp(\beta z), u) - Q(s)E(t)\\
\operatorname{Cov}(X(s), Z(t)) &= E_1(z\exp(\beta z), u) - Q(s)E_z(t)\\
\operatorname{Cov}(X(s), W(t)) &= E_1(z, u) - Q(s)E_1(z, t)\\
\operatorname{Cov}(Y(s), Z(t)) &= E(z\exp(2\beta z), u) - E(s)E_z(t)\\
\operatorname{Cov}(Y(s), W(t)) &= E_1(z\exp(\beta z), u) - E(s)E_1(z, t)\\
\operatorname{Cov}(Z(s), W(t)) &= E_1(z^2\exp(\beta z), u) - E_z(s)E_1(z, t)
\end{aligned} \tag{5.2}$$

for all $0 \le s, t \le M$, $u = \max(s, t)$.

PROOF. The proof of Lemma 5.1 consists of applying the standard theory of weak convergence given in Billingsley (1968). We first show that the finite dimensional distributions converge to a multivariate normal, and then prove tightness.

The vector $(X_n(t), Y_n(t), Z_n(t), W_n(t))$ can be represented as

$$(X_n, Y_n, Z_n, W_n) = n^{-1/2}\{\textstyle\sum_{i=1}^{n} (I_{[T_i \ge t, \Delta_i = 1]} - Q(t), \exp(\beta Z_i)I_{[T_i \ge t]} - E(t),$$
$$Z_i\exp(\beta Z_i)I_{[T_i \ge t]} - E_z(t), Z_i I_{[T_i \ge t, \Delta_i = 1]} - E_1(z, t))\}$$

Boundedness of the second moments follows directly from Assumption 3.1. Therefore, a simple application of the multivariate central limit theorem implies that the finite dimensional distribution of (X_n, Y_n, Z_n, W_n) are multivariate normal with covariance structure (5.2). As shown in Lemma A.2, the sequence of distributions induced by Y_n, Z_n are tight. Similarly, we can show that the sequence of distributions induced by X_n, W_n are tight. Consequently, (X_n, Y_n, Z_n, W_n) induces a tight sequence on the product space $D^4[0, M]$.

Arguing as in Pyke-Shorack (1968), or Breslow-Crowley (1974) we can replace (X_n, Y_n, Z_n, W_n) and (X, Y, Z, W) with a sequence of random functions having the same distribution for each n, but which also satisfies

$$\rho((X_n, Y_n, Z_n, W_n), (X, Y, Z, W)) \to 0, \quad \text{a.s.}, \tag{5.3}$$

where ρ is the Skorohod metric on $D^4[0, M]$. Therefore, if we show convergence of any function of these replaced sequences, it will be equivalent to convergence in distribution of the same function of the original sequences.

THEOREM 5.1. *The random function $\sqrt{n}[\hat{\Lambda}_0(t) - \Lambda_0(t)]$, $0 \le t \le M$ converges weakly to a Gaussian process defined by*

$$V(t) = A_1(t) + \cdots + A_3(t)$$

(5.4)
$$+ (C_1 + \cdots + C_6)\left[\left(\int_0^t - dQE_z/E^2\right)\Big/\int_0^{T_0} - dQ\mathrm{Var}(z \mid R(t))\right]$$

where

$$A_1(t) = \int_0^t Y\,dQ/E^2, \qquad A_2(t) = -\int_0^t X\,dE/E^2, \qquad A_3(t) = [X(0)/E(0) - X(t)/E(t)],$$

and

$$C_1 = W(0), \qquad C_2 = -X(0)E_z(0)/E(0), \qquad C_3 = -\int_0^{T_0} X\,dE_z/E$$

$$C_4 = \int_0^{T_0} XE_z dE/E^2, \qquad C_5 = \int_0^{T_0} ZdQ/E, \qquad C_6 = -\int_0^{T_0} YE_z dQ/E^2,$$

and (X, Y, Z, W) *is a joint Gaussian process with mean* 0 *and covariance structure* (5.2).

$V(t)$ *is a Gaussian process with mean zero and covariance structure given by*

(5.5)
$$\begin{aligned}&\mathrm{Cov}(V(s), V(t))\\ &= \int_0^s - dQ/E^2 + \left(\int_0^s - dQE_z/E^2\right)\left(\int_0^t - dQE_z/E^2\right)\Big/\int_0^{T_0} - dQ\mathrm{Var}(z \mid R(t))\end{aligned}$$

where $0 \le s \le t \le M$.

Proof. Let ρ_M denote the supremum metric on $[0, M]$ and let ρ denote the Skorohod metric. The proof consists of showing

(i) $C_{in} \to C_i$, a.s. $\quad i = 1, \cdots, 6$
(ii) $R_{in} \to_p 0$, $\quad i = 1, \cdots, 4$
(iii) $\rho(A_{in}(t), A_i(t)) \to 0$, a.s. $\quad i = 1, \cdots, 3$
(iv) $\rho(E_{in}(t)) \to_p 0$, $\quad i = 1, 2$.

Convergence to a continuous limit in ρ_M is equivalent to convergence in ρ. Therefore, since $(A_i(t), i = 1, \cdots, 3)$ and 0 are continuous almost surely, then we can replace ρ by ρ_M in the last two conditions.

Upon realizing that $E(t)$ is a nonincreasing function on the interval $[0, M]$, and since $E(M) > 0$ follows from the fact that $P(T \ge M) > 0$, we can then prove conditions (iii) and (iv) exactly as Breslow-Crowley (1974), Theorem 4. Conditions (i) and (ii) can be proved similarly since by assumption $P(T \ge T_0) > 0$.

Remark. Conditions (i) and (ii) are true even if we allow the $P(T \ge T_0)$ to be equal to zero. Although the detailed arguments are very lengthy, the idea is to represent the integrals of the remainder terms in formula (3.8) as $\int_0^{T_0-\theta} + \int_{T_0-\theta}^{T_0}$. Applying the same arguments as in the proof of Theorem 5.1, we can show that the first part of the integral converges in probability to zero as long as θ is positive. Using techniques similar to Crowley and Thomas (1975) and Crowley (1973), we can find $\theta > 0$ so that the second integral is as small as we want with large probability uniformly in n. Therefore, the remainder terms converge in probability to zero.

The evaluation of the covariance structure is a straightforward calculation of the covariances for the additive terms. The details of the calculations are deferred until Appendix 2.

Some interesting observations that result from the calculations of Appendix 2 are as follows: letting

(5.6)
$$V(t) = V_1(t) + \phi(t)V_2$$

where

$$V_1(t) = A_1(t) + A_2(t) + A_3(t),$$

$$V_2 = (C_1 + \cdots + C_6) \Big/ \int_0^{T_0} - dQ \operatorname{Var}(z \mid R(t)),$$

$$\phi(t) = \int_0^t - dQE_z/E^2,$$

then

(i). $V_1(t)$ is a mean zero, independent increments Gaussian process with covariance structure given by $\operatorname{Cov}(V_1(s), V_1(t)) = \int_0^s - dQ/E^2$, $s \le t$;

(5.7) (ii). V_2 is a mean 0 normal random variable with variance equal to $[\int_0^{T_0} - dQ \operatorname{Var}(z \mid R(t))]^{-1}$;

(iii). The variable V_2 is independent of the process $V_1(t)$, $0 \le t \le M$.

6. Complementary results. In this section we shall extend some of the results from the previous sections and outline their proofs. In Lemma 6.1 we calculate the asymptotic distribution for the cumulative hazard function for an individual with covariate z_0.

LEMMA 6.1. *The estimate for the cumulative hazard function evaluated at z_0, $(\Lambda_0(t)\exp(\beta z_0))$, is given by $\hat\Lambda_0(t)\exp(\hat\beta z_0)$. The random function $\sqrt{n}[\hat\Lambda_0(t)\exp(\hat\beta z_0) - \Lambda_0(t)\exp(\beta z_0)]$ converges weakly to a Gaussian process $V_{z_0}(t)$ which has mean 0 and covariance structure*

$$\operatorname{Cov}(V_{z_0}(s), V_{z_0}(t)) = \exp(2\beta z_0)\Bigg[\int_0^s - dQ/E^2 + \Bigg\{\Bigg(\int_0^s - dQ[E(z \mid R(\cdot)) - z_0]/E\Bigg) \times \Bigg(\int_0^t - dQ[E(z \mid R(\cdot)) - z_0]/E\Bigg) \Bigg/ \int_0^{T_0} - dQ \operatorname{Var}(z \mid R(\cdot))\Bigg\}\Bigg], \quad 0 \le s \le t \le M.$$

PROOF. The proof consists of expanding $\hat\Lambda_0(t)\exp(\hat\beta z_0)$ about $\Lambda_0(t)$ and β in a Taylor series expansion and noting that the joint distribution of $\sqrt{n}(\hat\Lambda_0(t) - \Lambda_0(t))$, $\sqrt{n}(\hat\beta - \beta))$ converges asymptotically to $(V_1(t) + \phi(t)V_2, V_2)$ (see (5.6)).

LEMMA 6.2. *Let the survival probability at time t for an individual with covariate z_0 be denoted by $S(t \mid z_0)$. Then*

$$S(t \mid z_0) = \exp - [\Lambda_0(t)\exp(\beta z_0)].$$

The estimate for $S(t \mid z_0)$ is given by $\exp - [\hat\Lambda_0(t)\exp(\hat\beta z_0)]$. The random function $\sqrt{n}\{\exp - [\hat\Lambda_0(t)\exp(\hat\beta z_0)] - S(t \mid z_0)\}$ converges weakly to a Gaussian process $S_{z_0}(t)$ which has mean 0 and covariance structure

$$\operatorname{Cov}(S_{z_0}(s), S_{z_0}(t)) = S(t \mid z_0)S(s \mid z_0)\operatorname{Cov}(V_{z_0}(s), V_{z_0}(t)), \qquad 0 \le s \le t \le M.$$

PROOF. This proof is a simple application of the δ-method, Rao (1965), (see, for example, Breslow-Crowley (1974), Theorem 5).

The above results can be extended to vector valued covariates. Therefore, if $\mathbf{z}' = (z_1, \cdots, z_p)$, $\boldsymbol\beta' = (\beta_1, \cdots, \beta_p)$, and the relationship of the hazard function to the covariates is given by (1.1), then

THEOREM 6.1. *The estimate of the underlying cumulative hazard function $\Lambda_0(t)$ is given by*

$$\hat\Lambda_0(t) = \textstyle\sum_{i \in D(t)} 1/\sum_{R(t_i)} \exp(\boldsymbol\beta'\mathbf{z}).$$

The random function $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$ converges weakly to a Gaussian process $\tilde{V}(t)$ which has mean zero and covariance structure given by

$$\text{Cov}(\tilde{V}(s), \tilde{V}(t)) = \int_0^s - dQ/E^2(\exp(\beta'\mathbf{z}), x) + \psi'(s)\Sigma^{-1}(\beta)\psi(t), \tag{6.1}$$

where

$$\psi'(t) = (\psi_1(t), \cdots, \psi_p(t)),$$

$$\psi_i(t) = \int_0^t - dQE(z_i\exp(\beta'\mathbf{z}), x)/E^2(\exp(\beta'\mathbf{z}), x)),$$

$$\Sigma(\beta) = (\sigma_{ij}, i, j = 1, \cdots, p)$$

$$\sigma_{ij} = \int_0^{T_0} - dQ[E(z_iz_j\exp(2\beta'\mathbf{z}), x) - E(z_i\exp(\beta'\mathbf{z})E(z_j\exp(\beta'\mathbf{z}), x)],$$

and

$\hat{\beta}$ is the estimate of the regressor variables given by Cox,

$\Sigma^{-1}(\beta)$ is the variance-covariance matrix of $\sqrt{n}(\hat{\beta} - \beta)$.

REMARK. As a consequence of Lemmas A.1 and A.2 estimates of the variances and covariances can be derived by the substitution of the appropriate empirical estimates in the variance and covariance formulas.

For example, by the appropriate substitution into (6.1) the estimate of the asymptotic variance for the cumulative hazard function estimate would be

$$\begin{aligned} &\int_0^t - d\hat{Q}/\hat{E}^2(\exp(\hat{\beta}'\mathbf{z}, x) + \hat{\psi}'(t)\hat{\Sigma}^{-1}\hat{\psi}(t) \\ &\qquad = n\{\textstyle\sum_{i\in D(t)} 1/(\sum_{R(t_i)} \exp(\hat{\beta}'\mathbf{z}_j))^2\} + \hat{\psi}'(t)\hat{\Sigma}^{-1}\hat{\psi}(t) \end{aligned} \tag{6.2}$$

where $\hat{\psi}'(t) = (\hat{\psi}_1(t), \cdots, \hat{\psi}_p(t))$, $\hat{\psi}_l(t) = \sum_{i\in D(t)} [\sum_{j\in R(t_i)} z_{jl}\exp(\hat{\beta}\mathbf{z}_j)/(\sum_{R(t_i)} \exp(\hat{\beta}\mathbf{z}_j))^2]$ and $\hat{\Sigma}^{-1}$ is the estimate of the variance-covariance matrix of $\hat{\beta}$ as derived by Cox (1972).

We note that (6.2) is the same as the asymptotic variance derived by Tsiatis (1978).

APPENDIX 1

7. Appendix 1. Let $\hat{E}(g(z), t)$ be the empitical estimate of $E(g(z), t)$, so that

$$\hat{E}(g(z), t) = \textstyle\sum_{j\in R(t)} g(Z_j)/n.$$

LEMMA A.1. *Let Z be a random variable such that $E(g^2(Z))$ is finite, where $g(z)$ is a continuous function, then*

$$\sup_{0\le t\le M} |\hat{E}(g(z), t) - E(g(z), t)| \to 0, \quad \text{a.s.}$$

PROOF. Without any loss in generality we may assume that $g(z)$ is positive. This is because $g(z)$ may be written as $g^+(z) - g^-(z)$, where $g^+(z) = \max(g(z), 0)$, $g^-(z) = \max(-g(z), 0)$ are both positive. Since $E(g^2(Z))$ is finite, this implies that $E(g(z))$ is bounded. The functions $\hat{E}(g(z), t)$ and $E(g(z), t)$ are both monotone decreasing. $E(g(z), t)$ is bounded and $\hat{E}(g(z), t)$ is bounded a.s., therefore the desired result can be proved exactly as in the Glivenko-Cantelli lemma (see Loève (1963) page 28).

The following lemma will prove useful in establishing consistency of certain estimates which appear in the main body of the paper.

LEMMA *A*.2. *Let $\hat{X}_n(t)$, $\hat{Y}_n(t)$, $\hat{Q}_n(t)$ be random functions on the interval $[0, M]$ that converge almost surely in sup norm to the functions $X(t)$, $Y(t)$, $Q(t)$. $X(t)$, $Y(t)$ are continuous functions on*

[0, M]. $Q(t)$ is a continuous subsurvival function, that is $Q(t)$ is a positive decreasing function $Q(0) \le 1$. Let $f(x, y)$ be a continuous function from $R^2 \to R$ such that the partial derivatives $\partial f/\partial x$, $\partial f/\partial y$, exist and are continuous on $R_X \times R_y$, where $R_X \times R_y$ denote the range space of $X(t)$, $Y(t)$ respectively, $0 \le t \le M$. Then

$$\sup_{0\le t\le M} \left| \int_0^t f(\hat{X}_n, \hat{Y}_n)(-d\hat{Q}_n) - \int_0^t f(X, Y)(-dQ) \right| \to 0, \text{ a.s.}$$

PROOF. The proof of Lemma A.2 is given by Aalen (1976, Lemma 6.1).
Let $Z_n(t) = \sqrt{n}[\hat{E}(g(z), t) - E(g(z), t)]$.

LEMMA A.3. *If Z is a random variable such that $E(g^2(Z))$ is finite, then the sequence of distributions induced by the random functions $Z_n(t)$ are tight.*

PROOF. Using the standard theory of weak convergence (see Billingsley (1968) page 128) it suffices to show

$$E[(Z_n(t) - Z_n(t_1))^2(Z_n(t_2) - Z_n(t))^2] \le C_4(F(t_2) - F(t_1))^2 \tag{7.1}$$

where

$$F(t) = E(g^2(Z)I_{[T<t]}), \qquad 0 \le t_1 \le t \le t_2 \le T_0.$$

Arguing as in Billingsley (1968) page 106, we note that the left-hand side of (7.1) can be written as

$$n^{-2}E[(\textstyle\sum_{\iota=1}^n U_{1\iota}^2)(\sum_{\iota=1}^n U_{2\iota})^2, \tag{7.2}$$

where

$$U_{1\iota} = g(Z_\iota)I_{1\iota} - E(g(Z)I_1), \qquad I_{1\iota} = I_{[T_\iota\in[t_1,t)]},$$

$$U_{2\iota} = g(Z_\iota)I_{2i} - E(g(Z)I_2), \qquad I_{2\iota} = I_{[T_\iota\in[t,t_2)]}.$$

Due to the symmetry and independence of the U_i's, (7.2) can be expresed as

$$n^{-2}[nE(U_1^2U_2^2) + n(n-1)E(U_1^2)E(U_2^2) + 2n(n-1)E^2(U_1U_2)]. \tag{7.3}$$

We can express

$$U_\iota^2 = [g(Z) - E(g(Z)I_\iota)]^2I_\iota + E^2(g(Z)I_\iota)\bar{I}_\iota, \qquad i = 1, 2,$$

where $\bar{I}_\iota = 1 - I_\iota$, therefore

$$\begin{aligned} E(U_1^2U_2^2) = {} & E^2(g(Z)I_1)E[[g(Z) - E(g(Z)I_2)]^2I_2] \\ & + E^2(g(Z)I_2)E[[g(Z) - E(g(Z)I_1)]^2I_1] \\ & + E^2(g(Z)I_1)E^2(g(Z)I_2)E(\bar{I}_1\bar{I}_2). \end{aligned} \tag{7.4}$$

The following inequality

$$E^2(g(Z)I_\iota) \le E(g^2(Z)I_\iota), \qquad i = 1, 2, \tag{7.5}$$

follows from Schwartz's inequality. Using (7.4) and (7.5) we get that

$$E(U_1^2U_2^2) \le 3E(g^2(Z)I_1)E(g^2(Z)I_2) \le 3E^2[g^2(Z)(I_1 + I_2)]; \tag{7.6}$$

also

$$E(U_1^2)E(U_2^2) \le E(g^2(Z)I_1)E(g^2(Z)I_2) \le E^2[g^2(Z)(I_1 + I_2)], \tag{7.7}$$

and

$$E^2(U_1U_2) = E^2(g(Z)I_1)E^2(g(Z)I_2) \le E(g^2(Z)I_1)E(g^2(Z)I_2) \le E^2[g^2(Z)(I_1 + I_2)]. \tag{7.8}$$

Applying inequalities (7.6)–(7.8) we get $n^{-2}E[(\sum_{i=1}^{n} U_{1i})^2(\sum_{i=1}^{n} U_{2i})^2] \le 6(F(t_2) - F(t_1))^2$. The proof is complete upon setting $C_4 = 6$.

PROOF OF LEMMA 4.1. The strong consistency of $\hat{\beta}$, the asymptotic normality of $\sqrt{n}(\hat{\beta} - \beta)$ and the use of Slutsky's theorem would prove the desired result upon establishing that the statistic

$$K(t) = \int_0^t d\hat{Q}[\hat{E}(z^2e^{\beta^* z}, x)/\hat{E}^2(e^{\beta^* z}, x) - 2\hat{E}^2(ze^{\beta^* z}, x)/\hat{E}^3(e^{\beta^* z}, x)]$$

is bounded in probability.

Choose $\delta > 0$ such that $E(Z^2e^{\beta' Z})$ and $E(Z^2e^{2\beta' Z})$ are finite where $\beta' = \beta + \delta$. This follows from Assumption 3.1. We establish that

$$\hat{E}(z^2e^{\beta^* z}, x) \le \hat{E}(z^2e^{\beta^* z}, 0) \le \hat{E}(z^2e^{\beta' z}, 0) \tag{7.9}$$

$$\hat{E}^2(ze^{\beta^* z}, x) \le \hat{E}(z^2e^{2\beta^* z}, x) \le \hat{E}(z^2e^{2\beta^* z}, 0) \le \hat{E}(z^2e^{2\beta' z}, 0). \tag{7.10}$$

The first inequality in (7.10) follows by Schwartz's inequality, and the last inequality in both (7.9) and (7.10) is guaranteed with arbitrarily large probability when the sample size is large enough because of the consistency of $\hat{\beta}$. Since the empirical estimates $\hat{E}(z^2e^{\beta' z}, 0)$ and $\hat{E}(z^2e^{2\beta' z}, 0)$ converge in probability to the bounded quantities $E(Z^2e^{\beta' Z})$ and $E(Z^2e^{2\beta' Z})$ respectively, then this implies that $K(t)$ will be bounded in probability as long as $E(e^{\beta^* z}, x)$ is bounded away from 0 uniformly for $x \in [0, M]$. This follows because $\hat{E}(e^{\beta^* z}, x)$ is a nonincreasing function of x and by assumption $P(T \ge M) > 0$.

APPENDIX 2

8. Appendix 2. *Calculation of the covariance structure.* The covariance structure $\text{Cov}(V(s), V(t))$ will be calculated in three parts;

(i) $\text{Cov}(V_1(s), V_1(t))$, where $V_1(s) = A_1(s) + \cdots + A_3(s)$, $0 \le s \le t \le M$,

(ii) $\text{Var}(V_3)$, where $V_3 = C_1 + \cdots + C_6$;

(iii) $\text{Cov}(V_1(s), V_3)$, $0 \le s \le M$.

(i). Write $\text{Cov}(V_1(s), V_1(t))$ as $\text{Var}(V_1(s)) + \text{Cov}(V_1(s), V_1(t) - V_1(s))$ where $s \le t$. We use repeatedly the relationships (2.9), (5.2) and integration by parts.

(a). $\text{Var}(V_1(s)) = \text{Var}(\int_0^s Y dQ/E^2 - \int_0^s X\, dE/E^2 + X(0)/E(0) - X(s)/E(s))$ which is equal to

$$2\int_0^s dQ(r)/E^2(r)\left\{\int_r^s \text{Cov}(Y(r), Y(u))\, dQ(u)/E^2(u) - \int_r^s \text{Cov}(Y(r), X(u))\, dE(u)/E^2(u) + \text{Cov}(Y(r), X(0))/E(0) - \text{Cov}(Y(r), X(s))/E(s)\right\} \tag{8.1}$$

$$\begin{aligned} &+ 2\left\{\int_0^s dE(r)/E^2(r) \int_0^r \text{Cov}(X(r), X(u))\, dE(u)/E^2(u)\right.\\ &- \int_0^s dQ(r)/E^2(r) \int_0^r \text{Cov}(X(r), Y(u))\, dE(u)/E^2(u)\\ &\left. - \int_0^s \text{Cov}(X(r), X(0))\, dE(r)/E^2(r)E(0) + \int_0^s \text{Cov}(X(r), X(s))\, dE(r)/E^2(r)E(s)\right\} \end{aligned} \tag{8.2}$$

$$+ \text{Var}(X(0))/E^2(0) + \text{Var}(X(s))/E^2(s) - 2\text{Cov}(X(0), X(s))/E(0)E(s). \tag{8.3}$$

We note that the covariances of (5.2) are given as a difference of two terms, a single valued term and a product term. Therefore, if we substitute the appropriate covariances of (5.2) into (8.1) through (8.3) then

$$\text{Var}(V_1(s)) = \text{Var}^{(1)}(V_1(s)) - \text{Var}^{(2)}(V_1(s)),$$

where $\operatorname{Var}^{(1)}(V_1(s))$ is the evaluation of (8.1)–(8.3) substituting the single valued term of the covariances, whereas $\operatorname{Var}^{(2)}(V_1(s))$ is the evaluation of (8.1)–(8.3) substituting the product term of the covariances.

It is easily seen that

$$\operatorname{Var}^{(2)}(V_1(s)) = \left[\int_0^s E\,dQ/E^2 - \int_0^s Q\,dE/E^2 + Q(0)/E(0) - Q(s)/E(s)\right]^2. \tag{8.4}$$

Integrating by parts we get

$$\int_0^s dQ/E = Q(s)/E(s) - Q(0)/E(0) + \int_0^s Q\,dE/E^2.$$

Therefore, $\operatorname{Var}^{(2)}(V_1(s)) = 0$.

Substituting the single valued terms of (5.2) into (8.1) we get

$$2\int_0^s dQ(r)/E^2(r)\left\{\int_r^s E(\exp(2\beta z), u)\,dQ(u)/E^2(u) - \int_r^s E_1(\exp(\beta z), u)\,dE(u)/E^2(u) + E_1(\exp(\beta z), r)/E(0) - E_1(\exp(\beta z), s)/E(s)\right\}. \tag{8.5}$$

By using (2.9) we get

$$\int_r^s E(\exp(2\beta z), u)\,dQ(u)/E^2(u) = \int_r^s dE_1(\exp(\beta z), u)/E(u)$$
$$= E_1(\exp(\beta z), s)/E(s) - E_1(\exp(\beta z), r)/E(r) + \int_r^s E_1(\exp(\beta z), u)\,dE(u)/E^2(u).$$

Therefore, (8.5) is equal to

$$-2\int_0^s E_1(\exp(\beta z), r)\,dQ(r)/E^3(r) + 1/E(0)\int_0^s E_1(\exp(\beta z), r)\,dQ(r)/E^2(r). \tag{8.6}$$

Substituting into (8.2) we get

$$-2\int_0^s Q(r)\,dE(r)/E^3(r) + 2\int_0^s E_1(\exp(\beta z), r)\,dQ(r)/E^3(r) - 2/E(0)\int_0^s E_1(\exp(\beta z), r)\,dQ(r)/E^2(r) + 2Q(s)/E(s)[1/E(0) - 1/E(s)]. \tag{8.7}$$

Substituting into (8.3) we get

$$Q(0)/E^2(0) + Q(s)/E^2(s) - 2Q(s)/E(0)E(s). \tag{8.8}$$

Adding (8.6), (8.7), (8.8) we get

$$-2\int_0^s Q(r)\,dE(r)/E^3(r) - Q(s)/E^2(s) + Q(0)/E^2(0). \tag{8.9}$$

Integrating by parts

$$-2\int_0^s Q(r)\,dE(r)/E^3(r) = Q(s)/E^2(s) - Q(0)/E^2(0) + \int_0^s - dQ(r)/E^2(r);$$

therefore, (8.9) is equal to $\int_0^s - dQ(r)/E^2 d(r)$. Thus we show that $\operatorname{Var}(V_1(s)) = \int_0^s - dQ(r)/E^2(r)$.

(b). In order to complete the evaluation of $\mathrm{Cov}(V_1(s), V_1(t))$ we calculate $\mathrm{Cov}(V_1(s), V_1(t) - V_1(s))$. As in part (a) this is done by evaluating the covariances with the single valued terms and the product terms of (5.2), say $\mathrm{Cov}^{(1)} - \mathrm{Cov}^{(2)}$. Also, similar to part (a) $\mathrm{Cov}^{(2)}$ is equal to

$$\mathrm{Cov}^{(2)}(V_1(s), V_1(t) - V_1(s)) = \left[\int_0^s E\, dQ/E^2 - \int_0^s Q\, dE/E^2 + Q(0)/E(0) - Q(s)/E(s)\right]$$

$$\times\left[\int_0^t E\, dQ/E^2 - \int_0^t Q\, dE/E^2 + Q(s)/E(s) - Q(t)/E(t)\right] = 0.$$

$$Cov^{(1)}(V_1(s), V_1(t) - V_1(s)) = \mathrm{Cov}^{(1)}\Big(\int_0^s Y\, dQ/E^2 - \int_0^s X\, dE/E^2 + X(0)/E(0)$$

$$- X(s)/E(s), \int_s^t Y\, dQ/E^2 - \int_s^t X\, dE/E^2$$

$$+ X(s)/E(s) - X(t)/E(t)\Big)$$

$$= \int_0^s dQ/E^2\Big\{\int_s^t E(\exp(2\beta z), u)\, dQ(u)/E^2$$

$$- \int_s^t (\exp(\beta z), u)\, dE/E^2 \tag{8.10}$$

$$+ E_1(\exp(\beta z), s)/E(s) - E_1(\exp(\beta z), t)/E(t)\Big\}$$

$$- \int_0^s dE/E^2\Big\{\int_s^t E_1(\exp(\beta z), u)\, dQ/E^2 \tag{8.11}$$

$$- \int_s^t Q\, dE/E^2 + Q(s)/E(s)$$

$$- Q(t)/E(t)\Big\} + [1/E(0) - 1/E(s)]\Big[\int_s^t E_1(\exp(\beta z), u)\, dQ/E^2$$

$$- \int_s^t Q\, dE/E^2 + Q(s)/E(s) - Q(t)/E(t)\Big]. \tag{8.12}$$

By (2.9), and integration by parts we get

$$\int_s^t E(\exp(2\beta z), u)\, dQ(u)/E^2(u) = \int_s^t dE_1(\exp(\beta z), u)/E(u)$$

$$= E_1(\exp(\beta z), t)/E(t) - E_1(\exp(\beta z), s)/E(s) + \int_s^t E_1(\exp(\beta z), u)\, dE(u)/E^2(u);$$

therefore (8.10) is equal to zero. Since $\int_0^s dE/E^2 = [1/E(0) - 1/E(s)]$, then (8.11) + (8.12) is also equal to zero. Hence,

$$\mathrm{Cov}^{(1)}(V_1(s), V_1(t)) - V_1(s)) = (8.10) + (8.11) + (8.12) = 0$$

Combining the results from parts (a) and (b), we show that

$$\operatorname{Cov}(V_1(s), V_1(t)) = \int_0^s - dQ(u)/E^2(u).$$

The calculations in parts (ii) and (iii) are mechanically identical to those in part (i); therefore we will only state the results and leave the details to the reader.

(ii) The variance of V_3 is equal to

$$\operatorname{Var}(V_3) = \int_0^{T_0} - dQ \operatorname{Var}(z \mid R(t)) = \int_0^{T_0} - dQ\{E(z^2 \exp(\beta z), t)/E(t) - [E_z(t)/E(t)]^2\}.$$

(iii) The covariance

$$\operatorname{Cov}(V_1(s), V_3) = 0 \qquad \text{for all} \quad 0 \leq s \leq M.$$

Acknowledgment. I would like to thank Dr. John Crowley for his help and encouragement.

REFERENCES

[1] AALEN, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3** 15–27.
[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.
[3] BRESLOW, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30** 89–99.
[4] BRESLOW, N. AND CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.* **2** 437–453.
[5] COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc. Ser. B* **34** 187–220.
[6] COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.
[7] CROWLEY, J. (1973). Nonparametric analysis of censored survival data with distribution theory for the k-sample generalized Savage statistics. Unpublished Ph.D. Dissertation, Univ. Washington, Seattle.
[8] CROWLEY, J. AND THOMAS, D. R. (1975). Large sample theory for the log rank test. Technical Report No. 415, Depart. Statist. Univ. Wisconsin, Madison.
[9] EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565.
[10] KALBFLEISCH, J. D. AND PRENTICE, R. L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60** 267–278.
[11] LOÉVE, M. (1963). *Probability Theory.* Van Nostrand, Princeton.
[12] PRENTICE, R. L. AND GLOECKLER, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* **34** 57–67.
[13] PYKE, R. AND SHORACK, G. (1968). Weak convergence of a two-sample empirical process and a new approach to the Chernoff-Savage theorems. *Ann. Math. Statist.* **39** 755–771.
[14] RAO, C. R. (1965). *Linear Statistical Inference and its Applications.* Wiley, New York.
[15] TSIATIS, A. A. (1978). A heuristic estimate of the asymptotic variance of the survival probability in Cox's regression model. Technical Report No. 524, Depart. Statist. Univ. Wisconsin, Madison

ST. JUDE CHILDREN'S HOSPITAL
BIOSTATISTICS DIVISION
332 NORTH LAUDERDALE
MEMPHIS, TENNESSEE 38101