Lifetime Data Analysis, 9, 195–210, 2003

Mixed Discrete and Continuous Cox Regression Model

ROSS L. PRENTICE rprentic@fhcrc.org
Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA

JOHN D. KALBFLEISCH jdkalbfl@umich.edu
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

Received February 21, 2002; Revised April 17, 2002; Accepted May 24, 2002

Abstract. The Cox (1972) regression model is extended to include discrete and mixed continuous/discrete failure time data by retaining the multiplicative hazard rate form of the absolutely continuous model. Application of martingale arguments to the regression parameter estimating function show the Breslow (1974) estimator to be consistent and asymptotically Gaussian under this model. A computationally convenient estimator of the variance of the score function can be developed, again using martingale arguments. This estimator reduces to the usual hypergeometric form in the special case of testing equality of several survival curves, and it leads more generally to a convenient consistent variance estimator for the regression parameter. A small simulation study is carried out to study the regression parameter estimator and its variance estimator under the discrete Cox model special case and an application to a bladder cancer recurrence dataset is provided.

Keywords: Cox regression, counting process, martingale, tied failure times

1. Introduction

The Cox (1972) hazard function regression model,

$$\lambda\{t; Z(t)\} = \lambda_0(t) \exp\{X(t)'\beta\}, \tag{1}$$

for a failure time variate $T > 0$, is well established as a central tool for the regression analysis of absolutely continuous failure times. In (1), $Z(t) = \{z(u), u < t\}$ is the history of a covariate process z at times less than t, λ_0 is an unspecified 'baseline' hazard function, $X(t)' = \{X_1(t), \ldots, X_p(t)\}$ is a modeled regression p-vector with elements comprised of functions of $Z(t)$ and terms involving products of these functions with functions of t. Typically, interest resides in inference on both the regression vector $\beta' = (\beta_1, \ldots, \beta_p)$, sometimes referred to as the relative risk parameter in acknowledgment of the multiplicative form of (1), and the cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$.

The model (1) was presented for absolutely continuous failure times. Various proposals have been made to extend (1) to accommodate discrete components, since tied failure times commonly arise in applications. These proposals, elaborated below, have either dropped the multiplicative form (1) for the discrete hazard rates, or have made *ad hoc* adjustments to the continuous failure time estimation procedure. Here, instead, we retain the multiplicative form (1) even at times where the cumulative hazard function is not continuous, giving a model

$$\Lambda\{dt; Z(t)\} = \Lambda_0(dt) \exp\{X(t)'\beta\}, \tag{2}$$

where for example $\Lambda_0(dt) = \Lambda_0(t) - \Lambda_0(t^-)$ if t is a mass point of the failure distribution, while $\Lambda_0(dt) = \{d\Lambda_0(t)/dt\}dt = \lambda_0(t)dt$ at a continuity point of the failure distribution. We will develop estimation procedures for β and Λ_0 in (2), using counting process methods.

2. Methods for Handling Tied Failure Times

Cox (1972) suggested that (1) be relaxed to the discrete logistic model

$$\Lambda\{dt; Z(t)\} = \frac{\Lambda_0(dt) \exp\{X(t)'\beta\}}{1 + \Lambda_0(\Delta t)[\exp\{X(t)'\beta\} - 1]} \tag{3}$$

to accommodate discrete components to the failure time distribution and tied failure time data, where $\Lambda_0(\Delta t) = \Lambda_0(t) - \Lambda_0(t^-)$. His partial likelihood argument extended readily to (3) upon conditioning on the number of failures at each distinct failure time. However, if there are d failures at a time point t, the partial likelihood factor at t involves a summation over all subsets of size d from the r individuals at risk at that t. This gives unwieldy calculations if d and $r - d$ are large, although there are some ways of streamlining these or closely related calculations (Gail et al., 1981). However, it can be noted that the regression coefficient in (3) has a logistic regression rather than a hazard ratio interpretation or, in epidemiologic parlance, has an odds ratio rather than a relative risk interpretation.

Kalbfleisch and Prentice (1973) considered instead a discrete failure time model obtained by grouping the failure times under (1) according to a fixed partition of the time axis. This approach leads to a mixed discrete/continuous hazard rate model

$$\Lambda\{dt, Z(t)\} = 1 - \{1 - \Lambda_0(dt)\}^{\exp\{X(t)'\beta\}}. \tag{4}$$

for an appropriately redefined Λ_0 function. The regression parameter in (4) is the same as in the underlying continuous model (1) when the modeled covariate is time-independent (i.e., $X(t) \equiv X$). Note also in this special case that (4) gives a survivor function

$$F\{t; Z(t)\} = \prod_0^t \{1 - \Lambda_0(dt)\}^{\exp(X'\beta)} = F_0(t)^{\exp(X'\beta)}$$

in the so-called Lehmann class for both discrete and continuous failure time variables. Using the continuous model, again with time-independent covariates, Kalbfleisch and Prentice (1980, p.74) obtain a marginal likelihood for β that accommodates ties and performs well in analyzing data from (4). This marginal likelihood is the time-independent special case of

$$L(\beta) = \prod_{i=1}^{k} \exp\{S(t_i)'\beta\} \sum_{P\varepsilon Q_i} \prod_{r=1}^{d_i} \left[\sum_{\ell\varepsilon R(t_i,\, p_r)} \exp\{X_\ell(t_i)'\beta\} \right]^{-1} \tag{5}$$

where $t_1 < \cdots < t_k$ are the ordered failure times in the sample, $S(t_i)$ is the sum of the covariate vectors $X(t_i)$ for the d_i 'subjects' failing at t_i, Q_i is the set of permutations over these d_i failing subjects, $P = (p_1, \ldots, p_{d_i})$ is an element of Q_i, $R(t_i)$ is the set of subjects at risk at time t_i and $R(t_i, p_r)$ is the set difference $R(t_i) - \{p_1, \ldots, p_{r-1}\}$. This likelihood arises by breaking the d_i ties at t_i in all possible $d_i!$ ways. Estimation based on (5) is also computationally difficult if some of the d_i's are large. Delong et al. (1994) developed an integral expression for the factors of (5) that can reduce computation time for factors having large d_i values, but numerical integration is required. To date, there is no computationally convenient implementation of the grouped continuous hazard model (4).

Because of these computational difficulties, most software packages use as a default maximization of an approximate likelihood. Breslow (1974) suggested the widely used approximate partial likelihood

$$L(\beta) = \prod_{i=1}^{k} \exp\{S(t_i)'\beta\} \Bigg/ \left[\sum_{\ell\varepsilon R(t_i)} \exp\{X_\ell(t_i)'\beta\} \right]^{d_i}. \tag{6}$$

A somewhat better approximation to (5) is due to Efron (1977),

$$L(\beta) = \prod_{i=1}^{k} \exp\{S(t_i)'\beta\} \Bigg/ \prod_{r=0}^{d_i-1} \left[\sum_{\ell\varepsilon R(t_i)} \exp\{X_\ell(t_i)'\beta\} - r d_i^{-1} \sum_{\ell\varepsilon D(t_i)} \exp\{X_\ell(t_i)'\beta\} \right], \tag{7}$$

where $D(t_i)$ denotes the set of subjects failing at t_i. The estimators $\hat{\beta}$ that maximize (6) or (7) are readily calculated, but both have some asymptotic bias under the model (4), and the corresponding matrices $\{-\partial^2 \log L(\hat{\beta})/\partial\hat{\beta}\,\partial\hat{\beta}'\}^{-1}$ are not consistent estimators of the covariance of $\hat{\beta}$.

3. Mixed Discrete and Continuous Relative Risk Model

As mentioned above, another generalization of the continuous failure time model (1) is (2), which we shall refer to as the mixed discrete and continuous Cox model. The regression parameter in (2) retains a natural and useful relative risk interpretation, even if the failure time distribution includes discrete elements. In general, (2) does not arise by grouping failure times from the continuous model (1), though models (2), (3), and (4), all coincide in the very special case in which baseline hazard function for the discrete variate is constant across failure times and the modeled regression variable is binary. The principal constraint on (2) is that the (discrete) hazard at any mass point of the failure distribution must be equal to or less than one. This is not a practically important constraint in the typical setting where the number of failures is a small fraction of the number of individuals at risk at most failure times. However, it is important that the application of (2) not be adversely affected by violations, or near violations of this constraint.

Let $N_i(t)$ be the right continuous counting process that takes value zero at $t = 0$ and jumps by one at an observed failure time for the ith subject. We also define, under (2), the cumulative intensity process

$$\Lambda_i(t) = \int_0^t Y_i(s) \exp\{X_i(s)'\beta\}\Lambda_0(ds)$$

where $Y_i(s)$ takes value one if the ith subject is at risk for an observed failure at time s^-, and value zero otherwise. Let $\mathcal{F}(t) = \{N_i(u), Y_i(u^+), Z_i(u^+), u \leq t, i = 1, \ldots, n\}$ specify the counting, censoring and covariate processes up to time t. For inference on β in (2), consider the Doob-Meyer decomposition

$$N_i(t) = \Lambda_i(t) + M_i(t), \; i = 1, \ldots, n. \tag{8}$$

The difference $M_i = N_i - \Lambda_i$ is a square integrable martingale with respect to the filtration or history $\mathcal{F}(t)$ under mild conditions. For example, it is sufficient that sample paths for each Y_i and X_i be left continuous with right hand limits.

An estimating function, which reduces to the partial likelihood score process (Cox, 1975; Andersen and Gill, 1982) if failure times are absolutely continuous, can be written

$$U(\beta, t) = \int_0^t \sum_{i=1}^n \left\{ Y_i(s)X_i(s) - \sum_{\ell=1}^n X_\ell(s)p_\ell(s) \right\} N_i(ds) \tag{9}$$

where

$$p_\ell(s) = Y_\ell(s) \exp\{X_\ell(s)'\beta\} \Big/ \sum_{j=1}^n Y_j(s) \exp\{X_j(s)'\beta\}.$$

Just as in the absolutely continuous special case, substitution from (8) into (9) leads to

$$U(\beta, t) = \int_0^t \sum_{i=1}^n \left\{ Y_i(s)X_i(s) - \sum_{\ell=1}^n X_\ell(s)p_\ell(s) \right\} M_i(ds) \tag{10}$$

so that U is a stochastic integral of a predictable process with respect to a square integrable martingale. Hence, U is itself a square integrable martingale with respect to $\{\mathcal{F}(t)\}$. Note that the overall score $U(\beta, \infty)$ is precisely $\partial \log L(\beta)/\partial\beta$ from (6). Hence, one can expect the Breslow estimator $\hat{\beta}$ to be consistent for β in (2) under suitable regularity conditions.

Similarly, using (8), a natural estimate of $\Lambda_0(t)$, for given β is the Nelson-Aalen-type estimator,

$$\hat{\Lambda}_0(\beta, t) = \int_0^t N.(ds) \Big/ \sum_{\ell=1}^{n} Y_\ell(s) \exp\{X_\ell(s)'\beta\}$$

$$= \int_0^t I\{Y.(u) > 0\}\Lambda_0(du) + \int_0^t M.(ds) \Big/ \sum_{\ell=1}^{n} Y_\ell(s) \exp\{X_\ell(s)'\beta\},$$

where a dot denotes summation from 1 to n and ratios 0/0 are defined to take value zero. Hence, $\hat{\Lambda}_0(t)$ can be expected to be consistent for $\Lambda_0(t)$ in (2) under regularity conditions that ensure that $I\{Y.(u) > 0\}$ converges to one uniformly for $u\varepsilon$ [0, t].

Since U is a martingale under (2), and hence has mean zero with uncorrelated increments, the variance of $U(t)$ can be written

$$V(\beta, t) = \int_0^t V(\beta, ds) = \int_0^t E\{U(\beta, ds)U(\beta, ds)' \,|\, \mathcal{F}(s^-)\}.$$

If the martingales M_i, $i = 1, \ldots, n$ are orthogonal (as follows for example from independent failure time and independent censoring assumptions) then

$$V(\beta, ds) = \sum_{i=1}^{n} W_i(s)W_i(s)'E\{M_i(ds)M_i(ds)' \,|\, \mathcal{F}(s^-)\}$$

$$= \sum_{i=1}^{n} W_i(s)W_i(s)'Y_i(s)[1 - \Lambda_0(\Delta s)\exp\{X_i(s)'\beta\}]\Lambda_0(ds)\exp\{X_i(s)'\beta\}, \tag{11}$$

where $W_i(s) = Y_i(s)X_i(s) - \sum_{\ell=1}^{n} X_\ell(s)p_\ell(s), i = 1, \ldots, n.$ Note that the factor in square brackets takes value one if s is a continuity point of the failure distribution.

Substituting $\hat{\Lambda}_0(\beta, \cdot)$ for Λ_0 in (11) gives an 'estimator' of $V(\beta, t)$ at specified β that can be written $\hat{V}(\beta, t) = \int_0^t \hat{V}(\beta, ds)$ where

$$\hat{V}(\beta, ds) = \sum_{i=1}^{n} W_i(s)W_i(s)'Y_i(s)\exp\{X_i(s)'\beta\}\hat{\Lambda}_0(\beta, ds)$$

$$- \sum_{i=1}^{n} W_i(s)W_i(s)'Y_i(s)\exp\{2X_i(s)'\beta\}\ \hat{\Lambda}_0(\beta, \Delta s)\hat{\Lambda}_0(\beta, ds). \tag{12}$$

The first term in (12) is the contribution at time s to $-\partial^2 \log L(\beta)/\partial\beta\, \partial\beta^T$ from (6), while the second term corrects the variance contribution for discreteness.

Expression (12) is somewhat unsatisfactory in that a variance correction is made even if $N.(\Delta s) = 1$; that is, even if there are no tied failure times at time s. Hence, $\hat{V}$ does not reduce to the usual score process variance estimator in the special case of absolutely continuous failure times, though the difference between the two estimators will generally disappear as sample size increases. A variance estimator that reduces to the usual estimator with absolutely continuous failure times, and that has somewhat better small sample properties arises by replacing $\alpha_0(ds) = \Lambda_0(\Delta s)\Lambda_0(ds)$ in (11) by the 'unbiased' estimator

$$\hat{\alpha}_0(\beta, ds) = \frac{\{N.(\Delta s) - 1\}N.(ds)}{\left[\sum_{\ell=1}^n Y_\ell(s) \exp\{X_\ell(s)'\beta\}\right]^2 - \sum_{\ell=1}^n Y_\ell(s) \exp\{2X_\ell(s)'\beta\}},$$

which arises from consideration of $E\left\{\sum_{i \neq j} N_i(ds)N_j(ds) \,\middle|\, \mathcal{F}(s^-)\right\}$.

The resulting variance estimator is $\hat{V}(\beta, t) = \int_0^t \hat{V}(\beta, ds)$ where

$$\hat{V}(\beta, ds) = \sum_{i=1}^n W_i(s)W_i(s)'Y_i(s) \exp\{X_i(s)'\beta\}\hat{\Lambda}_0(\beta, ds) - \sum_{i=1}^n W_i(s)W_i(s)'Y_i(s) \exp\{2X_i(s)'\beta\}\hat{\alpha}_0(\beta, ds) \tag{13}$$

For example, the contribution at time s to the variance of the score test for $\beta = 0$ can be written

$$\hat{V}(0, ds) = \sum_{i=1}^n Y_i(s)\{X_i(s) - \bar{X}(s)\}\{X_i(s) - \bar{X}(s)\}' \frac{N.(ds)[Y.(s) - N.(\Delta s)]}{Y.(s)[Y.(s) - 1]},$$

where $\bar{X}(s) = \sum Y_i(s)X_i(s)/Y.(s)$. In the special case in which $X(s) = X$ is comprised of indicator variables for p of $p + 1$ samples, (13) is precisely the standard hypergeometric covariance contribution at time t, usually obtained by conditioning on $N.(\Delta s)$.

A Taylor expansion of $\partial \log L(\beta)/\partial\beta$ from (6) leads to a variance estimator for the Breslow estimator $\hat{\beta}$ under (2) that can be written

$$\left\{-\partial^2 \log L(\hat{\beta})/\partial\hat{\beta}\,\partial\hat{\beta}^T\right\}^{-1} \hat{V}(\hat{\beta}) \left\{-\partial^2 \log L(\hat{\beta})/\partial\hat{\beta}\,\partial\hat{\beta}^T\right\}^{-1} \tag{14}$$

where $\hat{V}(\beta) = \hat{V}(\beta, \infty)$ is based on (12) or (13).

A sketch of the proof of the consistency and asymptotic normality of $\hat{\beta}$, of the consistency of variance estimators for $\hat{\beta}$ arising from (12) and (13), and of the consistency and asymptotic Gaussian distribution for $\hat{\Lambda}_0(\hat{\beta}, \cdot)$ is given in the Appendix.

4. Simulation Evaluation

Data were generated from the discrete model (2) with $\Lambda_0(\Delta t)$ either 0.1 or 0.2 at integer values $t = 1, 2, \ldots$ and $\Lambda_0(\Delta t) = 0$ otherwise, and with $X(t) = X$ either binary with values

-0.5 or 0.5 with probability 0.5, or normal with mean zero and variance 0.25. Censoring times were generated from an exponential distribution having a mean of 10 if $\Lambda_0(dt) = 0.1$, and a mean of 5 if $\Lambda_0(dt) = 0.2$ at integer values of t, so that there is a censoring probability of about 51–53% at each sampling configuration. Relative risk parameter values $\beta = 0$ and $\beta = \log 2 = 0.693$ were considered at each sample configuration. Ten thousand simulations were carried out with sample sizes of $n = 50$ or $n = 100$.

Table 1 shows sample means and variances for $\hat{\beta}$ along with the mean of the usual approximate likelihood variance estimator $V_1 = -\partial^2 \log L(\hat{\beta})/\partial\hat{\beta}\, \partial\hat{\beta}^T$ from (6), the consistent variance estimator V_2 from (12) and (14), and the preferred consistent estimator V_3 from (13) and (14). Table 1 also shows the empirical coverage rate for nominal 95% confidence intervals $(\hat{\beta} - 1.96\, V_j^{\frac{1}{2}}, \hat{\beta} + 1.96\, V_j^{\frac{1}{2}})$ using each variance estimator $j = 1, 2, 3$.

At sample size $n = 50$ one can see a slight upward bias in the Breslow estimator $\hat{\beta}$ at $\beta = \log 2$. The approximate partial likelihood variance $V1$ somewhat exceeds the sample variance for $\hat{\beta}$ especially at $\Lambda_0(\Delta t) = 0.2$. The consistent variance estimator $V2$ over-corrects this excess, while incorporation of the unbiased estimator of $\Lambda_0(\Delta t)\Lambda_0(dt)$ in $V3$ substantially reduces the overcorrection. There were two samples (out of 80,000)

Table 1. Summary statistics for estimation of the regression parameter β in the discrete special case of the Cox model (2). Statistics are based on 10,000 simulations of samples of sizes 50 and 100 at each of eight sampling configurations.

$\Lambda_0(dt)$	β	X	Avg $\hat{\beta}$	Sample var $\hat{\beta}$	$V1^{\dagger}$	$V2^{\dagger}$	$V3^{\dagger}$	$CI1^{\circ}$	$CI2^{\circ}$	$CI3^{\circ}$
					$n = 50$					
0.1	0.0	B*	0.005	0.183	0.189	0.160	0.171	96.0	94.5	95.0
0.1	0.0	N*	0.005	0.190	0.201	0.169	0.180	95.9	93.7	94.5
0.1	0.693	B	0.715	0.193	0.204	0.174	0.185	96.7	94.7	95.6
0.1	0.693	N	0.735	0.200	0.216	0.178	0.190	96.5	94.0	94.8
0.2	0.0	B	0.001	0.166	0.194	0.148	0.158	97.3	94.4	95.4
0.2	0.0	N	−0.001	0.181	0.206	0.154	0.165	97.0	93.5	94.4
0.2	0.693	B	0.728	0.183	0.213	0.164	0.174	97.6	94.9	95.6
0.2	0.693	N	0.713	0.179	0.220	0.157	0.168	97.4	93.2	94.2
					$n = 100$					
0.1	0.0	B	0.002	0.082	0.088	0.076	0.079	96.0	94.3	94.8
0.1	0.0	N	0.000	0.086	0.091	0.079	0.082	95.8	94.0	94.5
0.1	0.693	B	0.707	0.089	0.094	0.082	0.085	95.7	94.3	94.8
0.1	0.693	N	0.705	0.090	0.097	0.082	0.085	96.2	94.0	94.4
0.2	0.0	B	0.001	0.076	0.091	0.070	0.073	97.0	94.2	94.6
0.2	0.0	N	−0.003	0.078	0.094	0.072	0.075	96.9	93.9	94.4
0.2	0.693	B	0.706	0.081	0.098	0.076	0.079	97.1	94.6	95.0
0.2	0.693	N	0.707	0.078	0.100	0.073	0.075	97.5	94.0	94.5

* B refers to a binary covariate, N to a normal covariate. Each has mean 0.0 and standard deviation 0.5.

† $V1$ is average of variance estimates from approximate partial likelihood (6), $V2$ is average of corrected variance estimators using (12), $V3$ is average of variance estimators using (13).

° $CI1$, $CI2$ and $CI3$ are empirical coverage rates for intervals $(\hat{\beta} - 1.96\, V^{\frac{1}{2}}, \hat{\beta} + 1.96\, V^{\frac{1}{2}})$ for $V = V1$, $V2$ and $V3$, respectively.

at $n = 50$ that failed to converge within 10 iterations, one at $\Lambda_0(\Delta t) = 0.1$, $\beta = \log 2$, and one at $\Lambda_0(\Delta t) = 0.2$, $\beta = \log 2$ in the binary covariate special case. These samples are excluded from the summary statistics shown in Table 1.

At $n = 100$, all 80,000 runs converged by 10 iterations. The upward bias in $\hat{\beta}$ at $\beta = \log 2$ is much less, and the overestimation of the variance of $\hat{\beta}$ by the approximate partial likelihood procedure (6) is quite evident, especially at $\Lambda(\Delta t) = 0.2$ where there are many tied failure times. The consistent variance estimator $V2$ again somewhat over-corrects, while there is a good correspondence between the average of the variance estimators ($V3$) and the sample variance. For $V3$, the empirical coverage rates are between 94.4 and 95.0 (with standard error of about 0.05) across the eight sampling configurations. Hence, it seems appropriate to recommend this variance estimator ($V3$) for the Breslow estimator.

5. Bladder Tumor Recurrence Illustration

Byar (1980) discusses a randomized trial, conducted by the Veteran's Administration Cooperative Urological Group, among patients having superficial bladder tumors. One question of interest concerned the effect of the treatment thiotepa on the rate of tumor recurrence. Tumors present at baseline were removed transurethrally prior to randomization. In addition to the effect of treatment, there was interest in the relationship of recurrence rate to the number of pre-randomization tumors, and to the size of the largest such tumor.

Table 2, abstracted from Andrews and Herzberg (1985, pp. 254–299) shows some data from this trial, including the possibly right censored time to first post-randomization recurrence. There were 48 patients assigned to the placebo group of whom 29 experienced at least one recurrence, and 38 patients assigned to thiotepa of whom 18 experienced at least one recurrence, over a trial follow-up period that averaged 31 months. Recurrence times were recorded in months resulting in some tied recurrence times, including 8 tied recurrence times at each of two and three months.

Table 3 shows some analyses of these data using the discrete and continuous Cox model (2). The Breslow estimator is shown for each of three regression variables, along with corresponding standard deviation estimates from the approximate partial likelihood (6), and from the corrected variance estimators using (12) or (13). The variance corrections can be seen to be quite small in this illustration. Using the standard deviation estimates from (13) one obtains a standardized test statistic of $-0.517/0.308 = -1.68$ for treatment, and corresponding test statistics of 3.22 and 0.69 for the number of baseline tumors, and the diameter of the largest such tumor, respectively. Hence, there is suggestive evidence for a benefit of thiotepa ($p = 0.09$) and strong evidence for association of recurrence risk with the number of baseline tumors.

The right side of Table 3 repeats these analyses following a grouping of the recurrence times into six month intervals. Now, with a large number (28) of recurrences in the first grouping interval one can see somewhat greater conservatism in the approximate partial likelihood standard deviation estimates. For example, the standardized statistic for testing

Table 2. Bladder tumor recurrence data adapted from Andrews and Herzberg (1985, pp. 254–259).

Initial Tumors[1]		Recurrence	Initial Tumors		Recurrence[2]
Number	Size	Time	Number	Size	Time
		Placebo Group			
1	1	0*	1	5	2
1	3	1*	2	1	3
2	1	4*	1	3	12
1	1	7*	1	2	32*
5	1	10*	2	1	34*
4	1	6	2	1	36*
1	1	14*	3	1	29
1	1	18*	1	2	37*
1	3	5	4	1	9
1	1	12	5	1	16
3	3	23*	1	2	41*
1	3	10	1	1	3
1	1	3	2	6	6
3	1	3	2	1	3
2	3	7	1	1	9
1	1	3	1	1	18
1	2	26*	1	3	49*
8	1	1	3	1	35
1	4	2	1	7	17
1	2	25	3	1	3
1	4	29*	1	1	59*
1	2	29*	3	2	2
4	1	29*	1	3	5
1	6	28	2	3	2
		Thiotepa Group			
1	3	1*	8	3	26
1	1	1*	1	1	38*
8	1	5	1	1	22
1	2	9*	6	1	4
1	1	10*	3	1	24
1	1	13*	3	2	41*
2	6	3	1	1	41*
5	3	1	1	1	1
5	1	18*	1	1	44*
1	3	17	6	1	2
5	1	2	1	2	45*
1	1	17	1	4	2
1	1	22*	1	4	46*
1	3	25*	3	3	49*
1	5	25*	1	1	50*
1	1	25*	4	1	4
1	1	6	3	4	54*
1	1	6	2	1	38
2	1	2	1	3	59*

[1] Initial number of tumors of eight denotes 8 or more. Size denotes diameter of largest such tumor in centimeters.
[2] Recurrence times are measured in months. An asterisk denotes right censoring.

Table 3. Cox model (2) analysis of bladder tumor recurrence data.

	Recurrence Times from Table 1				Recurrence Times Grouped into 6-Month Intervals			
		Std. Dev. Estimates from				Std. Dev. Estimates from		
Regression Variable	$\hat{\beta}$	(6)	(12)	(13)	$\hat{\beta}$	(6)	(12)	(13)
Treatment (0-placebo; 1-thiotepa)	−0.517	0.316	0.305	0.308	−0.471	0.309	0.272	0.275
Number of baseline tumors	0.235	0.076	0.071	0.073	0.204	0.074	0.055	0.057
Size (*cm*) of largest baseline tumor	0.068	0.101	0.097	0.098	0.067	0.102	0.088	0.089

for no effect of thiotepa takes value $-0.471/0.209 = -1.52$ ($p = 0.13$) based on (6), as compared to -1.73 ($p = 0.08$) based on (12) and -1.71 ($p = 0.09$) based on (13), with somewhat different implications concerning the suggestion of benefit from thiotepa.

6. Summary and Discussion

A simple generalization (2) of the Cox regression model has been proposed to accommodate continuous, discrete and mixed continuous/discrete failure times. The Breslow (1974) estimators of the regression parameter and cumulative baseline hazard function, though usually regarded as computationally convenient approximations to estimators under the grouped continuous model (4), are shown to be consistent under the discrete/continuous relative risk model (2). The variance estimator obtained by regarding (6) as a partial likelihood is inconsistent, but simply calculated consistent estimators can be obtained using martingale theory. These variance estimators, and corresponding nominal confidence intervals based on asymptotic normal approximations appear to have adequate performance in moderate sized samples, especially if unbiased estimators of $\Lambda_0(dt)$ and $\Lambda_0(\Delta t)\Lambda_0(dt)$, and the variance estimator from (13) is used. The score statistic variance estimator (13) also has the advantage of reducing to the familiar hypergeometric variance in the special case of a logrank test to compare the survival curves for several populations.

The simulation study and the illustration (Sections 4 and 5) are consistent in supporting the appropriateness of the Breslow estimator under model (2), and the adequacy of the corresponding variance estimator based on (6) unless a noteworthy fraction (e.g., 10% or more) of study subjects fail at specific failure times. For example, the variance correction in the illustration was of little practical importance even though 8 of the 86 study subjects recurred at each of two months, and three months from randomization. On the other hand, the variance correction developed here is easily implemented and avoids concern about undue conservatism in tests and confidence intervals when using the mixed discrete and continuous Cox model (2).

As mentioned above, model (2) requires some constraint on the covariates and relative risk at a mass point $\{\Lambda_0(\Delta t) > 0\}$ in order to assure a valid discrete hazard. For example, in the simulations leading to Table 1 with $\Lambda_0(\Delta t) = 0.2$ and $\beta = \log 2$, the $N(0, 1)$ covariates need to be slightly restricted ($X < 4.64$) to assure that the relative risk factor is less than 5.0.

Since the estimating function (9) involves the regression parameter, but not the baseline hazard function, the constraint that (2) should not exceed unity does not explicitly come into play in testing and estimation on β. For example, it is well-known that the Breslow estimator that solves (8) does not experience numerical difficulties even if the number of ties, or the fraction of the risk set that fails, is large at some follow-up times. It can happen that the hazard rate estimator $\hat{\Lambda}_0(ds)\exp\{X(s)'\beta\}$ exceeds one at extreme covariate values, but this is also the case with absolutely continuous data under the Cox model since a step function estimator of Λ_0 is utilized. However, with tied failure times, but not with absolutely continuous failure times, the estimated hazard rate

$$\hat{\Lambda}_i(ds) = \hat{\Lambda}_0(ds)\exp\{X_i(s)'\beta\}$$

can exceed one at covariate values $X_i(s)$ for individuals in the risk set $R(s)$, and the contribution of such individuals to the variance increment (11) at time s need not be positive. Hence, it would be prudent to modify the contribution of subject i to (12) and (13) to the maximum of the given values and zero in order to acknowledge the hazard rate constraint.

Our simulations study illustrates that the estimation procedures proposed in this paper can be expected to perform well even if the actual and estimated discrete hazards are not small and even if the variance contributions are not constrained in the manner just described. However, the model (2) itself seems less natural if there are only a few distinct failure times, and we recommend that the use of (2), the Breslow estimator, and the corrected variance estimator given here, be restricted to settings in which the discrete hazards are expected to be much less than unity at most failure times.

Acknowledgments

This work was supported by CA 53996 from the National Cancer Institute and a grant from the Natural Science and Engineering Research Council of Canada. The authors would like to thank Mark Mason for computational assistance.

Appendix

Asymptotic Distribution Theory for the Discrete and Continuous Cox Model (2)

A sketch of the asymptotic results stated previously is given here by adapting the arguments of Andersen and Gill (1982) (hereafter AG) for the absolutely continuous special case. Like AG we assume a finite follow-up period $[0, \tau]$, and without loss of generality set $\tau = 1$. For simplicity, we also restrict attention to univariate failure times. Both of these restrictions can likely be avoided.

Extending the AG notation, define

$$S^{(j,k)}(\beta, t) = n^{-1} \sum_{\ell=1}^{n} X_\ell(t)^{(j)} Y_\ell(t) \exp\{X_\ell(t)' k\beta\}$$

for $j = 0, 1, 2$ and $k = 1, 2$, where $X(t)^{(0)} = 1$, $X(t)^{(1)} = X(t)$ and $X(t)^{(2)} = X(t)X(t)'$. Also define: $E(\beta, t) = S^{(1,1)}(\beta, t)S^{(0,1)}(\beta, t)^{-1}$; $C_1(\beta, t) = S^{(2,1)}(\beta, t)S^{(0,1)}(\beta, t)^{-1} - E(\beta, t)E(\beta, t)'$; and $C_2(\beta, t) = \{S^{(2,2)}(\beta, t) - S^{(1,2)}(\beta, t)E(\beta, t)' - E(\beta, t)S^{(1,2)}(\beta, t)' + E(\beta, t)E(\beta, t)' S^{(0,2)}(\beta, t)\}S^{(0,1)}(\beta, t)^{-2}$.

Consider the following conditions:

(i) (finite interval). $\Lambda_0(1) < \infty$

(ii) (asymptotic stability). There exists a neighborhood $\mathcal{B}$ of β_0 and functions $s^{(j,k)}$, $j = 0, 1, 2$, $k = 1, 2$ such that $\| S^{(j,k)}(\beta, t) - s^{(j,k)}(\beta, t) \| \xrightarrow{p} 0$, uniformly in $\mathcal{B} \times [0, 1]$, where $\| \cdot \|$ denotes supremum over the absolute values of the elements of the array.

(iii) (regularity). For all $(\beta, t)\varepsilon\mathcal{B} \times [0, 1]$ and $k = 1, 2$, $s^{(1,k)}(\beta, t) = \partial s^{(0,k)}(\beta, t)/\partial\beta$, $s^{(2,k)}(\beta, t) = \partial s^{(1,k)}(\beta, t)'/\partial\beta$; $s^{(j,k)}$ is a continuous function of $\beta\varepsilon\mathcal{B}$, uniformly in $t\varepsilon[0, 1]$ and $s^{(j,k)}$ is bounded on $\mathcal{B} \times [0, 1]$, for $j = 0, 1, 2$ and $k = 1, 2$; $s^{(0,1)}$ is bounded away from zero on $\mathcal{B} \times [0, 1]$. Also define the matrices

$$\Sigma(t) = \int_0^t c_1(\beta_0, u)s^{(0,1)}(\beta_0, u)\Lambda_0(du),$$

$$\Omega(t) = \Sigma(t) - \Delta(t) = \Sigma(t) - \int_0^t c_2(\beta_0, u)s^{(0,1)}(\beta_0, u)^2\Lambda_0(\Delta u)\Lambda_0(du),$$

where $e = s^{(1,1)}/s^{(0,1)}$, $c_1 = s^{(2,1)}/s^{(0,1)} - e\, e'$ and $c_2 = \{s^{(2,2)} - s^{(1,2)}e' - e\, s^{(1,2)\prime} + e\, e' s^{(0,2)}\}\{s^{(0,1)}\}^{-2}$, and require $\Sigma(1)$ to be positive definite.

(iv) (Lindeberg). For any $\varepsilon > 0$

$$\int_0^1 n^{-1} \sum_{i=1}^{n} \left|X_{ij}(u)\right|^2 I(n^{-1/2}\left|X_{ij}(u)\right| > \varepsilon)\{1 - \Lambda_i(\Delta u)\}\Lambda_i(du) \xrightarrow{p} 0$$

for $j\varepsilon\{1, \ldots, p\}$, where $\Lambda_i(du) = Y_i(u)\exp\{X_i(u)'\beta_0\}\Lambda_0(du)$.

(v) (orthogonality). The martingales $M_i = N_i - \Lambda_i$, $i = 1, \ldots, n$ are orthogonal (e.g., independent failure mechanisms and independent censoring).

THEOREM: *Under the discrete and continuous model (2) and conditions (i) to (v), $\hat\beta$ solving $U(\beta, 1) = 0$ is consistent for the true β_0, and $n^{1/2}(\hat\beta - \beta_0)$ converges in distribution to a mean zero normal distribution with variance $\Sigma(1)^{-1}\Omega(1)\,\Sigma(1)^{-1}$ as $n \to \infty$.*

Proof: The consistency of $\hat{\beta}$ follows from considering

$$L(\beta, t) = \int_0^t \sum_{i=1}^n \Big\{ Y_i(u) X_i(u)'\beta - \log S^{(0,1)}(\beta, u) \Big\} N_i(du), \quad \text{and}$$

$$A(\beta, t) = \int_0^t \sum_{i=1}^n \Big\{ Y_i(u) X_i(u)'\beta - \log S^{(0,1)}(\beta, u) \Big\} \Lambda_i(du).$$

Even though $(N_1, \dots, N_n)$ is not a multivariate counting process, since there may be multiple jumps at a given time, the process

$$n^{-1}[\{L(\beta, t) - L(\beta_0, t)\} - \{A(\beta, t) - A(\beta_0, t)\}] \tag{A1}$$

is a square integrable martingale for each β. Since the (predictable) covariation process for M_i has value $\{1 - \Lambda_i(\Delta u)\}\Lambda_i(du)$ at time u, and the M_i's are orthogonal by assumption (v) the covariation process B for (A1) is given by

$$nB(\beta, t) = n^{-1} \sum_{i=1}^n \int_0^t G_i(\beta, u) G_i(\beta, u)' \{1 - \Lambda_i(\Delta u)\} \Lambda_i(du),$$

where $G_i(\beta, u) = Y_i(w) X_i(u)'(\beta - \beta_0) - \log \big\{ S^{(0,1)}(\beta, u) / S^{(0,1)}(\beta_0, u) \big\}$. Under conditions (i), (ii) and (iii), $nB(\beta, t)$ converges to a finite quantity. Lenglart's inequality (AG, Appendix 1) then shows that $n^{-1}\{L(\beta, t) - L(\beta_0, t)\}$ has the same probability limit as $n^{-1}\{A(\beta, t) - A(\beta_0, t)\}$. The latter is readily shown to converge in probability to a function,

$$\int_0^t \left[s^{(1,1)}(\beta, u)'(\beta - \beta_0) - \log \left\{ \frac{s^{(0,1)}(\beta, u)}{s^{(0,1)}(\beta_0, u)} \right\} s^{(0,1)}(\beta_0, u) \right] \Lambda_0(du)$$

that has first derivative zero at $\beta = \beta_0$, and second derivative the negative of the positive definite matrix $\Sigma(1)$ at $\beta = \beta_0$ and $t = 1$. It follows as in AG that $\hat{\beta} \xrightarrow{P} \beta_0$.

Consider now the standardized score process

$$n^{-1/2} U(\beta_0, t) = \sum_{i=1}^n \int_0^t n^{-1/2} \{ Y_i(u) X_i(u) - E(\beta_0, u) \} \, M_i(du)$$

which defines a square integrable martingale with covariation process given by

$$\int_0^t n^{-1} \sum_{i=1}^n \{ Y_i(u) X_i(u) - E(\beta_0, u) \} \{ Y_i(u) X_i(u) - E(\beta_0, u) \}' \, \{1 - \Lambda_i(\Delta u)\} \Lambda_i(du)$$

$$= \int_0^t C_1(\beta_0, u) S^{(0,1)}(\beta_0, u) \Lambda_0(du) - \int_0^t C_2(\beta_0, u) S^{(0,1)}(\beta_0, u)^2 \Lambda_0(\Delta u) \Lambda_0(du).$$

This covariation process converges in probability to the positive semidefinite matrix $\Omega(t) = \Sigma(t) - \Delta(t)$. The Lindeberg condition (iv) can now be used to show that the conditions for Rebolledo's central limit theorem (e.g., Andersen et al., 1993, p.83) are fulfilled, so that $\{n^{-1/2}U(\beta_0, t), t\varepsilon[0, 1]\}$ converges to a mean zero Gaussian process with variance matrix $\Omega(t)$ at time t. Specifically, for the Rebolledo theorem to apply we need to show that for each j,

$$\int_0^t n^{-1} \sum_{i=1}^{n} \left| X_{ij}(u) - E_j(\beta_0, u) \right|^2 I\left\{ n^{-1/2} \left| X_{ij}(u) - E_j(\beta_0, u) \right| > \varepsilon \right\} \{1 - \Lambda_i(\Delta u)\}\Lambda_i(du) \xrightarrow{p} 0,$$

where the subscript j denotes the jth element of the vector. Now

$$\int_0^1 n^{-1} \sum_{1}^{n} \left| E_j(\beta_0, u) \right|^2 I\left\{ n^{-1/2} \left| E_j(\beta_0, u) \right| > \varepsilon \right\} \{1 - \Lambda_i(\Delta u)\}\Lambda_i(du) \xrightarrow{p} 0$$

on the basis of conditions (i)–(iii), so that condition (iv) and an elementary inequality used by AG (also Andersen et al., p.499) gives the desired result.

A Taylor expansion of $U(\beta, 1)$ about β_0, evaluated at $\hat{\beta}$ gives

$$n^{-1/2}U(\beta_0, 1) = n^{1/2}\hat{\Sigma}(\beta^*, 1)(\hat{\beta} - \beta_0)$$

where $\hat{\Sigma}(\beta, t) = \int_0^t C_1(\beta, u)\{N.(du)/n\}$, and β^* is on the line segment between $\hat{\beta}$ and β_0. Proof that $\hat{\Sigma}(\beta^*, 1)$ is consistent for Σ follows as in AG (p.1108). This establishes the stated asymptotic distribution for $n^{1/2}(\hat{\beta} - \beta_0)$.

As noted above, the argument in AG establishes that $\hat{\Sigma}(\beta^*, t)$ and hence $\hat{\Sigma}(\hat{\beta}, t)$ are consistent estimators of $\Sigma(t)$. A similar approach shows that

$$\hat{\Delta}(\beta^*, 1) = \int_0^1 C_2(\beta^*, t)\left\{ \frac{N.(\Delta t)N.(dt)}{n^2} \right\}$$

to be consistent for Δ. Specifically, consider

$$\begin{aligned}
\left\| \hat{\Delta}(\beta^*, 1) - \Delta(1) \right\| &\leq \left\| \int_0^1 \{C_2(\beta^*, t) - c_2(\beta^*, t)\} \frac{N.(\Delta t)}{n} \frac{N.(dt)}{n} \right\| \\
&+ \left\| \int_0^1 \{c_2(\beta^*, t) - c_2(\beta_0, t)\} \frac{N.(\Delta t)}{n} \frac{N.(dt)}{n} \right\| \\
&+ \left\| \int_0^1 c_2(\beta_0, t)\left\{ \frac{N.(\Delta t)}{n} \frac{N.(dt)}{n} - S^{(0,1)}(\beta_0, t)^2 \Lambda_0(\Delta t)\Lambda_0(dt) \right. \right. \\
&- \left. \left. \frac{S^{(0,1)}(\beta_0, t)}{n} \Lambda_0(dt) + \frac{S^{(0,2)}(\beta_0, t)}{n} \Lambda_0(\Delta t)\Lambda_0(dt) \right\} \right\| \\
&+ \left\| \int_0^1 c_2(\beta_0, t)\left\{ S^{(0,1)}(\beta_0, t)^2 - s^{(0,1)}(\beta_0, t)^2 \right\}\Lambda_0(\Delta t)\Lambda_0(dt) \right. \\
&+ \left. \int_0^1 c_2(\beta_0, t)\frac{S^{(0,1)}(\beta_0, t)}{n}\Lambda_0(dt) - \int_0^1 c_2(\beta_0, t)\frac{S^{(0,2)}(\beta_0, t)}{n}\Lambda_0(\Delta t)\Lambda_0(dt) \right\| .
\end{aligned}$$

The first term on the right side of this expression converges in probability to zero since the stability and regularity conditions imply that $\| C_2(\beta^*, t) - c(\beta^*, t) \|$ converges uniformly to zero in $\mathcal{B} \times [0, 1]$ and $\int_0^1 \frac{N.(\Delta t)}{n} \frac{N.(dt)}{n} \leq 1$. The regularity conditions and the consistency of $\hat{\beta}$ imply that the second term converges in probability to zero. The integral of the quantity in curly brackets in the third term is a square integrable martingale with covariation process that converges at time t to $s^{(0,1)}(\beta_0, t)^2 \Lambda_0(\Delta u) \Lambda_0(du)$. Hence, Lenglart's inequality can be used as in AG to show that this term also converges to zero in probability, as does the final term in view of the stability and regularity conditions. This shows the consistency of the score variance estimators arising from (12) and the consistency of the corresponding $\hat{\Sigma}(\hat{\beta}, 1)^{-1} \hat{\Omega}(\hat{\beta}, 1) \hat{\Sigma}(\hat{\beta}, 1)^{-1}$ as a variance estimator for $n^{1/2}(\hat{\beta} - \beta_0)$. It follows that (12) in conjunction with (14) over a finite follow-up interval provides a valid variance estimator for $\hat{\beta}$. A simple argument establishes that the difference between the score statistic variance estimator based on (13) and that based on (12) converges in probability to zero. This justifies (13) in conjunction with (14) as a variance estimator for $\hat{\beta}$ over a finite follow-up interval.

For brevity we will not go through the development of the asymptotic Gaussian distribution for the cumulative baseline hazard function estimator defined by $\hat{\Lambda}_0(\hat{\beta}, t)$, $t \varepsilon [0, 1]$. Instead we merely note that $\hat{\Lambda}_0(\hat{\beta}, \cdot)$ is strongly consistent for Λ_0 over [0, 1] from writing

$$\begin{aligned}
\left| \hat{\Lambda}_0(\hat{\beta}, t) - \Lambda_0(t) \right| &= \left| \int_0^t S_0^{-1}(\hat{\beta}, u) I\{Y.(u) > 0\} \frac{N.(du)}{n} - \int_0^t \Lambda_0(du) \right| \\
&\leq \left| \int_0^t \left\{ S_0^{-1}(\hat{\beta}, u) - s_0^{-1}(\hat{\beta}, u) \right\} I\{Y.(u) > 0\} \frac{N.(du)}{n} \right| \\
&\quad + \left| \int_0^t \left\{ s_0^{-1}(\hat{\beta}, u) - s_0^{-1}(\beta_0, u) \right\} I\{Y.(u) > 0\} \frac{N.(du)}{n} \right| \\
&\quad + \left| \int_0^t s_0^{-1}(\beta_0, u) I\{Y.(u) > 0\} \left\{ \frac{N.(du)}{n} - \frac{s_0(\beta_0, u)}{n} \Lambda_0(du) \right\} \right| \\
&\quad + \left| \int_0^t [I\{Y.(u) > 0\} - 1] \frac{\Lambda_0(du)}{n} \right|.
\end{aligned}$$

Each term on the right hand side is readily seen to converge in probability to zero, uniformly for $t \varepsilon$ [0, 1].

References

P. K. Andersen and R. D. Gill, "Cox's regression model for counting processes: a large sample study," *Ann. Statist.* vol. 10 pp. 1100–1120, 1982.

P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding, *Statistical Models Based on Counting Processes*, Springer-Verlag: New York, 1993.

D. F. Andrews and A. M. Herzberg, *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, Springer-Verlag: New York, 1985.

N. E. Breslow, "Covariance analysis of censored survival data," *Biometrics* vol. 30 pp. 89–99, 1974.

D. P. Byar, "The Veteran's Administration study of chemoprophylaxis of recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa," in *Bladder Tumors and Other Topics in Urological Oncology*, (M. Pavone-Macaluso, P. H. Smith, and F. Edsmyn, eds.) Plenum: New York, pp. 363–370, 1980.

D. R. Cox, "Regression models and life tables (with discussion)," *J. R. Statist. Soc.* vol. B-34 pp. 187–220, 1972.

D. R. Cox, "Partial likelihood," *Biometrika* vol. 62 pp. 269–276, 1975.

D. M. Delong, G. H. Guirguis, and Y. C. So, "Efficient computation of subset selection probabilities with application to Cox regression," *Biometrika* vol. 81 pp. 607–611, 1994.

B. Efron, "Efficiency of Cox's likelihood function for censored data," *J. Amer. Statist. Assoc.* vol. 72 pp. 557–565, 1977.

M. H. Gail, J. H. Lubin, and L. V. Rubinstein, "Likelihood calculations for matched case-control studies and survival studies with matched death times," *Biometrika* vol. 68 pp. 703–707, 1981.

J. D. Kalbfleisch and R. L. Prentice, "Marginal likelihoods based on Cox's regression and life model," *Biometrika* vol. 60 pp. 267–278, 1973.

J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley: New York, 1980.