Regression with Censored Data
Author(s): Rupert Miller and Jerry Halpern
Reviewed work(s):
Source: *Biometrika,* Vol. 69, No. 3 (Dec., 1982), pp. 521-531
Published by: Biometrika Trust
Stable URL: http://www.jstor.org/stable/2335987
Accessed: 29/11/2011 02:49

Regression with censored data

By RUPERT MILLER and JERRY HALPERN

Division of Biostatistics, Stanford University Medical Center, California, U.S.A.

SUMMARY

There are four regression techniques currently available for use with censored data which do not assume particular parametric families of survival distributions. They are due to (i) Cox (1972), (ii) Miller (1976), (iii) Buckley & James (1979), and (iv) Koul, Susarla & Van Ryzin (1981). These four methods are described, and their performances compared on the updated Stanford heart transplant data. Conclusions on the usefulness of the four procedures are drawn.

Some key words: Censored data; Linear model; Proportional hazards model; Regression.

1. INTRODUCTION

Within the past decade a variety of techniques have been proposed for handling regression problems in which the dependent variable is subject to censoring. Some of these techniques rely upon normal theory, but others make virtually no assumption about the underlying distribution. The purpose of this paper is to compare four methods of the latter type.

One of these techniques (Cox, 1972) bases its approach on the proportional hazards model. If $F(t; x)$ and $f(t; x)$ are the underlying distribution and density functions for the survival time T when the vector of independent variables is x, the proportional hazards model assumes that the hazard rate $\lambda(t; x) = f(t; x)/\{1 - F(t; x)\}$ is given by

$$\lambda(t; x) = \lambda_0(t)\, e^{x\beta},$$

where β is the vector of regression coefficients and $\lambda_0(t)$ is the hazard rate when $x = 0$. This model is equivalent to assuming Lehmann alternatives for $1 - F(t; x)$.

The other three methods, due to Miller (1976), Buckley & James (1979) and Koul, Susarla & Van Ryzin (1981), are based on the standard linear model $F(t; x) = F(t - \alpha - x\beta)$ with

$$E(T \mid x) = \alpha + x\beta, \tag{1}$$

where α is the intercept and β is the vector of regression coefficients for the independent variables in x. If T is measured on a log scale so that $T = \log U$ where U is the actual survival time, then (1) corresponds to an accelerated time model.

The four different estimators are described in §2. Mention is also made of other available regression estimators for censored data. The results of applying these estimators to the updated Stanford heart transplant data are given in §3. In §4 conclusions are drawn about these estimators based on their performances on this data set.

2. ESTIMATORS

2·1. *Generalities*

Let $T_1, ..., T_n$ be the survival times for n patients. These may not all be observable due to censoring times $C_1, ..., C_n$ for each of the patients. The observable quantities are

$$Y_i = T_i \wedge C_i, \quad \delta_i = I(T_i \leqslant C_i), \quad x_i = (x_{i1}, ..., x_{ip}),$$

where '$\wedge$' denotes the minimum, I is the indicator function, and x_i is the transposed vector of independent variables associated with patient i. The random censoring model adopted in this paper assumes that, given the x_i, the T_i and C_i are independently distributed according to the continuous distributions $F(t; x_i)$ and $G(t; x_i)$, respectively.

In the above framework the survival times T_i are censored on the right by the censoring times C_i. They could just as well be all left censored with $Y_i = T_i \vee C_i$ and $\delta_i = I(C_i \leqslant T_i)$, where '$\vee$' denotes the maximum. The direction of the time axis is simply reversed in all four estimators. However, the estimators of Cox and Koul *et al.* cannot be applied when there are both left and right censoring in the data. Miller's and Buckley & James's estimators can be extended to this situation by utilizing the distribution estimator of Turnbull (1974, 1976).

2·2. *Cox's estimator*

Cox (1972, 1975) proposed a partial likelihood approach to estimating β since the function $\lambda_0(t)$ being unknown prevents a full likelihood analysis. The patients in the risk set $\mathscr{R}(t)$ are those still alive and in the study at time $t-$. If it is known that a patient dies at time t, then the conditional probability that it is patient i among those at risk is $\exp(x_i\beta)/\Sigma_{j \in \mathscr{R}(t)} \exp(x_j\beta)$. If $y_{(1)} < ... < y_{(n)}$ are the ordered observations, then the partial likelihood is

$$L_c = \prod_{i=1}^{n} \left(e^{x_{(i)}\beta} \bigg/ \sum_{j \in \mathscr{R}(y_{(i)})} e^{x_j\beta} \right)^{\delta_{(i)}}, \tag{2}$$

where $x_{(i)}$, $\delta_{(i)}$ are associated with $y_{(i)}$. The value of β maximizing (2) is obtained by solving on a computer for the root of the equations $\partial \log L_c/\partial \beta = 0$.

Expression (2) for the partial likelihood requires that the uncensored survival times be all distinct. Several modifications to account for ties have been suggested in the literature. The one used in this paper is the approximate likelihood called by Peto (1972) P_{rough}, and also described by Kalbfleisch & Prentice (1972, (iv)) and Breslow (1974, (8)). Ties between uncensored and censored observations are handled by considering the uncensored survival times as preceding the censored ones.

Tsiatis (1981) and K. Bailey in a 1979 University of Chicago doctoral dissertation have proved the asymptotic normality of $\hat{\beta}$, the value of β maximizing (2). The asymptotic covariance matrix is the matrix of second partial derivatives of the logarithm (2) or its modification for ties.

Efron (1977) and Oakes (1977) have established that Cox's estimator is nearly fully efficient.

There have been various proposals for estimating the survival function

$$S(t; x) = 1 - F(t; x) = \exp \left\{ -e^{x\beta} \int_0^t \lambda_0(u) \, du \right\}. \tag{3}$$

The method adopted in this paper is to estimate $\lambda_0(t)$ by Breslow's (1974) estimator

$$\hat{\lambda}_0(t) = d_{u(i)}\left\{(y_{u(i)} - y_{u(i-1)}) \sum_{j \in \mathcal{R}(y_{u(i)})} e^{x_j \beta}\right\}^{-1}$$

for $y_{u(i-1)} < t < y_{u(i)}$, where $y_{u(1)} < y_{u(2)} < \dots$ are the ordered distinct uncensored observations and $d_{u(i)}$ is the number of deaths at $y_{u(i)}$, and then to substitute $\hat{\lambda}_0(t)$ and $\hat{\beta}$ into (3). This estimator $\hat{S}(t; x)$, which was proposed by C. Link in a 1979 Stanford University technical report and Tsiatis (1981), differs slightly from the Kaplan & Meier (1958) product-limit estimator if $\hat{\beta} = 0$.

2·3. *Miller's estimator*

Miller (1976) suggested minimizing the sum of squares

$$n \int e^2 \, d\hat{F}(e; a, b) \tag{4}$$

with respect to a and the vector b, where $\hat{F}(e; a, b)$ is the product-limit estimator based on δ_i, $\hat{e}_i = \hat{e}_i(a, b) = y_i - a - x_i b$ for $i = 1, \dots, n$. Specifically,

$$1 - \hat{F}(e; a, b) = \prod_{\hat{e}_{(i)} \leqslant e} (1 - d_{(i)}/n_{(i)})^{\delta_{(i)}},$$

where $\hat{e}_{(1)} < \hat{e}_{(2)} < \dots$ are the ordered distinct values of $\hat{e}_i$, $n_{(i)}$ is the number at risk at $\hat{e}_{(i)}-$, $d_{(i)}$ is the number dying at $\hat{e}_{(i)}$, and $\delta_{(i)} = 1$ if $d_{(i)} > 0$, $= 0$ otherwise. Expression (4) is a generalization of the usual sum of squares $\Sigma (y_i - a - x_i b)^2$ for uncensored data.

It is difficult to locate the infimum of (4) because it is a discontinuous function of b. Therefore, Miller proposed using an iterative sequence to calculate the estimate of the regression coefficient vector β:

$$\hat{\beta}_{k+1} = \{(X - \bar{X}^w)^{\mathrm{T}} W(\hat{\beta}_k) (X - \bar{X}^w)\}^{-1} (X - \bar{X}^w) W(\hat{\beta}_k) y, \tag{5}$$

where

$$X = ((x_{ij})), \quad \bar{X}^w = \left(\left(\sum_{i=1}^{n} w_i(\hat{\beta}_k) x_{ij}\right)\right),$$
$$W(\hat{\beta}_k) = \mathrm{diag}\{w_i(\hat{\beta}_k)\}, \quad y = (y_1, \dots, y_n)^{\mathrm{T}}. \tag{6}$$

The limit of the sequence $\hat{\beta}_k$ for $k = 0, 1, \dots$, is the estimate of β. Because of discontinuities in the weights as functions of $\hat{\beta}_k$, the sequence may become trapped in a loop. If the values in the loop are not far apart, an average value over the loop can be used for the estimate.

The weight $w_i(\hat{\beta}_k)$ in (6) is the size of the jump assigned to $\hat{e}_i^0 = \hat{e}_i(0, \hat{\beta}_k) = y_i - x_i \hat{\beta}_k$ by the Kaplan & Meier estimator applied to $\hat{e}_1^0, \dots, \hat{e}_n^0$. Inclusion of an intercept estimate $\hat{\alpha}_k$ in the estimate of e_i is unnecessary since the weights are invariant under location shifts in the data. If the largest $\hat{e}_i^0$ is censored, the mass $1 - \hat{F}(+\infty; 0, \hat{\beta}_k)$ is unassigned to any $\hat{e}_i^0$. The convention which seems to work best for these estimators is to normalize the weights so that they sum to one in (6) and (8), but to assign the remaining mass $1 - \hat{F}(+\infty; 0, \hat{\beta}_k)$ to the largest $\hat{e}_i^0$ in (7).

Only the uncensored y_i received nonzero weights in (5). For this reason it makes sense to use as a starting value $\hat{\beta}_0$ the ordinary unweighted least squares estimator applied to just the uncensored data.

For the limiting value $\hat{\beta}$ the associated estimate of the intercept is

$$\hat{\alpha} = \sum_{i=1}^{n} w_i(\hat{\beta}) \, (y_i - x_i \hat{\beta}). \qquad (7)$$

If the variation in $\hat{\beta}$ due to an estimate of β rather than the true β being used in the computation of the weights is ignored, then an estimate of the covariance matrix for $\hat{\beta}$ is

$$\left\{ \sum_{i=1}^{n} w_i^2(\hat{\beta}) \, (y_i - \hat{\alpha} - x_i \hat{\beta})^2 \right\} \{ (X - \bar{X}^w)^{\mathrm{T}} \, W(\hat{\beta}) \, (X - \bar{X}^w) \}^{-1}. \qquad (8)$$

For $\hat{\beta}$ to be consistent it is necessary that the censoring distributions $G(t; x)$ satisfy

$$G(t; x) = G(t - \alpha - x\beta). \qquad (9)$$

This assumption requires that, as x changes, the censoring distributions shift along the same line as the survival distributions.

2·4. *Buckley & James's estimator*

The Buckley & James (1979) estimator exploits the following linear relationship:

$$E\{ \delta_i \, Y_i + (1 - \delta_i) \, E(T_i \,|\, T_i > Y_i) \,|\, x_i \} = \alpha + x_i \beta. \qquad (10)$$

Buckley & James substitute an estimate for the conditional expectation $E(T_i \,|\, T_i > y_i)$ based on the Kaplan & Meier estimator into the variable $y_i = \delta_i y_i + (1 - \delta_i) E(T_i \,|\, T_i > y_i)$ and then solve the usual least squares normal equations iteratively.

Specifically, if $\delta_i = 1$, let $\hat{y}_i(\hat{\beta}_k) = y_i$, but if $\delta_i = 0$, let

$$\hat{y}_i(\hat{\beta}_k) = x_i \hat{\beta}_k + \left\{ \sum_{j: \hat{e}_j^0 > \hat{e}_i^0} w_j(\hat{\beta}_k) \, \hat{e}_j^0 \right\} \bigg/ \{ 1 - \hat{F}(\hat{e}_i^0; \, 0, \hat{\beta}_k) \}.$$

The convention in computing the Kaplan & Meier estimator $\hat{F}$ is to always assign the remaining mass to the largest $\hat{e}_j^0$ if it is censored. The regression estimator $\hat{\beta}_{k+1}$ at the $(k+1)$st step is the usual least squares estimator

$$\hat{\beta}_{k+1} = \{ (X - \bar{X})^{\mathrm{T}}(X - \bar{X}) \}^{-1}(X - \bar{X})^{\mathrm{T}} \, \hat{y}(\hat{\beta}_k), \qquad (11)$$

where the matrix $\bar{X}$ has elements $n^{-1} \Sigma_i x_{ij}$ and $\hat{y}(\hat{\beta}_k) = \{ \hat{y}_1(\hat{\beta}_k), \ldots, \hat{y}_n(\hat{\beta}_k) \}^{\mathrm{T}}$. The iteration is continued until $\hat{\beta}_k$ converges to a limiting value $\hat{\beta}$ or becomes trapped in a loop like the Miller estimator.

Since the estimator (11) uses a value for the dependent variable at every x_i, it seems sensible to take for the starting $\hat{\beta}_0$ the least squares estimator

$$\{ (X - \bar{X})^{\mathrm{T}}(X - \bar{X}) \}^{-1}(X - \bar{X})^{\mathrm{T}} y$$

which treats all the observations as uncensored whether they are uncensored or not.

For the limiting value $\hat{\beta}$ the associated estimate of the intercept is

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^{n} \{ \hat{y}_i(\hat{\beta}) - x_i \hat{\beta} \}.$$

In spirit this technique is a nonparametric version of the EM algorithm introduced by Dempster, Laird & Rubin (1977). The censored data points are replaced by their expectations and then the sum of squares is minimized.

In estimating the variability Buckley & James recommend using

$$\hat{\sigma}_u^2 = \frac{1}{n_u - 2} \sum_{i=1}^{n} \delta_i \left(\hat{e}_i^0 - \frac{1}{n_u} \sum_{i=1}^{n} \delta_j \hat{e}_j^0 \right)^2,$$

where n_u is the number of uncensored observations. Their estimator for the covariance matrix of $\hat{\beta}$ is

$$\hat{\sigma}_u^2 \{ (X - \bar{X}^u)^{\mathrm{T}} \Delta (X - \bar{X}^u) \}^{-1}, \tag{12}$$

where $\bar{X}^u$ is the matrix with elements $n_u^{-1} \Sigma_i \delta_i x_{ij}$ and $\Delta = \mathrm{diag}\,(\delta_i)$. There is no firm theoretical foundation for (12), but its validity has been empirically substantiated through Monte Carlo simulations.

In unpublished work James & Buckley have compared the efficiency of their estimator to the maximum likelihood estimator for normal and nonnormal distributions.

2·5. *Koul, Susarla & Van Ryzin's estimator*

Koul *et al.* (1981) based an estimator on a linear relationship different from (10), namely

$$E[\delta_i\, Y_i \{ 1 - G(Y_i;\, x_i) \}^{-1} \mid x_i] = \alpha + x_i \beta. \tag{13}$$

Under the assumption $G(t;\, x_i) \equiv G(t)$, Koul *et al.* substitute an estimate for $G(t)$ into the variable $y_i = \delta_i y_i \{ 1 - G(y_i) \}^{-1}$ and then solve the usual least squares normal equations. The relation (13) does not require $G(t;\, x_i)$ to be independent of x_i, but there is no way of estimating each $G(t;\, x_i)$ from the data without imposing assumptions on $G(t;\, x)$ as a function of x.

The product-limit estimator with the roles of t_i and c_i reversed could be used to estimate the common censoring distribution $G(t)$. However, Koul *et al.* suggest the following estimator which emanates from Bayesian considerations:

$$1 - \hat{G}(t) = \prod_{y_i \leqslant t} \left\{ \frac{1 + n^+(y_i)}{2 + n^+(y_i)} \right\}^{1 - \delta_i}, \tag{14}$$

where $n^+(y)$ is the number of y_j greater than y. Ties between censored and uncensored observations are handled in the usual fashion with the roles of censored and uncensored being reversed.

The product-limit estimator and the estimator (14) can be unstable for large values of t, because the data are so sparse for large t. Therefore, Koul *et al.* suggest truncating the large observations by defining $\hat{y}_i = \delta_i y_i \{ 1 - \hat{G}(y_i) \}^{-1} I(y_i \leqslant M_n)$, where I is the indicator function. They make some remarks on the selection of M_n in their paper.

The great advantage of the estimator of Koul *et al.* is that no iteration is required. Specifically,

$$\hat{\beta} = \{ (X - \bar{X})^{\mathrm{T}} (X - \bar{X}) \}^{-1} (X - \bar{X})^{\mathrm{T}} \hat{y}, \quad \hat{\alpha} = n^{-1} \Sigma_i (\hat{y}_i - x_i \hat{\beta}).$$

Standard least squares computer routines can be used once the vector $\hat{y}$ has been computed.

Koul *et al.* derive complicated integral expressions for the asymptotic covariance matrix of $\hat{\beta}$.

2·6. Other estimators

A computer routine which minimizes the sum of squares for all the uncensored observations and only the censored ones falling above the estimated regression line has been introduced by J. Friedman and W. Stuetzle in a 1981 Stanford University technical report.

For normal linear models Schmee & Hahn (1979) introduced an estimator which is identical in spirit to Buckley & James's estimator. The difference is that Schmee & Hahn base the computation of the conditional expectation $E(T \,|\, T > y)$ in (10) on the assumption of a normal distribution. Nelson & Hahn (1972, 1973) treated the case where there are replicate observations for each value of the independent variable. Linear combinations of order statistics are used to estimate the normal parameters. Amemiya (1973) applied maximum likelihood theory to the normal linear model with censoring at a fixed upper, or lower, level, type I censoring. Amemiya extended earlier work by Tobin (1958) and Glasser (1965). Heckman (1976) gave an *ad hoc* two-stage estimator in the normal linear model.

Regression estimators with censored data have also been proposed for the exponential distribution where $\lambda(t; x) \equiv \lambda(x)$ for $t > 0$. Zippin & Armitage (1966) treated the linear model $\lambda(x) = (\alpha + x\beta)^{-1}$ in the special case $p = 1$, and Mantel & Myers (1971) studied the multidimensional case. Glasser (1967) considered the proportional hazards model $\lambda(x) = \lambda_0 \, e^{x\beta}$ in the special case $p = 1$. All three papers are based on the maximum likelihood approach.

3. RESULTS

The Stanford heart transplantation program was begun in October 1967. By February 1980, 184 patients had received heart transplants. A few of these had multiple transplants. Their survival times, uncensored or censored in February 1980, are displayed in Table 1 along with their ages at the time of the first transplant. The patients' T5 mismatch scores which measure the degree of tissue incompatability between the initial donor and recipient hearts with respect to HLA antigens are also included in Table 1.

The data on patients who were admitted to the program but did not receive transplants and the pretransplant waiting times for patients who did receive transplants were not collected for this study. Because of this there is no analysis of whether or not transplantation prolongs survival.

In analysing the T5 mismatch scores on the data up to March 1974 Miller (1976) and Crowley & Hu (1977) made a distinction between deaths primarily due to rejection of the donor heart by the recipient's immune system and nonrejection related deaths. This distinction is no longer maintained because histological information is not always available and the judgement of rejection death is largely subjective.

For 27 of the 184 transplant patients the T5 mismatch scores are missing because the tissue typing was never completed. For the regression analyses reported here just the 157 patients with complete records were used. Of these 157 patients 55 were still alive, i.e. were censored as of February 1980, and 102 were deceased, i.e. were uncensored.

Table 2 gives the estimated coefficients and their estimated standard deviations in the multiple regression of the base 10 logarithm of the time to death against age and T5 mismatch score. For the Buckley–James and Miller estimators the initial estimates, $\hat{\beta}_0$, the one-step estimates, $\hat{\beta}_1$, and the final estimates, $\hat{\beta}$, are all given. No truncation was

used in computing the estimates of Koul *et al.* and no standard deviations could be calculated for them.

Due to its nonsignificance the T5 mismatch score was deleted from the analysis. The estimated regression coefficients and standard deviations with age alone for the different methods are practically identical to those for age in Table 2.

Figure 1(a) graphs the survival times against the ages of the patients at transplant. The dashed line is the Buckley–James estimated regression line. The solid line is the estimate of the median survival time from Cox's method. That is, at each x the value of t for which $\hat{S}(t; x)$ equals $\frac{1}{2}$ is plotted. The median estimator was selected because it is much easier to compute at each point than the mean estimator. Also, since the time scale is irrelevant to Cox's method the median estimator seems more appropriate.

From Table 2 and other analyses not presented here we concluded that the Cox and Buckley–James estimators are the superior estimators. Full discussion of this will be given in the next section. However, the fits of these two models to the data are not ideal. This is particularly evident in Fig. 1(a) for Cox's model at low values of x where the median estimator exceeds all the data. Also, the Cox and Buckley–James estimators do not agree on the statistical significance of the age effect. The estimated slope from Cox's method is more than 2·5 standard deviations away from zero, whereas that for Buckley & James's method is less than 2 standard deviations. Note that the two estimators have opposite signs because of the parameterizations.

In an attempt to achieve better fit a quadratic age model $\beta_1(\text{age}) + \beta_2(\text{age})^2$ was tried. In this analysis the 5 patients with survival times less than 10 days were deleted. The logarithmic transformation was not perfect in symmetrizing the data, and it was felt that these few very low values might have distorted the Buckley–James regression estimates and inflated the error variance. The results for the Cox and Buckley–James methods applied with the quadratic model are presented in Table 3. The data and the fitted regression functions are displayed in Fig. 1(b).

The agreement between the two estimation methods is very much better for the quadratic model. The degrees of significance of the linear and quadratic regression coefficients are very similar for the two estimators. Also, the mean and the median regression functions fit the data very much better. The estimates for the extreme younger and older ages are more reasonable.

Can any distinction be drawn between the proportional hazards model and the linear model? Does one fit the data better than the other? To examine this question we plotted the ordinary residuals for the linear model and the generalized residuals for the proportional hazards model. Neither graph indicated any serious model deficiency. The linear residuals have a nonpronounced U-shape, which indicates some slight model irregularity, but both models fit rather well.

4. Discussion

From their performances on the updated Stanford heart transplant data we feel that the Cox and Buckley & James estimators are the two most reliable regression estimators for use with censored data. The other two estimators, those of Koul *et al.* and of Miller, have methodological weaknesses.

The estimator of Koul *et al.* is based on the assumption that the censoring distribution $G(t)$ does not vary with x. The peculiarity of the dependent variable values used in this technique makes it sensitive to departures from this assumption. Each censored

Table 1. *Stanford heart transplant data February* 1980; *first column, patient number; second column, T, survival time (days); third column, dead* = 1, *alive* = 0; *fourth column, age at time of first transplant; fifth column,* T5 *mismatch score*

No.	T	Dead/alive	Age	T5	No.	T	Dead/alive	Age	T5	No.	T	Dead/alive	Age	T5
1	15	1	54	1·11	51	1384	1	46	1·41	101	1393	0	46	0·95
2	3	1	40	1·66	52	544	1	52	1·94	102	1202	1	38	9999
3	46	1	42	0·61	53	29	1	53	1·08	103	1378	0	41	1·65
4	623	1	51	1·32	54	48	1	53	3·05	104	1373	0	41	1·38
5	126	1	48	0·36	55	297	1	42	0·60	105	274	1	31	0·58
6	64	1	54	1·89	56	1318	1	48	1·44	106	31	1	33	0·36
7	1350	1	54	0·87	57	1352	1	54	0·68	107	1341	0	50	1·13
8	23	1	56	2·05	58	50	1	46	2·25	108	42	1	19	0·63
9	279	1	49	1·12	59	547	1	49	0·81	109	381	1	45	0·98
10	1024	1	43	1·13	60	431	1	47	0·33	110	1264	0	52	0·64
11	10	1	56	2·76	61	68	1	51	1·33	111	1262	0	34	1·68
12	39	1	42	1·38	62	26	1	52	0·82	112	1261	0	47	0·82
13	730	1	58	0·96	63	161	1	43	1·20	113	47	1	36	0·16
14	1961	1	33	1·06	64	14	1	40	9999	114	1193	0	24	1·15
15	136	1	52	1·62	65	2313	0	26	0·46	115	626	1	53	1·74
16	1	1	54	0·47	66	1634	1	23	1·78	116	48	1	51	0·99
17	836	1	44	1·58	67	146	1	45	0·16	117	1150	1	32	2·25
18	60	1	64	0·69	68	48	1	28	0·77	118	45	1	48	0·65
19	3695	0	40	0·38	69	2127	1	35	0·67	119	1116	0	14	0·54
20	1996	1	49	0·91	70	263	1	49	0·48	120	1107	0	18	0·25
21	0	1	41	0·87	71	2106	0	40	0·86	121	1102	0	39	1·35
22	47	1	62	0·87	72	293	1	43	0·70	122	195	1	39	0·73
23	54	1	49	2·09	73	2025	0	30	1·44	123	30	1	34	0·84
24	51	1	50	9999	74	2006	0	15	1·26	124	1040	0	43	0·50
25	2878	1	49	0·75	75	2000	0	45	1·46	125	993	0	30	0·95
26	3410	0	45	0·98	76	1995	0	47	1·65	126	950	0	46	9999
27	44	1	36	0·0	77	1945	0	38	1·28	127	729	1	49	1·10
28	994	1	48	0·81	78	65	1	55	0·69	128	121	1	45	9999
29	51	1	47	1·38	79	731	1	38	0·42	129	202	1	48	1·24
30	1478	1	36	1·35	80	1866	0	49	0·51	130	841	0	48	0·86
31	254	1	48	1·08	81	538	1	49	2·76	131	834	0	49	9999
32	897	1	46	9999	82	1846	0	44	0·83	132	265	1	49	1·22
33	148	1	47	9999	83	68	1	35	0·85	133	1	1	21	0·47
34	51	1	52	1·51	84	1778	0	27	0·70	134	793	0	19	1·98
35	323	1	48	1·82	85	1722	0	40	0·95	135	328	1	34	1·02
36	3021	0	38	0·98	86	928	1	50	1·12	136	781	0	20	1·12
37	66	1	49	0·66	87	1718	0	39	1·77	137	752	0	43	1·50
38	2984	0	32	0·19	88	22	1	27	1·64	138	738	0	41	0·53
39	2723	1	32	1·93	89	40	1	42	1·59	139	86	1	12	1·26
40	550	1	48	0·12	90	7	1	28	1·00	140	132	1	46	1·09
41	66	1	51	1·12	91	1638	0	48	0·43	141	663	0	36	0·47
42	65	1	45	1·68	92	1612	0	51	1·25	142	660	0	42	0·75
43	227	1	19	1·02	93	25	1	52	0·53	143	221	1	35	1·04
44	2805	0	48	1·20	94	1534	1	44	1·71	144	90	1	38	1·00
45	25	1	53	1·68	95	1547	0	50	0·18	145	619	0	47	0·90
46	631	1	26	1·46	96	1271	1	32	1·05	146	618	0	50	0·82
47	2734	0	47	0·97	97	44	1	46	1·71	147	576	0	53	2·25
48	12	1	29	0·61	98	1247	1	41	0·43	148	563	0	41	9999
49	63	1	56	2·16	99	1232	1	18	0·70	149	36	1	45	0·20
50	2474	1	52	1·70	100	191	1	42	1·74	150	549	0	40	2·53

Table 1 (*cont.*)

No.	T	Dead/ alive	Age	T5	No.	T	Dead/ alive	Age	T5	No.	T	Dead/ alive	Age	T5
151	548	0	30	0·47	161	136	1	55	9999	171	231	0	52	9999
152	541	0	47	0·43	162	406	0	39	1·18	172	145	1	50	0·96
153	534	0	20	9999	163	391	0	27	1·17	173	188	0	52	9999
154	169	1	51	1·89	164	374	0	47	9999	174	176	0	29	1·72
155	122	1	51	1·33	165	50	1	50	0·50	175	138	1	41	9999
156	382	1	36	9999	166	139	1	51	0·96	176	149	0	21	9999
157	468	0	24	1·39	167	322	0	36	1·73	177	119	0	20	9999
158	464	0	38	2·07	168	292	0	43	1·40	178	107	0	46	9999
159	10	1	13	1·49	169	278	0	41	0·98	179	98	0	19	9999
160	5	1	20	9999	170	22	1	45	9999	180	89	0	27	9999
										181	60	0	13	9999
										182	56	0	27	9999
										183	2	0	39	9999
										184	1	0	27	9999

For T5 mismatch score 9999 denotes missing value.

Table 2. *Regression estimates and standard deviations,* SD, *for* $\log_{10}$ *of time to death versus age at transplant and* T5 *mismatch score with* $n = 157$ *Stanford heart transplant patients*

Estimator	Intercept $\hat{\alpha}$	SD $(\hat{\alpha})$	Age $\hat{\beta}_1$	SD $(\hat{\beta}_1)$	T5 $\hat{\beta}_2$	SD $(\hat{\beta}_2)$
Cox			0·030	0·011	0·167	0·183
Buckley–James						
Initial	2·78		−0·007		−0·034	
One-step	3·14		−0·013		−0·011	
Final	3·23	0·35	−0·015	0·008	−0·003	0·134
Miller						
Initial	2·03		0·001		0·061	
One-step	2·58		−0·002		0·060	
Final†	2·57	0·51	−0·001	0·011	0·072	0·191
	2·54	0·36	0·000	0·008	0·040	0·135
Koul-Susarla–Van Ryzin	0·72		0·024		0·251	

† Loop with two values.

observation is decreased to zero, and each uncensored observation is inflated by the factor $\{1 - \hat{G}(y_i)\}^{-1}$. In some cases the observations are being moved away from the regression line in order to estimate it. With the Stanford heart transplant data the proportionately larger number of censored observations in the younger ages creates a proportionately larger number of zeros for small x. This gives a positive slope to the estimated regression line whereas everyone feels that increasing age is deleterious and the slope should be negative. The implication for other data sets is that if the censored observations are not spread evenly over the ranges of the independent variables they can tip the estimated regression plane in a false direction or unduly enhance its slope.
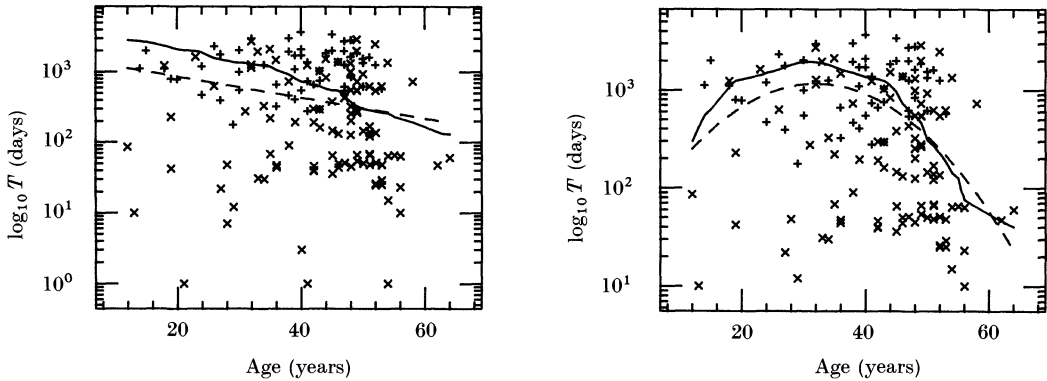
Fig. 1. (a) (left). Scatterplot of $\log_{10}$ survival time (days) versus age at transplant (years) for 157 Stanford heart transplant patients. Patients denoted by × are deceased; those by + were alive in February 1980. Dashed line, Buckley & James linear regression line; solid line, Cox proportional hazards regression median.

(b) (right). As Fig. 1(a) but for quadratic regression with 152 patients who survived at least 10 days.

Table 3. *Regression estimates and standard deviations for* $\log_{10}$ *of time to death versus age and age squared at time of transplant with* $n = 152$ *Stanford heart transplant patients who survived at least* 10 *days*

Estimator	Intercept		Age		Age²	
	$\hat{\alpha}$	SD $(\hat{\alpha})$	$\hat{\beta}_1$	SD $(\hat{\beta}_1)$	$\hat{\beta}_2$	SD $(\hat{\beta}_2)$
Cox			-0.146	0.055	0.0023	0.0007
Buckley–James	1.35	0.71	0.107	0.037	-0.0017	0.0005

Also, while integral expressions have been obtained for the asymptotic variances and covariances of Koul *et al.* estimators when $p > 1$, computational expressions are currently available only when $p = 1$. Also, further guidelines and experience are needed in the proper choice of the truncation point M_n.

Miller's estimator can also be thrown off by the censoring pattern. The asymptotic consistency of the estimator requires condition (9) on the censoring distributions, and this condition is not satisfied for the Stanford heart transplant data. While the Miller estimator does not seem to be biased so much by violations of (9) as the estimator of Koul *et al.* by violations of $G(t; x) \equiv G(t)$, it is influenced by the censoring pattern.

Both the Buckley & James and Miller iterative sequences of estimators can become trapped in loops and fail to converge. This happened in Table 1 for the Miller estimator. Our studies show that the loops are less frequent and less severe for the Buckley & James estimator than the Miller estimator. Why this should be the case was discussed by Buckley & James (1979).

In conclusion, the Cox and Buckley & James estimators are not dependent on particular censoring patterns and have proved to be reliable estimators. Theoretical validation is lacking for Buckley & James's variance estimator, but use of it seems justified since it gives empirically sensible results and is supported by Monte Carlo studies. Both the Cox and Buckley & James estimators require about the same amount of programming and computing time. The choice between them should depend on the appropriateness of the proportional hazards model or the linear model for the data.

REFERENCES

AMEMIYA, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica* **41**, 997–1016.

BRESLOW, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

BUCKLEY, J. & JAMES, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429–36.

COX, D. R. (1972). Regression models and life-tables (with discussion). *J. R. Statist. Soc.* B **34**, 187–202.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.

CROWLEY, J. & HU, M. (1977). Covariance analysis of heart transplant survival data. *J. Am. Statist. Assoc.* **72**, 27–36.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B **39**, 1–22.

EFRON, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Am. Statist. Assoc.* **72**, 557–65.

GLASSER, M. (1965). Regression analysis with dependent variable censored. *Biometrics* **21**, 300–7.

GLASSER, M. (1967). Exponential survival with covariance. *J. Am. Statist. Assoc.* **62**, 561–8.

HECKMAN, J. J. (1976). The common structure of statistical methods for truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Ann. Econ. Soc. Meas.* **5**, 475–92.

KALBFLEISCH, J. & PRENTICE, R. L. (1972). Contribution to discussion of paper by D. R. Cox. *J. R. Statist. Soc.* B **34**, 215–6.

KAPLAN, E. L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.* **53**, 457–81.

KOUL, H., SUSARLA, V. & VAN RYZIN, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist.* **9**, 1276–88.

MANTEL, N. & MYERS, M. (1971). Problems of convergence of maximum likelihood iterative procedures in multiparameter situations. *J. Am. Statist. Assoc.* **66**, 484–91.

MILLER, R. G. (1976). Least squares regression with censored data. *Biometrika* **63**, 449–64.

NELSON, W. & HAHN, G. J. (1972). Linear estimation of a regression relationship from censored data. Part I—Simple methods and their applications. *Technometrics* **14**, 247–69.

NELSON, W. & HAHN, G. J. (1973). Linear estimation of a regression relationship from censored data. Part II—Best linear unbiased estimation and theory. *Technometrics* **15**, 133–50.

OAKES, D. (1977). The asymptotic information in censored survival data. *Biometrika* **64**, 441–8.

PETO, R. (1972). Contribution to discussion of paper by D. R. Cox. *J. R. Statist. Soc.* B **34**, 205–7.

SCHMEE, J. & HAHN, G. J. (1979). A simple method for regression analysis with censored data. *Technometrics* **21**, 417–32.

TOBIN, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24–36.

TSIATIS, A. (1981). A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93–108.

TURNBULL, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Am. Statist. Assoc.* **69**, 169–73.

TURNBULL, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Statist. Soc.* B **38**, 290–5.

ZIPPIN, C. & ARMITAGE, P. (1966). Use of concomitant variables and incomplete survival information in the estimation of an exponential survival parameter. *Biometrics* **22**, 665–72.