DOI: http://dx.doi.org/10.24327/ijcmes2017.07100023

Available online at http://journalijcmes.com

INTERNATIONAL JOURNAL OF CURRENT MEDICAL SCIENCES

Research Article

ISSN: 2320- 8147

International Journal of Current Medical Sciences- Vol. 7, Issue, 10, pp. 313-319, October, 2017

COX'S PROPORTIONAL HAZARDS MODEL FOR HEART TRANSPLANT DATA

Christuraja R* and Ravichandran M.K

Department of Statistics, Annamalai University, Annamalainagar – 608002

ARTICLE INFO

Article History:

Received 20th July, 2017
Received in revised form 15th August, 2017
Accepted 27th September, 2017
Published online 28th October, 2017

Key words:

Hazards model, Heart transplant data,

ABSTRACT

Long before human-to-human transplantation was ever imagined by the public, scientists were conducting pioneering medical and surgical research that would eventually lead to today's transplantation successes. The Cox proportional hazards model has become the model of choice in the analysis of time to event data in survival analysis. It is evident that Cox proportional hazard model is not always appropriate and the Accelerated Failure Time (AFT) model provides a better alternative for variable selection in survival analysis. If the effects of treatment are to accelerate (or delay) the event of interest rather than having a longer term impact, in the context of the trial duration, on the occurrence of the event, then the accelerated failure time model should replace the proportional hazards model as the model of choice. For a detailed study refer to Kay and Kinnersley (2002). In this paper it is proposed to study the Cox's Proportional Hazards and AFT Model for Heart Transplantation Data. Numerical examples are also provided.

INTRODUCTION

The Cox proportional hazards model has become the model of choice in the analysis of time to event data in survival analysis. It is evident that Cox proportional hazard model is not always appropriate and the Accelerated Failure Time (AFT) model provides a better alternative for variable selection in survival analysis. If the effects of treatment are to accelerate (or delay) the event of interest rather than having a longer term impact, in the context of the trial duration, on the occurrence of the event, then the accelerated failure time model should replace the proportional hazards model as the model of choice. For a detailed study refer to Kay and Kinnersley (2002).

Let 'T' be the 'failure time', the response variable, and x be the corresponding covariate vector. Suppose that we are interested in making inferences about the effect from x on the response variable T. If there is no censored observation at the time of the analysis, most likely we would regress T or a transformation of it directly on the covariate x. this approach is quite appealing to practitioners owing to its ease of interpretation. However, if there are censored observations in the data, we usually do not take this conventional approach to examine the covariate effect. In fact, we almost exclusively use the proportional hazards model with the partial likelihood principle to draw inferences about the covariate effect, refer to Cox (1972). However, since the proportional hazard model specifies that the effect of the covariate x is to act multiplicatively on the hazard function; it is not easy to interpret, for example, the estimates of regression parameters. So, if the ordinary linear regression model can handle censored observations, it would be a useful alternative to the Cox model in the survival analysis. In fact, with the presence of censored observations, the linear regression analysis with log T as the response variable has been studied extensively by many authors. The linear regression model with censored observation suggested by Miller (1976), Prentice (1978), Buckley and James (1979) and Koul *et. al.* (1981) with the log-transformation is called the AFT model. For a more detailed discussion on AFT, refer to Kalbfleisch and Prentice (1980). Miller and Halpern (1982) has been discussed the four regression techniques for censored data as suggested by Cox (1972), Miller (1976), Buckley and James (1979) and Koul *et. al.* (1981). These four methods are compared based on the updated Stanford heart transplant data. Computational expression are only available when p=1 in the case of Koul *et. al*. (1982). The expressions for the asymptotic

Corresponding author:* **Christuraja R
Department of Statistics, Annamalai University, Annamalainagar – 608002

variances and co-variances have been obtained for Koul *et. al*. estimators when p>1.

The Cox, Buckley and James estimators are not dependent on particular censoring patterns and have proved to be reliable estimators. Theoretical validation is lacking for Buckley and 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. Unfortunately, the original Buckley – James method has no theoretical justification and does not provide a reliable numerical method for implementation. Therefore, it has seldom been used in practice. AFT model has been discussed by many authors cited in the recent literature. Long before human-to-human transplantation was ever imagined by the public, scientists were conducting pioneering medical and surgical research that would eventually lead to today's transplantation successes. In this paper it is proposed to study the Cox's Proportional Hazards and AFT Model for Heart Transplantation Data. Numerical examples are also provided.

Regression Techniques

The four regression techniques are available for the use with censored data which do not assume particular parametric families of survival distributions are listed under the following heads

1. Cox's approach (1972)
2. Miller's approach (1976)
3. Buckley-James approach (1979)
4. Koul, Susarla and Van Ryzin approach (1981)

Cox's Approach

Suppose that on each individual one or further measurements are available, say on variables $z_1,\ldots,z_p$. Assume that the failure-times are continuously distributed and the possibility of ties can be ignored. For the j^{th} individual let the values of z be $z_j = (z_{ij},\ldots,z_{pj})$. The z's may be function of time. The proportional hazards models suggested by Cox (1972) is of the form,

$$\lambda(t;z) = \exp(z\beta)\lambda_0(t) \quad \ldots (1)$$

where β is a p × 1 vector of unknown parameters and $\lambda_0(t)$ is an unknown function giving the hazard function for the standard set of conditions z = 0.

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 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 $\exp(x_i)/\sum_{jR(t)}\exp(x_j\beta)$ If $y_{(1)} < \ldots < y_{(n)}$ are the ordered observations, then the partial likelihood is

$$L_c = \prod_{i=1}^{n}\left(e^{x_i\beta}/\sum_{jR(y_{(i)})}e^{x_j\beta}\right)^{\delta_{(i)}} \quad \ldots (2)$$

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

There have been various proposals for estimating the survival function

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

To estimate $\lambda_0(t)$ by using Breslow's (1974), the estimator

$$\lambda_0(t) = d_{u(i)}\left\{(y_{u(i)} - y_{u(i-1)})\sum_{jR(y_{(i)})}e^{x_j\bar{\beta}}\right\}^{-1}$$

for $y_{u(i-1)} < t < y_{u(i)}$, where $y_{u(1)} < y_{u(2)} < \ldots$ are the ordered distinct uncensored observations and $d_{u(i)}$ is the number of deaths at $y_{u(i)}$, and then to substitute $\lambda_0(t)$ and $\hat{\beta}$ into equ. (3). This estimator, $\widehat{S}(t;x)$ which was proposed by Link (1979) and Tsiatis (1981).

Miller's Approach

The linear regression model for n pairs of variates (x_i, y_i) (i = 1, ...,n) is

$$y_i = \alpha + \beta x_i + e_i \quad \ldots (4)$$

where the random variables e_i are assumed to be independent and identically distributed according to the distribution F. Sometimes the condition that F be the normal distribution is added to the model, but it is only assumed that F has mean μ= 0 and finite variance σ^2. The estimation of α and β when the y_i are subject to censoring. The primary type of censorship to be considered is single censoring on the right, i.e. the observable variable is not y_i but

$$Y_i = \begin{cases} y_i & (y_i \le c_i) \\ y_i^+ = c_i & (y_i > c_i) \end{cases} \quad \ldots (5)$$

It is known whether each observation is censored or uncensored, and the presence or absence of a ' + ' on y_i denotes this. When it is necessary to specify the structure of the c_i, it will be assumed that the c_i are random variables independently distributed of each other and the e_i, but not necessarily of the x_i.

Miller's Estimator

Miller (1976) suggested minimizing the sum of squares

$$n\int e^2 d\widehat{F}(e;\,a,\,b) \quad \ldots (6)$$

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 ,...,n. Specifically,

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

where $\hat{e}_{(1)} < \hat{e}_{(2)} < \ldots$ are the ordered distinct values of $\hat{e}_i$, $n_{(i)}$ is the number at risk at $\hat{e}_{(1)} - \hat{e}_{(2)}$ is the number dying at $\hat{e}_{(1)}$, and $\delta_{(i)} = \begin{cases} 1, & if\ d_i > 0 \\ 0, & othetwise \end{cases}$. Equ. (6.6) 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 equ.(6) 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:

$$\hat{\beta}_{k+1} = \left\{\left(X - \bar{X}^{w}\right)^T W\left(\hat{\beta}_k\right)\left(X - \bar{X}^{w}\right)\right\}^{-1}\left(X - \bar{X}^{w}\right) W\left(\hat{\beta}_k\right) y \quad \ldots (7)$$

Where

$$X = \left(\left(x_{ij}\right)\right),\ \bar{X}^{w} = \left(\left(\sum_{i=1}^{n} w_i\left(\hat{\beta}_k\right) x_{ij}\right)\right) \quad \ldots (8)$$

The limit of the sequence $\hat{\beta}_k$ for k = 0,1, ..., 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.

Buckley-James Approach

The Buckley-ames approach differs from that of Miller (1976) in that the normal equations rather than the sum of squares of residuals are modified and this appears to overcome the inconsistency problems in Miller's approach.

$Z_i = \min(y_i, t_i)$ (i = 1, 2,..., n)

where $y_1,..., y_n$ are the survival times; $t_1, ..., t_n$ are censor values, together with indicator variables

$$\delta_i = \begin{cases} 1 & if\ \ y_i < t_j \quad (uncensored) \\ 10 & if\ \ y_i > t_j \quad (censored) \end{cases}$$

The t_i's are not in general assumed to be random variables. Cox (1972) showed how covariates may be introduced into the distribution-free analysis of such data through the proportional hazards model

$$\lambda(y;\ x) = \lambda_0(y)\ e^{-\beta x}$$

where λ is the hazard function for the survival times with covariates x, λ_0 is an unspecified function and β an unknown parameter vector. Consider the linear regression model

$$y_i = \alpha + \beta x_i + \varepsilon_i \ (i = 1, ..., n) \quad \ldots (9)$$

where the ε_i are independently and identically distributed with unspecified distribution function F, mean zero and finite variance. The model given in equ. (9) will often provide an adequate approximation, particularly after suitable transformation of the survival times and possibly of the covariates. Model given in equ. (9) was studied by Miller (1976), who suggested using as estimates of α and β those values a and b which minimize

$$\int \varepsilon^2 d\hat{F}_{a,b}(\varepsilon)$$

Where

$$\hat{F}_{a,b}(\varepsilon) = 1 - \prod_{i;e_{(i)} \le \delta} \left(\frac{n-i}{n-i+1}\right)^{\delta_i}$$

is the nonparametric product limit estimator of F based on the censored and uncensored residuals $e_i(a, b) = z_i - a - bx_i$

Buckley-James Estimator

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

$$E\left\{\delta_i Y_i + (1 - \delta_i) E\left(T_i | T_i > Y_i\right) | x_i\right\} = \alpha + x_i \beta \quad \ldots (10)$$

Buckley – James substitute an estimate for the conditional expectation $E(T_i | T_i > y_i)$ based on the Kaplan and 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

$$\delta_i = \begin{cases} 1, \hat{y}_i(\hat{\beta}_k) = y_i \\ 0, \hat{y}_i(\hat{\beta}_k) = x_i\hat{\beta}_k + \left\{\sum_{j;\hat{e}_j^0 > \hat{e}_i^0} w_i(\hat{\beta}_k)\hat{e}_i^0\right\} \Big/ \{1 - \hat{F}(\hat{e}_i^0; 0, \hat{\beta}_k)\} \end{cases}$$

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

$$\hat{\beta}_k = \{(X - \bar{X})^T (X - \bar{X})\}^{-1} (X - \bar{X})^T \hat{y}(\hat{\beta}_k) \quad \ldots (11)$$

where the matrix $\bar{X}$ has elements $n^{-1}\Sigma_i x_{ij}$ and $\hat{y}_i(\hat{\beta}_k) = \{\hat{y}_1(\hat{\beta}_k), \ldots, \hat{y}_n(\hat{\beta}_k)\}^T$. 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 equ.(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

$$\left\{\left(X - \overline{X}\right)^T \left(X - \overline{X}\right)\right\}^{-1} \left(X - \overline{X}\right)^T y$$

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

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

Table 1 Heart transplant data

Patient No.	Time of Accept.	Age (X_2)	Surgery (X_1)	Transplant Time (t)	Survival Status	Survival Time (Y)	T_5
1.	0.12	30.84	0		1	50	0.61
2.	0.25	51.84	0		1	6	1.23
3.	0.27	54.30	0	1	1	16	1.11
4.	049	40.26	0	36	1	36	1.66
5.	0.61	20.79	0		1	18	1.09
6.	0.70	54.60	0		1	3	1.24
7.	0.78	50.87	0	51	1	675	1.32
8.	0.84	45.35	0		1	40	0.98
9.	0.86	47.16	0		1	85	0.65
10.	0.86	42.50	0	12	1	58	0.61
11.	0.87	47.98	0	26	1	153	0.36
12.	0.96	53.19	0		1	8	1.68
13.	0.97	54.57	0	17	1	81	1.89
14.	0.97	54.01	0	37	1	1387	0.87
15.	0.99	53.82	1		1	1	1.46
16.	1.07	49.45	0	28	1	308	1.12
17.	1.08	20.33	0		1	36	1.65
18.	1.09	56.85	0	20	1	43	2.05
19.	1.13	59.12	0		1	37	1.46
20.	1.33	55.28	0	18	1	28	2.76
21.	1.34	43.34	0	8	1	1032	1.13
22.	1.46	42.78	0	12	1	51	1.38
23.	1.53	58.36	0	3	1	733	0.96
24.	1.57	51.80	0	83	1	219	1.62
25.	1.57	33.22	0	25	0	1800	1.06
26.	1.58	3.054	0		0	1401	1.44
27.	1.59	8.79	0		1	263	1.46
28.	1.68	54.02	0	71	1	72	0.47
29.	1.79	50.43	0		1	35	0.82
30.	1.88	44.91	0	16	1	852	1.58
31.	1.89	54.89	0		1	16	1.65
32.	1.91	64.41	0	17	1	77	0.69
33.	2.16	48.90	0	51	0	1587	0.91
34.	2.20	40.55	0	23	0	1572	1.22
35.	2.31	46.47	0		1	12	2.09
36.	2.51	48.93	0	46	1	100	0.87
37.	2.57	61.50	0	19	1	66	0.87
38.	2.59	41.47	0	1	4.50	5	1.38
39.	2.63	50.52	0	2	1	53	0.78
40.	2.65	48.48	1	41	0	1408	0.98
41.	2.88	45.30	1	58	0	1322	1.65
42.	2.89	36.44	0		1	3	0.18
43.	3.06	46.39	1		1	2	1.71
44.	3.16	42.58	1		1	40	1.38
45.	3.26	36.18	0	1	1	45	0.00
46.	3.28	48.61	1	2	1	996	0.81
47.	3.34	47.10	0	21	1	72	1.38
48.	3.35	56.04	0		1	9	1.71
49.	3.38	36.65	1	36	0	1142	1.35
50.	3.38	45.89	1	83	1	980	0.75
51.	3.38	48.73	0	32	1	285	1.08
52.	3.56	41.25	0		1	102	0.87
53.	3.75	47.34	0	41	1	188	0.91
54.	3.75	47.79	0		1	3	0.12
55.	3.85	52.45	0	10	1	61	1.51
56.	3.92	38.74	0	67	0	942	0.98
57.	3.95	41.26	0		1	149	1.41
58.	3.98	48.02	1	21	1	343	1.82
59.	3.99	41.38	1	78	0	916	0.19
60.	4.13	49.05	0	3	1	68	0.66
61.	4.18	52.56	0		1	2	0.97
62.	4.19	39.35	0		1	69	1.74
63.	4.20	32.66	0	27	0	842	1.93
64.	4.34	48.82	1	33	1	584	0.12
65.	4.43	51.29	0	12	1	78	1.12
66.	4.47	53.21	0		1	32	1.71
67.	4.48	19.55	0	57	1	285	1.02
68.	4.52	45.24	0	3	1	68	1.68
69.	4.67	47.99	0	10	0	670	1.20
70.	4.71	53.00	0	5	1	30	1.68
71.	4.80	47.41	0	31	0	620	0.97
72.	4.87	26.73	0	4	0	596	1.46

4.95	56.33	0	27	1	90	2.16
4.97	29.17	0	5	1	17	0.61
5.00	52.18	0		1	2	0.98
5.01	52.08	1	46	0	545	1.70
5.02	41.11	0		1	21	1.74
5.09	48.70	0	210	0	515	0.81
5.17	53.78	0	67	1	96	1.08
5.18	52.89	0	6	0	445	1.94
5.28	52.89	0	6	0	445	1.94
4.08	29.20	0		0	428	1.65
5.32	53.31	0	32	1	80	3.05
5.33	42.72	0	37	1	334	0.60
5.35	47.98	0		1	5	1.24
5.42	48.92	0	8	0	397	1.44
5.47	46.25	0	60	1	110	2.25
5.49	54.36	0	31	0	370	0.68
5.51	51.05	0	139	1	207	1.33
5.51	52.03	1	160	1	186	0.82
5.53	47.59	0		1	340	0.16
5.57	44.98	0	310	0	340	0.33
5.78	47.75	0	28	0	265	1.20
5.95	43.84	1	4	1	165	1.24
5.98	40.28	0	2	1	16	0.46
6.01	26.65	0	13	0	180	1.78
6.14	23.62	0	21	0	131	0.77
6.20	28.63	0	96	0	109	2.53
6.23	49.83	0		1	21	1.71
6.35	35.06	1	38	0	39	0.77

Source: Crowley and Hu (1977).

Time of acceptance	Years since Jan. 1, 1967
Matching	1 = number of mismatches, 2 = HLA–A2 mismatch, 3 = mismatch score
Survival status	1 = dead, 0 = censored
Previous surgery	1 -= yes, 0 = no.
Survival time, transplant time	In days from acceptance

In estimating the variability Buckley and James recommend using

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

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

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

where $\bar{X}^u$ is the matrix with elements

$$n_u^{-1}\sum_{ij} X_{ij} \text{ and } \Delta = \text{diag}(\delta_i)$$

Koul, Susarla and Van Ryzin's approach

Let $\{T_i, i = 1,\ldots, n\}$ be n independent random variables satisfying

1. $T_i = \alpha + \beta x_i + \varepsilon_i$, $1 \le i \le n$ where $x_1,\ldots, x_n$ are known input variables, α and β are the parameters
2. $\varepsilon_1,\ldots,\varepsilon_n$ are independent and identically distributed random variables with zero mean.
3. $\delta_i = [T_i < Y_i]$ and $Z_i = \min(T_i, Y_i)$, $1 \le i \le n$.
4. $Y_1,\ldots,Y_n$ are independent and identically distributed random variables which are independent of $\varepsilon_1,\ldots, \varepsilon_n$.

The random variables $Y_1,\ldots,Y_n$ are called the censoring variables. When dealing with survival time data, one can take T_i to be '$\log_{10}$' or 'ln' of the survival time. The problem considered here is that of the estimation of (α, β) based on $(\delta_1,Z_1),\ldots,(\delta_n,Z_n)$.

Koul, Susarla & Van Ryzin's estimator

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

$$E\left[\delta_i Y_i\left\{1 - G\left(Y_i; x_i\right)\right\}^{-1} \mid x_i\right] = \alpha + x_i\beta \qquad \ldots (12)$$

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 given in equ. (12) 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:

$$I - \hat{G}(t) = \prod_{y_i \le t}\left\{\frac{1 + n^+(y_i)}{2 + n^+(y_i)}\right\}^{1-\delta_i} \qquad \ldots (13)$$

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 equ. (13) 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 $\hat{y}_i = \delta_i y_i\{1 - \hat{G}(y_i)\}^{-1} I(y_i < M_n)$, where I is the indicator function. The great advantage of the estimator of Koul *et al.* is that no iteration is required. Specifically,

$$\hat{\beta} = \left\{\left(X - \overline{X}\right)^T (X - X)\right\}^{-1} \left(X - \overline{X}\right)^T \hat{y}, \quad \hat{\alpha} = n^{-1}\sum_i \left(\hat{y}_i - x_i\beta\right)$$

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

Real Data example

For data analysis and interpretation the Stanford heart transplantation data suggested by Turnbull (1974), Crowley and Hu (1977), Miller and Halpern (1982) has been used.

1. Patient Number
2. Y – Survival time in days
3. $Y_1 = Log_{10}(Y)$
4. $X_1 = \begin{cases} 1 \text{ if dead} \\ 0 \text{ if alive} \end{cases}$
5. X_2 = Age at first transplant
6. $X_3 = X_2 \times X_2$
7. T_5 = Mismatch score

and are displayed for 100 patients whose ages between 40 to 50 with complete records, given in table 1. The multiple regression analysis (linear and quadratic) for the data set is shown in table 2 and table 3. Table 4 gives the estimated coefficients and their estimated standard deviations for the variables Y_1, X_1 and T_5 score based on the four approaches. The estimated regression coefficients together with their standard deviations for Log_{10} of time to death verses age and age squared at time of transplant with n=100 Standford heart transplant patients who survived at least 10 days are shown in table 1.

Table 2 Least Square Regression (Linear)

Predictor variables	Coefficient	Std error	Student's t	P- value
Constant	0.71122	0.66630	0.74	0.3420
X_1	– 0.60839	0.10370	– 4.02	0.0000
X_2	– 0.035142	0.01130	2.22	0.0267
X_3	– 0.08715	0.10350	– 0.74	0.2513

Overall F = 10.29
P – value = 0.0000
Adjusted R squared = 0.5685
R squared = 0.5977

From the table 2, p-value is significant since F value is also significant. R squared value is found to be 0.5977. It is observed that only 59.77% of the variation in Y_1 is influenced by the explanatory variables X_1, X_2 and T_5. Considering the individual regression co-efficient, it is found that X_1(Patient survival status) and X_2(Age at transplant) are significant.

Table 3 Least Square Regression (Quadratic)

Predictor variables	Coefficient	Std error	Student's t	p- value
Constant	1.75750	0.30213	4.00	0.0001
X_1	–0.50796	–0.06644	–5.95	0.0000
X_2	–0.29942	–0.19976	2.33	0.0160
X_3	–0.19266	–0.08655	–0.85	0.3402

Overall F = 14.32
P – value = 0.0000
Adjusted R squared = 0.6198
R squared = 0.6698

From the table 3, p-value is significant since F value is also significant. R squared value is found to be 0.6698. It is observed that only 66.98% of the variation in Y_1 is influenced by the explanatory variables X_1, X_3 and T_5. Considering the individual regression co-efficient, it is found that X_1(Patient survival status) and X_3 (Squared age at transplant) are significant.

Table 4 Regression estimates and standard deviations, SD, for log_{10} of time to death versus age at transplant and T_5 mismatch score

Estimator	Intercept		Age		T_5	
	$\hat{\alpha}$	SD ($\hat{\alpha}$)	$\hat{\beta}_1$	SD ($\hat{\beta}_1$)	$\hat{\beta}_2$	SD ($\hat{\beta}_2$)
Cox			0.018	0.014	0.151	0.179
Buckley – James	2.30	0.21	–0.023	0.004	–0.002	0.130
Miller	2.31	0.56	0.000	0.007	0.039	0.141
Koul - Susarla – Van Ryzin	0.57		0.018		0.131	

The regression estimates and standard deviations based on the four approaches, can be found out by using simple program suggested by Kalbfleisch and prentice (1980), MULCOX programs suggested by Lin (1990, 1993), the multiple regression (linear and quadratic) are found out using SPSS version 18.0 and SX software. The scatter plot of Log 10 survival times in days verses age at transplantation (years) for 100 cases for linear and quadratic form are shown in Fig. 1 and Fig. 2.

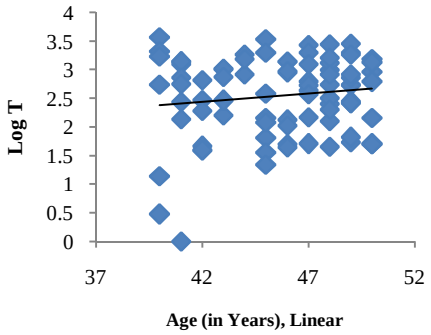

Fig.1

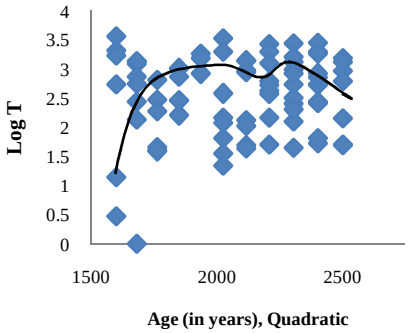

Fig.2

Table 5 Regression estimates and standard deviations for log10 of time to death versus age and age squared at time of transplant with n = 70 Stanford heart transplant patients who survived at least 10 days

Estimator	Intercept		Age		Age^2	
	$\hat{\alpha}$	SD ($\hat{\alpha}$)	$\hat{\beta}_1$	SD ($\hat{\beta}_1$)	$\hat{\beta}_2$	SD ($\hat{\beta}_2$)
Cox			–0.134	0.031	0.0010	0.0003
Buckley – James	1.18	0.74	0.112	0.033	–0.0035	0.0002

CONCLUSION

It is observed from table 5, miller's approach gives less reliable result comparing to the Koul *et. al*. approach. It is found that Cox and Beckley – James have proved to reliable estimates. From the real data example it is observed that for the multiple regression analysis (linear and quadratic) R-squared value is found to be 0.5977 (linear), 0.6698 (quadratic). It is found that in linear case X_1 (patient survival status) and X_2 (age at transplant) are the variables mainly influences the survival time in days. It is also found that in quadratic case X_1 (patient survival status) and X_3 (squared age) are two variables which influences the survival time in days. The goodness of fit test (R-squared value) for both linear and quadratic found to be the same. Among the four approaches Cox and Buckley – James estimator seems to be more consistence than the other two approaches.

Acknowledgement

The authors thank Dr.R.Sathiyamoorthi, Professor & Head of Statistics (Rtd.) for his valuable guidance in completing this work.

Reference

1. Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics*, 30, 89-99.
2. Buckley, J. and James, I. (1979). Linear Regression with Censored Data. *Biometrika*, 66(3), 429-36.
3. Cox, D.R. (1972). Regression model and life tables (with discussion). *Journal of the Royal Statistical Society*, B 34, 187-220.
4. Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *J. Am. Statist. Assoc*., 72, 27-36.
5. Kalbfleisch, J. and Prentice, R. (1980). The Statistical Analysis of Failure time Data, Wiley, New York.
6. Kay, R and Kinnersley, N. (2002). On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: A case study in influenza. *Drug Inform. J.*, 36, 571-579. Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right censored data. *Ann. Statist*., 9, 1276-1288.
7. Lin, D.Y. (1990). MULCOX: A computer program for the Cox regression analysis of multiple failure time variables. Computer Methods and Programs in Biomedicine, 32, 125-135.
8. Lin, D.Y. (1993). MULCOX2: a general computer program for the Cox regression analysis of multiple failure time variables. Computer Methods and Programs in Biomedicine, 40, 279-293.
9. Link, C. L. (1979). Confidence intervals for the survival function using Cox's proportional hazard model with covariates. Tech. Rep. 45, Division of Biostatistics, Stanford University
10. Miller, R. and Halpern, J. (1982). Regression with Censored Data. *Biometrika*, 69(3), 521-31.
11. Miller, R. and Halpern, J. (1982). Regression with Censored Data. *Biometrika*, 69(3), 521-31.
12. Miller, R. G. (1976). Least squares regression with censored data. *Biometrika*, 63(3), 449-464.
13. Prentice, R. L. and Gloeceler, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34, 57-67.
14. Tsiatis, A. (1981). A large sample study of Cox's regression model. *Ann. Stat.*, 9, 93-108.
15. Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Am. Statist. Assoc*., 69, 169-73.

****