Tuberculosis, 1995(2) 45 - 48

CO-VARIATE ANALYSIS OF TUBERCULOSIS DATA USING COX'S REGRESSION MODEL

P. Venkatesan*, K. Viswanathan@ and R. Prabhakar*

*Tuberculosis Research Centre (I.C.M.R.), Madras, India
@Dept. of Statistics, University of Madras, Madras, India

Abstract

A regression model which allows for analysis of censored survival data adjusting for continuous as well as discrete covariates and varying with time has been proposed by Cox. The hazard rate could be modelled as a function of both time and covariates and the hazard rate could be represented as the product of two terms, the first representing an unadjusted force of mortality which can be estimated non~parametrically and the second adjusting for the linear combination of a particular covariate profile. In this paper an attempt is made to demonstrate the value of this model with pulmonary tuberculosis data in quantifying the effects of disease, demographic and treatment variables.

Introduction

Many applications in medical research concern the relationship, between certain covariates and time to occurrence of certain event[1-5]. In clinical trials of tuberculosis several treatments are compared with rcspcct to the time to response, toxicity or relapse[6-10]. The covariates could include indicator, components (measures of response) for treatment as well as other prognostic and demographic characteristics.

The distribution of response time or event-free time can be represented in the usual way by means of density or distribution function. But the presence of censoring precludes this approach. This leads to the use of more specialised methods such as hazard function otherwise called as instantaneous failure rate, force of mortality or age-spccific failure rate. The notion of failure rate is basic and conceptually simple and provides a starting point for modelling for association between response, toxicity, relapse or death and various covariates. Many books on survival analysis deal this aspect exhaustivcly[11-14].

2. Cox's semi-parametric, regression model

The multiplicative hazard or proportional hazards mode1 is defined by

$$h(t;z) = ho(t)\ g(z\beta)$$

where h(t;z) is the hazard rate at time t for an individual with covariate vector z. All multiplicate hazard models share a common feature, that the hazard ratio of two individuals is proportional and is not dcpcndcnt on time. The generalisation of this characteristic called proportional hazards was introduced by Cox[15] and is given by

$$h(t;z) = ho(t)\ \exp\ (\gamma\beta)$$

without further spccification of the base line hazard. It is often reflected to as semi-paramctric model as it assumes no distributional shape for ho(t). The flcxibility of this model makes classical likelihood theory inapplicablc. Inference on this model is based on the partial likelihood principle[15,16]. The baseline hazard is estimated non-parametrically.

Correspondance : Dr. P. VENKATESAN, Department of Statistics, Tuberculosis Research Centre (I.C.M.R.), Spurtank Road, Chetput, Madras - 600 031, INDIA

The main reason for the popularity of-Cox's model is that (a) it allows for non-informative censoring [18], (b) a parametric form for the nominal failure rate ho(t) need not be specified and (c) various covariates represented by the vector z may be discrete, continuous or even time dependent.

A rigorous development of the asymptotic properties of the Cox's model has been given by Tsiatis[19] who concentrated on situations where the variables in z, are realization of a random vector. He noted that the components of z cannot be monotone with respect to failure times if the partial likelihood is to have a maximum.

The behaviour of covariate adjustment in small samples with censoring was studied by Johnson *et al.* [20] The grouped version of the Cox's model was first introduced by Kalbfleisch and Prentice[21] and maximum likelihood estimation was discussed by Prentice and Gloeckler. [22]

3. Application to pulmonary. tuberculosis data

The application to pulmonary tuberculosis data considered here is based on the works of Venkatesan.[23] The data consists of 261 patients taken from a study conducted at the Tuberculosis Research Clinic, Madras.[9] The patients belong to treatment groups of the controlled clinical study of short-course regimens described as below:

R-Series : Daily chemotherapy with rifampicin 12 mg/kg body-weight plus streptomycin sulphate 0.75, g plus isoniazid 400 mg plus pyrazinamide 40 mg/kg followed by twice weekly chemotherapy for 5 months with streptomycin 0.75 mg plus isoniazid 1.5 mg/kg plus pyrazinamide 0 mg/kg, the total duration belong 7 months.

Z-series : The same as the R series but without rifampicin, the total duration belong 7 months.

Of the 261 patients (132 R-series, 127 Z-series) 67% were males, 24% of the patients were less than 2.5 years of age, 63% were between 25 and 44 and 13% were more than 45 years. The patients in the two treatment series were almost similar with respect to age, sex, and weight. The response variable considered was time taken for the conversion of the positive sputum culture to negative from start of chemotherapy. Of over 30 disease and demographic characteristics available a set of 5 (age, sex, disease level, cavity level and smear grade on first collection) at the time of admission involving 14 components was chosen on the basis of a preliminary investigation, to illustrate the application.

The explanatory variables considered are shown in Table I. Except for age, all the variables are categorical and we have used dummy covariates to represent them. The smear grade refers to the pathological characteristic of tuberculosis, while the disease level and cavity level refer to radiological characteristic of tuberculosis and 9% of the observation are censored by end of the study. The results are presented in Table II.

TABLE I

Initial variables - pulmonary tuberculosis data

Variable	Description
Y	**Time to culture negativity (1-uncensored. O-censored)**
X1	**Age at diagnosis (years)**
X2	**Sex (1-male, 0-female)**
X3	**Regimen (1-Z series, 0-R series)**
	Extent of disease
X4	**EOD 1 : (1-limited, 0-other)**
X5	**EOD 2 : (1-moderate, 0-other)**
X6	**EOD 3 : (1-extensive/gross, 0-other) (Base line - slight)**
	Extent of cavity
X7	**EOC 1 : (1-slight, 0-other)**
X8	EOC 2 : (1-moderate, 0-other)
X9	EOC 3 : (1-extensive, 0-other) (Base line - nil)
X10	CSM 1 : (1scanty, 0-other)
X11	CSM 2 : (1-moderate, 0-other)
X12	CSM 3 : (1-heavy, 0-other) (Base line - negative)

TABLE 2

Radiological and Bacteriological findings on admission

Characteristics	Percentage of patients	
	R-series	Z-series
Extent of Disease		
Slight	5	8
Limited	24	19
Moderate	39	39
Extensive/ Gross	32	35
Extent of Cavity		
Nil	6	6
Slight	37	33
Moderate	35	30
Extensive	22	31
Smear Result (Ist Collection)		
Negative	11	8
Scanty (1+)	34	33
Heavy (3+)	38	46
	17	14
Total patients	132	129

The analysis which follows (Table III) was carried out using BMDP Package in VAX11/750 computer. The data is complex and the analysis meant to illustrate various points rather than to give exhaustive discussion of its features.

TABLE 3

M.l.e. lit of Cox's model to data on 261 pulmonary tuberculosis patients

Co-variate	B	Se(B)	Sig
REG	- 0.344	0.117	0.01
EOD1	0.101	0.083	
EOD2	0.210	0.210	NS
EOD3	0.314	0.255	
EOC1	0.100	0.073	
EOC2	0.348	0.238	NS
EOC3	0.369	0.258	
CSM1	0.162	0.140	
CSM2	0.231	0.154	NS
CSM3	0.259	0.170	

From the examination of the fitted model, it is clear that only the regimen covariate has significant effect on the response time. The estimated regression co-efficient for the z series is β = -0.344 with an estimated standard error of β from the observed information matrix is 0.117 given a t-value of 2.94 which is significant at 0.01 level. This suggests that the response time hazard rate is reduced by an estimated multiplicative factor exp (β) = 0.71 for deletion of rifampicin drug to the treatment. The estimates for the other covariates were not significant. But we observe from the table that the β 's increases as we move from the lower levels to higher levels for all the three covariates. For example, the hazard of response is prolonged by a factor exp(0.101) = 1.11 for a patient with limited disease. The factors for moderate and extensive diseases are 1.23 and 1.37 respectively. For a patient with two or more. covariates the hazard will be prolonged by exp(sum of the β 's). For example, the hazard of covariates for a patient having moderate disease, extensive cavity and smear grade heavy will be prolonged by a factor 2.31 for favourablc response. If this patient is treated with z series, the hazard will be reduced by a factor 0.71 in comparison to a patient in R - series.

4. Discussion

The proportional hazard model paradigm is flexible and convenient for covariate analysis of tuberculosis data and in many situations allow succinct representation of the approximate effect of covariates on the distribution of response time (for extensive discussion see Venkatesan[23]). The Cox's analysis is flexible in allowing hazard functions of different shapes. An important point particularly with strong regression effects concerns the appropriateness of proportional hazard assumption. When the population is heterogeneous and there is considerable variation in survival times, the constant proportional hazards at all times often seems unreasonable. In these situations accelerated failure time model or modified proportional hazard model often found useful.[23] The other possible departure one can make is the proportional odds[24] model which assumes that the hazards ratio converges monotonically.

5. Conclusion

Regression methodology provides useful tools for investigating medical data bases. It is our experience that with softwares such as a BMDP, GLIM and SAS, it is feasible to explore tuberculosis data and to fit and assess regression models with moderate amount of computer time. In cases, where it does not given satisfactory description, it provides a base line against which to look for other interesting features. In studies involving human subjects no

theoretical form for survival distributions can always be feasible and in such situations the Cox's regression model appears to be superior to other regression models.

References

1. Aitkin, M., Laird, N. and Francis, B. A reanalysis of the Standford heart transplant data (with discussion). J. Amer. Statist. Assoc. 78: 264, 1983.

2. Ciampi, A., Bush, R., Gospodarowicz, M. and Till, J.E. A classification or survival experience of non-Hodgkins Lymphoma patients. Cancer, 47: 621, 1981.

3. Gore, S.M., Pocock, S.J. and Kerr, G.R. Regression models and non-proportional hazards in the analysis of breast cancer survival. Appl. Statist. 33: 176, 1984.

4. Harrel, F.E., Lee, KL., Matcher, D.B. and Reichert, T.A. Regression models for prognostic prediction : advantages, problems and suggested solution. Cancer Treat Rep. 69: 1071, 1985.

5. Lawless, J.F. and Singhal, M. In : Biostatistics, I.B. MacNeil and G.S. Umprey, Eds. (D. Reidel Pub. Co.) 1987.

6. Tuberculosis Chemotherapy Centre, Madras. A concurrent comparison of home and sanatorium treatment of pulmonary tuberculosis in South India. Bull Wld Hlth Org. 23: 51, 1959.

7. Tuberculosis Chemotherapy Centre, Madras. A concurrent comparison of isoniazid plus PAS with 3 regimens of isoniazid alone in the domiciliary treatment of pulmonary tuberculosis in South India. Bull Wld. Hlth org. 23: 535. 1960.

8. Tuberculosis Chemotherapy Centre, Madras. A concurrent comparison of intermittent (twice week) isoniazid plus strepmycin and daily isoniazid plus PAS in the domiciliary treatment of pulmonary tuberculosis. Bull Wld. Hlth Org. 31: 247, 1964.

3. Tuberculosis Research Centre, Madras. Study of chemotherapy regimens of 5 and 7 month's duration in the treatment of sputum-positive patients with pulmonary tuberculosis in South India. Am. Rev. Respir. Dis 134: 27, 1986.

11. Cox, David, R. and Oakes, D. Analysis of Survival data. (Chapman and Hall, New York), 1984.

12. Elandt-Johnson, R.C. and Johnson, N.D. Survival models and data analysis. (Wiley. New York) 1980.

13. Kalbfleisch, J.D. and Prentice, R.L. The Statistical analysis of failure time data. (Wiley, New York) 1980.

14. Lawless, J.F. Statistical models and methods for life time data. (Wiley. New York) 1982.

15. Cox, D.R. Regression models and life tables (with discussion). J. R. Statist. Soc. B24: 406, 1972.

16. Cox, D.R. Partial likelihood. Biometrika. 62: 269, 1975.

17. Oakes, D. Survival times : Aspects of partial likelihood (with discussion). Int. Statist. Rev. 49: 199, 1981.

18. Lagakos, S.W. General right censoring and its impact on the analysis of survival data. Biometrics, 35: 139, 1979.

19. Tsiatis, A.A. A large sample study of Cox's regression model Ann. Statist 9: 93, 1981.

20. Johnson, M.E, Tollcy, H.D., Bryson, M.C. and Goldman, A.S. Covariate analysis of survival data small sample study of Cox's model. Biometrics, 38: 685, 1982.

21. Kalbfleisch, J.D. and Prentice, R.L Marginal likelihood based on Cox's regression and lift model. Biometrika, 60: 267, 1973.

22. Prentice, R.L. and Gloeckler. L.A. Regression analysis of grouped survival data with application to breast cancer data. Biometrics. 34: 57, 1978.

23. Venkatesan. P. A study on survival models and their biomedical applications. Ph.D. thesis, University of Madras, Madras, 1990.

24. Bennet, S. Analysis of survival data by the proportional hazards model. Statist. Med. 2: 273, 1953.