VOL. 10, NO 21, NOVEMBER, 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

SUBDISTRIBUTION TO COX'S MODEL FOR PARTLY INTERVAL-CENSORED DATA WITH APPLICATION TO AIDS STUDIES

F. A. M. Elfaki
Department of Science in Engineering, Faculty of Engineering, International Islamic University Malaysia, Kuala Lumpur, Malaysia
E-Mail: faizelfaki@yahoo.com

ABSTRACT

In this paper, we consider incomplete survival data that is, partly-interval failure time data where observed data include both exact and interval-censored observation on the survival time of interest. We presented a modification of Fine and Gray (1999) which proposed a class of estimation procedures for semi-parametric proportional hazards regression model for the subdistribution of a competing risks model using the partial likelihood principle and weighting techniques. The method is evaluated using simulation studies and illustrated by AIDS data set.

Keywords: Cox's model, competing risks, HIV, AIDS, subdistribution function.

1. INTRODUCTION

Suppose time to event random variable, or failure time $T_1, T_2, ..., T_n$ are independent and identically distributed as F_0. If all the random variables are observable, then it is well known that the semiparametric maximum likelihood estimator of F_0 is the empirical distribution function and it is asymptotically efficient. However many reliability and medical studies, observations are subject to censoring. The goal of this paper is to discuss a semi-parametric Cox's proportional hazards regression model with the subdistribution of F_0 based on incomplete partly interval-censored data in which some of the failure times are observed, but some of the failure times are subject to interval censoring [9, 12].

There is many cases of partly censored data, here we consider the case that is; for the some subjects, the exact failure times $T_1, T_2, ..., T_n$ are observed. But for the remaining subjects, only the information pertaining to their current status is available. That is for the ith subject in this group, we only know whether or not failure has occurred at the examination time U_i, so the observed data is

$$(\delta_i, U_i) \quad i = n_1 + 1, ..., n$$

where $\delta_i = 1$ if the unknown failure time $T_i \le U_i$ and $\delta_i = 0$ otherwise. Note that this censored model is different from doubly-censored data studied by [3, 4, 8].

General partly interval censored data often arise in follow-up studies. An example of such data is provided by the Framingham Heart Disease Study; see 10 for a description. In this study, time of the first occurrence of subcategory angina pectoris in coronary AIDS/ HIV disease patients are of interest. For some patients, time of the first occurrence of subcategory angina pectoris is recorded exactly. But for others, time is recorded only between two clinical examinations.

In the competing risks model, a unit is exposed to several risks simultaneously, but it is assumed that the eventual failure of the unit is due to only one of these risks, which is called "cause of failure" [1]. The cause specific hazard function is known as subdistribution function, also historically was known as the cumulative incidence function, the marginal probability function, the crude incidence or the absolute cause-specific risk [2]. In this paper, we propose the semi-parametric proportional hazard model of the subdistribution function for partly interval-censored of a competing risks survival data based on EM algorithm to estimate the parameters.

2. COMPETING RISKS MODEL FORMULATION

[6] developed a class of estimation procedures for semi-parametric proportional hazards regression model for the subdistribution of a competing risks model using the partial likelihood principle and weighting techniques. The main interest is the modeling of the cumulative incidence function for failure from say cause 1 conditional on the covariates, i.e $F_1(t;Z) = \Pr(T \le t, \varepsilon = 1/Z)$, and the hazard of the subdistribution as originally described by [7]. Gray constructed K-sample tests for differences in the cumulative incidence function based on integrated difference of nonparametric estimates of the within-group subdistribution hazard functions. The subdistribution hazard as defined by Gray is,

$$\lambda_i(t;Z) = \lim_{\Delta t \to 0} \Pr(t \le T < t + \Delta t, \varepsilon = 1/T \ge t, Z = z) / \Delta t,$$

$$= \frac{1}{1-F_1(t;Z)}.\frac{d}{dt}\left(F_i(t;Z)\right) \quad (1)$$

$$= \frac{-d}{dt}\left[\log\{1-F_1(t;Z)\}\right]$$

The cumulative incidence function and subdistribution hazard functions are estimable from the competing risks data [11]. We use Cox proportional hazards models to specify each $\lambda(t;Z)$ and assume that censoring is conditionally independent of the latent failure times for given Z. Then, under Cox model;

$$\lambda(t;Z) = \lambda_0(t)\exp(Z^T(t)\beta) \quad (2)$$

where $\lambda_0(t)$ is a completely unspecified, nonnegative function in t, β is regression coefficients and $Z(t)$ is the original time-dependent covariates (time-varying covariates). For simplicity, we restrict our attention to a time-independent covariates. Thus the regression coefficients and baseline hazard form the Cox model for F have straightforward interpretation that does not depend on the probabilistic structure of the subdistribution hazard and given as;

$$F(t;Z) = 1-\exp\left[-\int_0^t \lambda_0(s)\exp\{Z^T(s)\beta\}ds\right] \quad (3)$$

3. SEMI-PARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION

3.1. Complete data

The partial likelihood for the improper distribution, $F(t:Z)$ as proposed by [6] is;

$$L(\beta) = \prod_{i=1}^{n}\left[\frac{\lambda_0(t_i)\exp\left\{Z_i^T(t_i)\beta\right\}\Delta t_i}{\sum_{j\in R_{(t_i)}}\exp\left\{Z_i^T(t_i)\beta\right\}}\right]^{I(\varepsilon_i=1)} \quad (4)$$

where $R_{(t_i)}$ is the risk set at time of failure for the ith individual.

The log partial likelihood will be obtained from equation (4) and then an iterative process such as the EM algorithm or Newton-Raphson is adopted to solve this system of equations for β.

4. A WEIGHTED SCORE FUNCTION METHOD

To modify the second model of [6], let $\{T_i, C_i, Z_i, i = 1,...,n\}$ to be n independent copies of $\{T, C, Z\}$. However, one can only observe $X_i = \min(T_i, C_i)$ and $\Delta_i = I(C_i \le T_i)$ for $i = 1,...,n$. In the case when the survival distribution $G(.)$ of the censoring variable C does not depend on Z, the weight at time t proposed by [6] is, $w_i(t) = r_i(t)\hat{G}(t)/\hat{G}(X_i \wedge t)$ can make a simple modification of the weight at time t (5) as follows;

$$w_i(t) = \frac{r_i(t)G(t)}{G_{Z_i}(t)G_{Z_j}(t)} \quad (5)$$

where $G_Z(.)$ is the Kaplan-Meier estimator for the survival function and $r_i(t) = I(C_i \ge T_i \wedge t)$ is the vital status on individual i at time t. Censored individuals are observed until time C_i; thereafter, vital status is uncertain. If $r_i(t) = 0$, then $Y_i(t)$ and $N_i(t)$ are not observable. If $r_i(t) = 1$, then $Y_i(t)$ and $N_i(t)$ are observed data up to time t.

5. SIMULATION STUDIES

A simulation study was conducted to evaluate the finite sample performance of our proposed methods. Our simulation set-up is similar to that in [9]. We generated data from exponential distribution with $h_0(t) = 1$ under our proposed model $h(t/z) = h_0(t)\exp(z\theta)$. Examination times were generated to make the proportions of left, interval, and right-censored observations about equal. The sample size n is the sum of the number of exact data n_1 and the number of interval-censored data n_2. Following [9] we consider the range as; (25, 25) and (40, 10) for a sample of size 50, and (50, 100), (50, 150) and (50, 200) for the sample size 150, 200 and 250 respectively. We refer to (t, z) and (t_1, z_1) as the original data and exact data respectively. Table 1 show our results compared with one obtained by [9]. For each sample, we obtained the bias and the mean standard error. The estimation based on the exact data and original complete data. The results obtained by our two proposed models that is, the censoring complete model (CC) and a weighting technique model (W) look similar compare to the one obtained by [9].

Table-1. Comparison results obtained by our proposed model with [9] from simulation data from 1500 replication.

Proposed model (Weighting Technique Model (W)**)**								
(n_1, n_2)	**Biases**			**Standard errors**				
	$\hat{\beta}_E$	$\hat{\beta}_O$	$\hat{\beta}_{OC}$	$\hat{\sigma}_E$	$\hat{\sigma}_M$	$\hat{\sigma}_p$	s	$\hat{\sigma}_{OC}$
(25, 25)	0.242	0.112	0.073	1.132	0.495	0.509	0.602	0.417
(40, 10)	0.072	0.074	0.052	0.472	0.434	0.443	0.475	0.406
(50, 100)	0.088	0.042	0.022	0.402	0.289	0.287	0.311	0.207
(50, 150)	0.082	0.034	0.023	0.408	0.257	0.259	0.258	0.189
(50, 200)	0.085	0.017	0.012	0.413	0.226	0.237	0.238	0.165
Kim (2003)								
(25, 25)	0.258	0.103	0.062	1.140	0.522	0.531	0.617	0.424
(40, 10)	0.081	0.061	0.058	0.507	0.450	0.455	0.485	0.424
(50, 100)	0.079	0.030	0.019	0.426	0.297	0.298	0.301	0.231
(50, 150)	0.075	0.023	0.011	0.424	0.264	0.267	0.266	0.198
(50, 200)	0.074	0.014	0.008	0.428	0.240	0.243	0.241	0.177

Where $\hat{\beta}_E$ estimated bias from exact data; $\hat{\beta}_O$ estimated bias from observed data; $\hat{\beta}_{OC}$ estimated bias from original complete data, $\hat{\sigma}_E$ mean standard error estimate from the exact data; $\hat{\sigma}_M$ mean standard error estimate from the observed data; $\hat{\sigma}_p$ mean standard error estimate from the observed data by profile information; s sample standard deviation of regression parameter estimates from the observed data; $\hat{\sigma}_{OC}$, mean standard error estimate from the original complete data.

Table-2. Estimate obtained under the modification of [6] using EM algorithm for Sudan HIV/ADIS data.

First causes				
Eq.	β_1	β_2	var($\hat{\beta}$)	E(var)
W	0.632 (0.184)	0.601(.057)	0.013	0.013
CC	0.641 (0.184)	0.610(.057)	0.013	0.013
Other causes				
W	0.720 (0.145)	0.845(1.76)	0.0739	0.0789
CC	0.732 (0.145)	0.854(1.76)	0.0789	0.0789

6. EXAMPLE: APPLICATION TO HIV/AIDS IN SUDAN

The proposed method illustrated HIV/AIDS of hemophiliacs who were treated in two hospitals in Sudan. There were 550 patients of the study at risk for HIV infection through the contaminated blood factor. At the end of the study, there were 550 patients found to be HIV infected, but the infection times were interval-censored. Among them 120 progressed to AIDS (or related symptoms). The patients were classified into either the heavily treated group or lightly treated group according to the amount of blood received (when treated for hemophilia). The goal here to investigate the possible association between the treatment and the AIDS incubation time. We code the covariate $z_i = 0$ or $z_i = 1$ if the ith patient was lightly or heavily treated. To see the effect of covariates on development of complications, we fitted our proposed model that is competing risks model based on EM algorithm. Applying the procedures described in section 2, 3 and 4, we obtained the result as shown in Table-2. We conclude that the covariates do not have a significant different. However, it is confirmed that the heavily treated group had a significantly higher risk of

the onset of AIDS after HIV infection. In additional to that; our both model show similar results in the two causes of failure, in fact the first cause of failure is better compared to the second one with respect to the smaller variance and standard error of the estimation.

7. CONCLUSIONS

We have proposed a simple modification of estimating functions for partly-interval censored data using the semi-parametric Cox's proportional hazards regression models of the subdistribution of a two competing risks models namely, the censoring complete model (CC) and a weighting technique model (W). Simulations studies indicate that under the assumed modification models of [6], the weighted estimating equation with censored data can be as efficient as the censoring complete score function. However, both proposed models give similar results from the two simulation studies ([5]). EM algorithm was used to estimate the parameters of the model. The simulation studies strongly support the generalized missing information principle in a semi-parametric context and use of the generalized profile information for non-identically distributed samples. From the real data set, we find that the covariates do not have a significant difference. Fixing the age at diagnosis, a male has a lower hazard rate than female. Fixing gender, very young patients have a lower hazard rate than relatively young patients. Even with many exact observations (550), the additional interval-censored observations (126) help to give a more accurate estimate of the regression parameter. However, it is confirmed that the heavily treated group had a significantly higher risk of the onset of AIDS after HIV infection.

REFERENCES

[1] Aly, A. A. E, Kochar, S. C., and Mckeague, I. W. 1994. Some Tests For Comparing Cumulative Incidence Functions and Cause-Specific Hazard Rates. American. S. A. J. 89: 994-999.

[2] Benichou, J., and Gail, M. H. 1992. Estimates of Absolute Cause Specific Risk in Cohort Studies. Biometrics. 46: 813-826.

[3] Chang, M. N. 1990. Weak Convergence of a Self-Consistent Estimator of the Survival Function with Doubly Censored Data. Ann. Statist. 18: 391-404.

[4] Chang, M. N. and Yang, G. L. 1987. Strong Consistency of a Nonparametric Estimator of the Survival Function with Doubly Censored Data. Ann. Statist. 15: 1536-1547.

[5] Elfaki, F. A. M. 2004. Parametric and semi parametric competing risks models for statistical process control with reliability analysis. Ph.D., Thesis. University Putra Malaysia.

[6] Fine. J. P. and Gray. R. J. 1999. A Proportional Hazards Model for the Subdistribution of a Competing Risk. American. S. A. J. 94: 496-509.

[7] Gray, R. J. 1988. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. The Annals of Statistics. 80: 557-572.

[8] Gu, M. G. and Zhang, C. H. 1993. Asymptotic Properties of Self-Consistent Estimation Based on Doubly Censored Data. Ann. Statist. 21: 611-624.

[9] Kim. J. S. Maximim Likelihood Estimation for the Proportional Hazards Model with Perty Interval-Censored Data. J R. Statist. Soc. 2003, Series B 65: 489-502.

[10] Odell, P. M., Anderson, K. M. and D'Agostino, R. B. 1992. Maximum Likelihood Estimation for Interval-Censored Data using a Weibull-based Accelerated Failure Time Model. Biometrics. 48, 951-959.

[11] Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T., and Berslow, N. E., 1978. The Analysis of Failure Times in the Presence of Competing Risks. Biometrics, 34: 541-554.

[12] Zhao. X., Zhao. Q. Sun. J. and Kim S. J. 2008. Generralized Log-Rank Test for Partly Interval-Censored Failure Time Data. Biometrical Journal. 3:375-385.