A Correlation Technique to Reduce the Number of Predictors to Estimate the Survival Time of HIV/ AIDS Patients on ART

Vajala Ravi[1], Gurprit Grover[2], Rabindra Nath Das[3], M.K. Varshney[4] and Anurag Sharma[2,*]

[1]Department of Statistics, Lady Shri College, University of Delhi, Delhi-110024, India

[2]Department of Statistics, University of Delhi, Delhi-110007, India

[3]Department of Statistics, The University of Burdwan, Burdwan, West Bengal-713104, India

[4]Department of Statistics, Hindu College, University of Delhi, Delhi-110007, India

Abstract: Till now, many research papers have been published which aims to estimate the survivle time of the HIV/AIDS patients taking into consideration all the predictors viz, Age, Sex, CD4, MOT, Smoking, Weight, HB, Coinfection, Time, BMI, Location Status, Marital Status, Drug etc, although all the predictors need not to be included in the model. Since some of the predictors may be correlated/ associated and may have some influence on the outcome variable, therefore, instead of taking both the significantly correlated/ associated predictors, we may take only one of the two. In this way, we may be able to reduce the number of predictors without affecting the estimated survival time. In this paper we have tried to reduce the number of predictors by determining the highly positively correlated predictors and then evaluating the effect of correlation/ association on the survival time of HIV/AIDS patients. These predictors that we have considered in the starting are Age, Sex, State, Smoking, Alcohol, Drugs, Opportunistic Infections (OI), Living Status (LS), Occupation (OC), Marital Status (MS) and Spouse for the data collected from 2004 to 2014 of AIDS patients in an ART center of Delhi, India. We have performed one – way ANOVA to test the association between a quantitative and a categorical variable and Chi-square test to test between two categorical variables. To select one of the two highly correlated/ associated predictors, a suitable model is fitted keeping one predictor independent at a time and other dependent and the model having the smaller AIC is considered and the independent variable in the model is included in the modified model. The fitted models are logistic, linear and multinomial logistic depending on the type of the independent variable to be fitted. Then the true model (having all the predictors) and the modified model (with reduced number of predictors) are compared on the basis of their AICs and the model having minimum AIC is chosen. In this way we could reduce the number of predictors by almost 50% without affecting the estimated survival time with a reduced standard error.

Keywords: AIDS, AFT, Correlation, Chi- Square test, One- Way ANOVA.

1. INTRODUCTION

Human Immunodeficiency Virus (HIV) is a virus that attacks and destroys the infection-fighting CD4 cells of the body's immune system. Due to continuous loss of CD4 cells, it becomes very difficult for the immune system of the body to fight infections. As a result, the immune system of the patient damages progressively. Due to the progressively damaged immune system, the infected person becomes immunosuppressed and is, therefore, vulnerable to other opportunistic infections, especially tuberculosis [1]. There is another advanced and symptomatic form of HIV, known as, Acquired Immunodeficiency Syndrome (AIDS). In the modern world, this epidemic is considered as one of the most destructive health crises of modern times. This epidemic is destroying families and communities around the world, causing huge socio-economic burdens. It is assumed that worldwide 36 million persons are infected from HIV and this disease has caused 1.2 million deaths globally [2]. In India only, 2.5 million people were estimated to have been suffering from HIV till 2014 [3]. This virus can be transmitted through many ways which includes transmission of blood, semen, genital fluids, or breast milk of an infected person. Among all the modes of transmission, most common ways through which HIV is spread are unprotected sex or sharing drug injection equipment with an infected person. Many tremendous researches have been conducted in the field of HIV/AIDS, but there is currently no cure for this infection. There are, however, steps that can be taken to delay the onset of full blown AIDS and to reduce its progression. The most promising advance has been the advent of potent combination of therapy, the Anti-retroviral therapy (ART) in 1996. The ART can prolong the life of the infected patient by slowing down the wasting period as it boosts the CD4 count in the immune system.

In HIV dynamics, every patient is supposed to visit Anti- retroviral therapy (ART) center after four weeks, but actual time visit may differ from patient to patient and also time between visits may vary.

Several studies have been proposed for the estimation of HIV populations and underlying covariate effect on the hazard of death among HIV patients by

*Address correspondence to this author at the Department of Statistics, University of Delhi, Delhi-110007, India;
E-mail: anuragsharma532@gmail.com

E-ISSN: 1929-6029/18

using Cox proportional model (Ghate *et al*., 2011 [4]; Rai *et al*., 2013 [5]; Kee *et al*., 2009 [6]; Jerene *et al*., 2006) [7].

Hernan *et al*. (2005) developed a structural AFT model for estimating the effect of HAART on AIDS free survival in two prospective cohort studies of HIV infected individuals [8].

Xue *et al*. (2006) discussed a semi parametric accelerated failure time regression model for interval censored HIV/AIDS data [9].

Grover and Banerjee (2011) estimated survival of HIV-1 infected children for doubly and interval censored data [10].

Nawumbeni *et al*., (2014) compares the performance of Cox PH model and the Accelerated Failure Time (AFT) model using HIV/TB Co-infection Survival data [11].

Tarekegn (2011) conducted a retrospective study in which a total of 632 patients(316 in ART and pre-ART cohort) were followed for a median of 32.9 months in pre-HAART and 35.4 months in HAART. The study aimed to identify factors that increase the risk of TB in People Living with HIV/AIDS (PLWHA) [12].

Musenge *et al*., (2013) modeled the contribution of spatial analysis to understanding HIV/TB mortality in children using the structural equation modeling approach. They used multiple logit regression model with and without spatial household random effects [13].

Grover and Swain (2016) identified the independent predictors affecting the survival of HIV/AIDS infected patients on Antiretroviral Therapy (ART), an interval censored event time outcome [14].

Grover.G, and Varshney M.K (2012) estimated the probability of death of AIDS patients in the presence of competing risks. They identified and studied the effect of various independent risk that an AIDS patient is exposed to in day to day life [15].

Grover G and Vajala Ravi (2014) estimated the expected survival time of AIDS patients undergoing Antiretroviral therapy using generalized Poisson regression model. They used the parametric approach with and without covariates to analyze the survival data of HIV/AIDS patients. They analyzed the inclusion of covariates using Censored Generalized Poisson Regression Model (CGPR)[16].

In all these studies, all the possible predictors have been taken into the model which may result in the increased Standard error of the model. In this paper, we have proposed a novel correlation screening procedure for reducing the number of predictors to be included in the model without affecting the estimated survival time and also resulting in lower Standard error.

This research is aimed to:

- Model HIV/AIDS population on ART by using AFTM;
- Determine the significantly correlated predictors in the model;
- Isolate the significant predictors between two correlated predictors;
- Determine the effect of correlated predictors on the AFTM for estimating survival time of HIV/AIDS patients under ART;
- Compare the two models (One with all the predictors and other with reduced number of predictors) by comparing their AIC
- Comparing the proposed method with the existing variable reduction methods [17, 18].

Software R is used to fit the models and to compute the corrections/associations among predictors.

2. MATERIALS AND METHODS

Estimation of AFT Models

Suppose that we wish to estimate the survival time of HIV/AIDS patients under ART using AFT model with all possible predictors. Let T_i be a random variable representing survival time of i^{th} patient, then

$$\log(T_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots\dots\dots \beta_p x_{pi} + \sigma\varepsilon_i \tag{1}$$

where,

β_0 is the intercept

$\beta_i' s$ are the coefficients of "*p*" explanatory variables for i^{th} patient.

σ is the scale parameter

ε_i is a random variable used to model the deviation of values of $\log_e(T_i)$ from the linear part of model.

The parameters of AFT model are estimated by the maximum likelihood estimation method and by using Newton- Raphson procedure.

The correlation between a pair of variable is a measure of the linear association between two variables. If the correlation between two pairs is significant, then instead of taking both the variables, taking only one will be sufficient. For example, in the model $(1.1), X_1 \text{ and } X_2$ are significantly correlated/ associated, then keeping only of them in the model will fulfill the need of both the variables. So, if we suppose out of these "*p*" explanatory variables, "*r*" pairs of predictor are correlated/ associated, then, out of these "*r*" pairs, only one from each pair can be taken in the model, thereby, reducing the number of predictors in the model to "*p-r*". The reduced model will then be given by:-

$$\log(T_i) = \beta_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i} + \gamma_3 x_{3i} + \ldots\ldots\ldots \gamma_{p-r} x_{(p-r)i} + \sigma\varepsilon_i \quad (2)$$

where the notations have their usual meaning.

To test the independence between two categorical variables, chi- square test of independence is applied. If p-value is less than the level of significance (0.05 in our case), null hypothesis that variables are not independent is rejected.

This model fitted by proposed method is likely to have lower standard error due to the elimination of correlated predictors from the true model thereby resulting in better fit than the true model. Different models can be compared by their AICs where AIC is:-

$$AIC = -2LL + 2(a + c) \quad (3)$$

Where *LL* = Log-likelihood of the model, a = number of parameters of the assumed probability distribution (for example; *a* = 2 for Log-Logistic AFT model as there are two parameters involved) and *c*, the number of coefficients (excluding constant) in the final model. The model with smaller value of AIC can be considered as a better model compared to other models under consideration.

3. RESULTS AND DISCUSSIONS

To illustrate the methodology, the survival time of 767 AIDS patients from 2004 - 2014 from Ram Manohar Lohia Hospital, Delhi, were recorded, out of which 384 were censored and rest were uncensored. This hospital serves as a referral center for health centers in the District. The hospital has a unit for HIV/AIDS patients on treatment. The inclusion criteria included been patients who were given ART treatment, aged more than 18 years, both gender and willing to provide written consent. Patients who were not willing to participate and not able to provide written consent were all excluded from the study. Survival time for each patient is recorded in years. An Accelerated Failure time model, taking all the possible predictors, is fitted. Results are shown in the Table **1** below.

Table 1: Log-Likelihood for Comparing Different AFT Models

Distribution	Log- Likelihood	Degrees of Freedom
Exponential	-1086.9	18
Weibull	-798.3	18
Log- Normal	-918	18
Logistic	-662.7	18
Log- logistic	-836	18

We compared all these models using statistical criterion (likelihood ratio test and AIC). The nested AFT models can be compared using the likelihood ratio (LR) test. The exponential model, the Weibull model and the log-normal model are nested within the gamma model (Table **1**).

However, the LR test is not valid for comparing models that are not nested. In such cases, AIC is used to compare the models (Table **2**) (The smaller AIC is the better).

According to the LR test, the logistic model fits better (Table **1**). Also, from Table **2**, AIC is least for Logistic distribution. So, the Logistic AFT model appears to be an appropriate AFT model according to LR test and AIC compared with other AFT models. We also note that the Log- Normal and exponential model are poorer fits according to LR test and AIC.

The AFTM results have been presented in Table **3**. It has been observed that patients who are older have shorter survival time than patients who are younger (as Time ratio (TR) < 1 for Age), female patients and eunuchs had shorter survival time than their male counter parts (as TR <1 for both). Similarly, East and Central patients are expected to have longer survival time than North Indian patients (TR >1). Patients who don't smoke and drink alcohol are observed to have better survival time than the patients who smoke and drink (TR >1).

Table 2: Akaike Information Criterion (AIC) Values for AFT Models

Distribution	Log-Likelihood	DF	C	AIC
Exponential	-1086.9	18	1	2210.4
Weibull	-798.3	18	2	1636.6
Log- Normal	-918	18	2	1876
Logistic	-662.7	18	2	1349.4
Log- Logistic	-836	18	2	1712

Table 3: Results of Logistic AFTM Model for HIV/ AIDS Patients

Predictors	β	Std. Error	TR	95% C.I
Age	-0.0118	0.00804	0.98827	(1.30, 3.98)
Male			1	
Female	-0.7025	0.34583	0.49536	(-0.027 ,0.003)
Eunuch	-1.0992	0.00000	0.333142	(-1.099, 0.065)
North India			1	
Central India	6.8151	0.00000	1.513	(6.815, 7.086)
East India	0.1864	0.77448	1.204925	(-1.331 ,1.704)
Smoking- Yes			1	
Smoking- No	0.1252	0.22841	1.133381	(-0.322, 0.572)
ALCOHOL- Yes			1	
ALCOHOL- No	0.1433	0.16252	1.154103	(-0.175,0.461)
DRUGS- Never			1	
DRUGS- Past	5.8478	0.18161	1.474	(5.491, 6.203)
DRUGS- Yes	0.7756	1.0675	0.571879	(-1.295,2.847)
Opportunistic Infection- Viral			1	
Opportunistic Infection- Bacterial	-0.6424	0.14163	0.526036	(-0.919,-0.365)
Opportunistic Infection- Fungal	0.6009	0.26871	1.823728	(0.074,1.127)
Urban			1	
Rural	-0.5271	0.18308	0.59031	(-0.885,-0.168)
Government Employee			1	
Non- Working	0.5102	0.35621	1.665702	(-0.187,1.208)
Agricultural Labor	0.1055	0.18550	1.111221	(-0.258,0.469)
Regular Employee	0.529	0.21158	1.697212	(0.114,0.943)
Business Man	-0.5673	0.32049	0.56704	(-1.195,0.060)
Un Married			1	
Married	0.3221	0.54588	1.379957	(-0.747 ,1.391)
Spouse- Positive			1	
Spouse- Negative	0.352	0.14234	1.421864	(0.072,0.630)

Then we have applied One- way Anova and Chi-Squared tests to determine whether 2 variables are independent or not. One- way Anova is used to test the dependence between a categorical and numerical variable where Chi- Squared test is used to the test the correlation/ association between two categorical

Table 4: Result of the Tests to Test Dependence between Predictors (p-Value)

	Age	Sex	State	Smoking	Alcohol	Drugs	OI	LS	OC	MR	Spouse
Age											
Sex	**0.00**										
State	0.429	0.952									
Smoking	0.539	**0.00**	0.688								
Alcohol	**0.0284**	**0.00**	0.325	**0.00**							
Drugs	0.174	0.211	0.674	0.731	**0.00**						
OI	0.522	**0.017**	0.253	**0.001**	0.610	0.08					
LS	0.0989	0.791	**0.00**	**0.00**	0.651	0.237	0.073				
OC	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.001**	0.124	0.304			
MR	0.562	**0.008**	0.926	0.652	**0.037**	**0.00**	0.617	0.760	0.109		
Spouse	**0.00**	**0.00**	0.215	0.346	**0.00**	0.509	0.101	0.909	**0.00**	0.956	

variables. So, to test the correlation/ association of the predictor "Age" with other predictors, we have used One- way Anova whereas to test the dependence between other variables, we have used Chi- Squared test. Results of the tests are given in the Table **4** above.

The p- values which are bold in the table are the significant values. They signify that variables are significantly correlated. The pairs of predictors which are found to be significantly correlated/ associated are listed below in the Table **5** below.

Table 5: Significantly Correlated/Associated Predictors

Predictor	Significantly correlated predictors
Age	Sex, Alcohol, Occupation, Spouse
Sex	Smoking, Alcohol, OI, Occupation, MR, Spouse
State	Living Status, Occupation
Smoking	Alcohol, OI, Living Status, Occupation
Alcohol	Drugs, Occupation, MR, Spouse
Drugs	Occupation, Spouse
Spouse	Occupation

Next, among the significantly found correlated/ associated predictors, the predictor which is significantly dependent on another variable is determined by fitting two models, taking one predictor dependent and another independent at a time and then taking another one independent and one which was taken independent as dependent next time. Then, two models are compared on the basis of their AIC. The one which has lower AIC is chosen and the independent predictor of the chosen model is then included in the AFTM and the dependent predictor is left out (Table **4**. Logistic, Linear, Multinomial Logistic models are fitted depending upon the type of dependent variable. The fitted models along with the type of models and their corresponding AIC values are shown in the Table **6** below:-

So, we keep only those independent predictors in the AFTM whose model have lower AIC compared to the model in which the other predictor of the significantly correlated pair is taken as the independent variable. Finally, only 5 variables are selected, namely, Age, Opportunistic Infections, Living Status, Drugs, Occupation.

Then the survival times of the patients are estimated using AFTM with only 5 selected predictors. Again, for this model, Logistic model is found to be appropriate AFT model according to LR test and AIC. Results have been presented in the Table **7** below:-

Also, variables are also selected using LASSO and elastic-net variable selection methods. Six variables, viz, Age, Occupation, Opportunistic Infections, State, Sex and drugs are selected using LASSO method. Elastic- net method selects Age, Alcohol status, Smoking Status, Spouse, and Occupation. Also, the standard errors of the coefficients of predictors in the different models are compared. It is observed that the standard error of the coefficients chosen by the proposed method is less than the standard error of the coefficients in the true model and with the standard errors of the coefficients chosen by LASSO and Elastic- net methods as shown in the Table **8** below:-

Table 6: AIC Values for Choosing Suitable Predictors in the Modified Model

Model	Dependent	Independent	AIC
Logistic	Smoking	Sex	**482.14**
Multinomial Logistic	Sex	Smoking	870.95
Linear	Age	Sex	5460.234
Multinomial Logistic	Sex	Age	**894.86**
Logistic	Alcohol	Age	**1040.7**
Linear	Age	Alcohol	5481.82
Logistic	Alcohol	Sex	763.23
Multinomial Logistic	Sex	Alcohol	**640.86**
Logistic	Alcohol	Smoking	1005.2
Logistic	Smoking	Alcohol	**493.71**
Logistic	Alcohol	Drugs	999.95
Logistic	Drugs	Alcohol	**490.71**
Multinomial Logistic	OI	Sex	1388.008
Multinomial Logistic	Sex	OI	**912.48**
Logistic	Smoking	OI	**533.23**
Multinomial Logistic	OI	Smoking	1385.52
Logistic	Living Status	State	790.01
Multinomial Logistic	State	Living Status	**175.97**
Logistic	Smoking	Living Status	**514.86**
Logistic	Living Status	Smoking	788.7
Multinomial Logistic	Occupation	Age	**2244.59**
Linear	Age	Occupation	5465.58
Multinomial Logistic	Occupation	Age	1632.22
Multinomial Logistic	Sex	Occupation	**894.98**
Multinomial Logistic	Occupation	State	2289.32
Multinomial Logistic	State	Occupation	**196.62**
Logistic	Smoking	Occupation	**533.6**
Multinomial Logistic	Occupation	Smoking	2256.34
Logistic	Alcohol	Occupation	**1036.1**
Multinomial Logistic	Occupation	Alcohol	2065.44
Multinomial Logistic	Drugs	Occupation	**523.02**
Multinomial Logistic	Occupation	Drugs	2268.79
Logistic	Marital Status	Sex	**65.89**
Multinomial Logistic	Sex	Marital Status	914.83
Logistic	Marital Status	Alcohol	67.68
Logistic	Alcohol	Marital Status	1039
Logistic	Marital Status	Drugs	**46.78**
Logistic	Drugs	Marital Status	512.99
Logistic	Spouse	Age	**1026.7**
Linear	Age	Spouse	5446.81
Logistic	Spouse	Sex	948.45
Multinomial Logistic	Sex	Spouse	**804.47**
Logistic	Spouse	Alcohol	1047.5
Logistic	Alcohol	Spouse	**1026.1**
Logistic	Spouse	Occupation	**1054.7**
Multinomial Logistic	Occupation	Spouse	2193.49

Table 7: Results of Modified Logistic AFTM Model for HIV/ AIDS Patients

Predictors	β	Std. Error	TR	95% C.I
Age	-0.0128	0.00804	0.98727	(-0.028,0.002)
DRUGS- Never			1	
DRUGS- Past	5.9812	0.16881	1.9012	(5.650,6.312)
DRUGS- Yes	0.6801	1.05695	0.474171	(-1.412,2.772)
Opportunistic Infection- Viral			1	
Opportunistic Infection- Bacterial	-0.6289	0.14144	0.53317	(-0.906,-0.351)
Opportunistic Infection- Fungal	0.6732	0.26792	1.960507	(0.148,1.198)
Urban			1	
Rural	-0.5402	0.18111	0.58264	(-0.895,-0.185)
Government Employee			1	
Non- Working	0.1671	0.1747	1.181901	(-0.175,0.509)
Agricultural Labor	0.0523	0.1829	1.05368	(-0.306,0.410)
Regular Employee	0.5042	0.21123	1.655607	(0.090,0.918)
Business Man	-0.6129	0.31784	0.541779	(-1.235,0.010)

Table 8: Comparison of the Standard Error of the Coefficients of the Predictors Selected using Different Variable Selection Methods

Predictors	S.E (True model)	S.E (Proposed Method)	S.E (LASSO Method)	S.E (Net- Elastic Method)
Age	0.00804	0.00796	0.00799	0.00801
Male				
Female	0.34583		0.32554	
Eunuch	0		0	
North India				
Central India	0		-0.001	
East India	0.77448		0.75256	
Smoking- Yes				
Smoking- No	0.22841			0.20144
ALCOHOL- Yes				
ALCOHOL- No	0.16252			0.12564
DRUGS- Never				
DRUGS- Past	0.18161	0.16881	0.17512	
DRUGS- Yes	1.0675	1.05695	1.06114	
Opportunistic Infection- Viral				
Opportunistic Infection- Bacterial	0.14163	0.14144	0.14256	
Opportunistic Infection- Fungal	0.26871	0.26792	0.16852	
Urban				
Rural	0.18308	0.18111		
Government Employee				
Non- Working	0.35621	0.1747	0.1725	0.25879
Agricultural Labor	0.1855	0.1829	0.1810	0.1829
Regular Employee	0.21158	0.21123	0.22147	0.22346
Business Man	0.32049	0.31784	0.32247	0.31862
Un Married				
Married	0.54588			
Spouse- Positive				
Spouse- Negative	0.14234			0.13596

Also, it is observed that the model fitted by proposed method has less AIC as compared to the true model and LASSO and Elastic- Net models as shown in Table **9**:-

Table 9: Comparison of Proposed and Existing Variable Selection Methods on Basis of AIC Values

Model	No. of predictors	AIC
True	11	1349.4
Proposed Method	5	1316.4
LASSO method	6	1320.8
Net-Elastic method	5	1325.6

4. CONCLUSIONS

This study took a sample of seven hundred and sixty seven (767) patients who were diagnosed of HIV/AIDS within a period of 2004-2014. The prognostic factors were Age, Sex, Smoking, Drugs, Alcohol, Opportunistic Infections, Occupation, State, Living Status, Spouse and Marital Status. Since Logistic AFTM has the minimum AIC, therefore, it is considered to be the best fit model. Then the independence between each pair of prognostic factors is tested and significantly correlated pairs are selected. For these pairs, suitable models are fitted to identify the dependent and independent predictors. Finally, only five predictors, viz, Age, Drugs, Opportunistic Infections, Living Status and Occupations are selected. The survival times are then again estimated using AFTM. Then the proposed method is compared with existing methods with respect to their AIC values It is found that the model fitted with proposed method has the minimum AIC value. So, it can be said that the proposed method is a good fit as compared to the other existing models.

REFERENCES

[1] AIDS info Fact Sheet, http://aidsinfo.nih.gov/2012

[2] WHO: Global update on HIV treatment 2013: Results, Impact and Opportunities 2013.

[3] UNAIDS (2014): a GAP report 2014.

[4] Ghate M, Deshpande S, Tripathy S, Godbole S, Nene M, *et al*. Mortality in HIV infected individuals in Pune, India. Ind J Med Res 2011; 133: 414-420.

[5] Rai S, Mahapatra B, Sircar S, Raj PY, Venkatesh S, *et al*. Adherence to antiretroviral therapy and its effect on survival of HIV-infected individuals in Jharkhand, India. PloS One 2013. DOI: 10.1371/journal.pone.0066860

[6] Kee MK, Lee JH, Kim EJ, Lee J, Nam JG, *et al*. Improvement in survival among HIV-infected individuals in the Republic of Korea: Need for an early HIV diagnosis. BMC Infect Dis 2009; 9: 128-128. https://doi.org/10.1186/1471-2334-9-128

[7] Jerene D, Endale A, Hailu Y, Lindtjørn B. Predictors of early death in a cohort of Ethiopian patients treated with HAART. BMC Infect Dis 2006; 6: 136-136. https://doi.org/10.1186/1471-2334-6-136

[8] Hernán MA, Cole SR, Margolick J, Cohen M, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. J. Pharmacoepidemiol. Drug Safety 2005; 14: 477-491. https://doi.org/10.1002/pds.1064

[9] Xue H, Lam KF, Cowling BJ, Wolf FD. Semi-parametric accelerated failure time regression analysis with application to interval-censored HIV/AIDS data. Stat Med 2006; 25: 3850-3863. https://doi.org/10.1002/sim.2486

[10] Grover G, Banerjee T. Estimation of survival times of HIV-1 infected children for doubly and interval censored data. Electron J App Sta Anal 2011; 4: 155-163. DOI: 10.1285/i20705948v4n2p155

[11] Nawumbeni DN, Luguterah A, Adampah T. Performance of Cox Proportional Hazard and Accelerated Failure Time Models in the Analysis of HIV/TB Co-infection Survival Data. Research on Humanities and Social Sciences 2014; 4(21): 94-102.

[12] Tarekegn S. The Effect of HAART on Incidence of Tuberculosis among HIV Infected Patient in Hawassa University Referral Hospital, South Ethiopia Clinics in Rungwe District, Tanzania. Unpublished dissertation in partial fulfillment of the requirements for the degree of Master of Science, Tanzania 2011.

[13] Musenge E, Vounatsou P, Collinson M, Tollman S, Kaln K. The Contribution of the Spatial Analysis to Understanding HIV/TB Mortality in Children a Structural Equation Modeling Approach. Global Health Action 2013 PMCID: PMC35566702 DOI: 103402/gha.v6i0.19266.

[14] Grover G, Kumar SP. Determination of predictors associated with HIV/AIDS patients on ART using Accelerated Failure Time model for interval, American Journal of BioStatistics (USA) 2016; 6(1): ISSN: 1948-9897.

[15] Grover G, Gadpayle AK, Varshney MK. On the estimation of probability of death of AIDS patients in the presence of competing risks, Aligarh Journal of Statistics 2012; 32: 69-83.

[16] Grover G, Ravi V. On the estimation of expected survival time of AIDS patients undergoing Antiretroviral therapy using generalized Poisson regression model, Turkiye Klinikleri J BioStat (Turkey) 2014; 6(1): 538-541. ISSN: 2249-555X

[17] Tibshirani R. The lasso method for variable selection in the Cox model. Statistics in Medicine 1997; 16(4): 385-395. https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

[18] Zou H, Hastie T. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2005; 67(2): 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Received on 27-06-2018 Accepted on 27-08-2018 Published on 25-10-2018

https://doi.org/10.6000/1929-6029.2018.07.04.3