Estimation of Survival Time of HIV/AIDS Patients While Reducing the Number of Predictors using a Partial Correlation/Association Technique

Gurprit Grover and Anurag Sharma*

Department of Statistics, University of Delhi, Delhi- 110007, India
*Corresponding Author E-mail : anuragsharma532@gmail.com

http://dx.doi.org/10.13005/bpj/1932

(Received: 20 January 2020; accepted: 16 May 2020)

A number of techniques have been developed for variable reduction in the survival analysis. However, none of those techniques take into consideration the correlation among the predictors. This paper focuses on the reducing the number of predictors using the concept of partial correlation/association and then estimating the survival time of HIV/AIDS patients while. Partial correlation is preferred over pairwise correlation & multiple correlation as pairwise correlation takes into account only the linear relationship between two predictors at a time and multiple correlation may give misleading results if one variable is numerically related to other variable and both the variables again are taken simultaneously to find the association/correlation. This can be avoided by controlling the confounding variable using partial correlation coefficient. Partial correlation is used to assess the correlation between a continuous predictor and a categorical predictor when the controlling variable is also categorical. ANCOVA is used to determine the association between two categorical predictors when the controlling variable is also categorical. AFT models are used to estimate survival times with/without reduction in the number of predictors. It is observed that the estimated survival times are not affected by the reduction of predictors from 11 to 4. Also, the estimates obtained by the new proposed technique are found to be more efficient than the existing methods.

Keywords: AIDS; AFT; ANCOVA; HIV.

Presently, Human Immuno deficiency virus (HIV) has been assumed to have a measure of an international epidemic1, 2 and it’s notably most striking in Asia3. Worldwide about 36 Million persons have been estimated to be infected with HIV. It has also resulted in death of 1.2 million people globally4. Of all HIV infected people, 90% live in developing countries1. India has third largest HIV epidemic in the world. In India, 2.5 million individuals were estimated to have been living with HIV infection till 20144. All the above mentioned figures designate the enormous measure of HIV and as a result validate a crucial requirement to conduct a depth research of the calamity of various prognostic factors to which the HIV patients are left unprotected to.

Anti-retroviral therapy (ART) plays a crucial role in treating the patients infected with HIV/AIDS. It has the capability to reduce mortality as well as morbidity rates among HIV patients, and as a result help in improving the life quality of patients. This treatment (ART) has been available.
freely for patients since 2004 in India. At ART clinics, HIV patients have access to nutritional advice, testing and counseling (HTC) by experts. They are needed to go through a CD4 count test half yearly5. Moreover, the government is also rolling out several campaigns to aware people about their appointments to increase the overall attendance6.

With the use of ART and vast improvements in preventive facilities, HIV epidemic in India has slowed down considerably with a 19% reduction in new HIV infections (130,000 in 2013), and a reduction of 38% in deaths contributed by AIDS between 2005 and 20134. By the end of 2013, more than 700,000 people were receiving ART which is a great achievement in itself considering that it is the second largest number of people on treatment in any single country.

To check the intensity of HIV virus in the AIDS patients, one of the most important predictors is the CD4 cell counts. A healthy person is estimated to have 1200 CD4 cells per mm3 in blood. CD4 cells are deteriorated as immune system is attacked by HIV virus.

Center for Disease Control (CDC) recommended initiating Anti-retroviral therapy (ART) when the CD4 cells goes down 200 cells per . and Various observational studies and Two randomized controlled trials (RCTs) showed that initiation of ART at CD4 < 350 cells/mm3 results in significantly reduction of disease progression, mortality and the incidence of opportunistic diseases, especially TB and non-AIDS-defining conditions6.

Various studies have been proposed to estimate the survival and effects of underlying covariates on the survival among HIV patients by using Cox PH model (Ghate et al., 2011; Rai et al., 2013; Kee et al., 2009; Jerene et al., 2006)10. Markus Abigo Erango. (2017) predicted and compared the survival time of HIV/AIDS patients in three hospitals in Ethiopia. Three parametric accelerated failure time distributions: lognormal, logistic and Weibull are used to analyze, predict and compare survival probabilities of the patients. The results indicated that the empirical hazard rates of the three data sets reveal maximal peaks11. Jaya Chakravarty et al. (2013) analyzed the factors affecting the survival of patients under routine programme conditions in an ART centre in India five years after its initiation. She found poor survival in the first half year of therapy mainly in patients with severe immunosuppression. Thus, she emphasized the requirement for early enrolment into the programme12. Performance of Cox Proportional Hazard Model and the Accelerated Failure Time Model was compared by Derek Ngbandor Nawumbeni et al. (2014) using HIV/TB Co-infection Survival data. They concluded that the Accelerated Failure Time model has better predictive power as compared to the Cox model13.

Emeka E. Orisakwe et al., (2012) determined whether knowledge of HIV and the attitude of patients referred for HCT correlated with a willingness to test for HIV14. Colins Kingoum Nubed (2016) assessed the KAPs of senior secondary school students in Fako Division, Cameroon, on HIV/AIDS. This cross-sectional study was conducted on 464 students between 13 to 25 years of age. Participants were selected by systematic quota random sampling from secondary schools in Fako, from April to June 201415. Hanhui Ye et al., (2016) conducted an analysis to analyze the correlation between AIDS restriction and metabolic pathway gene expression. They showed that HIV-1 postentry cellular viral from AIDS restriction genes can be coexpressed in human transcriptome microarray datasets16. Grover and Swain (2013) tried to estimate the survival time of HIV/AIDS patients who underwent Antiretroviral Therapy in Delhi. They used Non-Parametric Maximum Likelihood Estimation method (NPMLE) for interval censored data and KM survival time for left, right and mid-point imputation to estimate the survival of these patients17. Richard Hafner et. al (1999) studied the linear relationship between Mycobacterium avium complex (MAC) infection of blood and bone marrow in human immunodeficiency virus—infected patients before and during treatment18. Sergio Rueda (2016) conducted a systematic review and series of meta-analyses on the association between HIV related stigma and health among people living with HIV. They conducted a structured search on 6 electronic databases for journal articles reporting associations between HIV-related stigma and health related outcomes published between 1996 and 201319. Ravi et al (2018) determined highly positively correlated predictors and reduced the number of predictors using pairwise correlation / association and then estimated the survival time patients on
ART. Grover G et al. (2018) examined the effect of reduction of predictors affecting the survival of patients using a multiple correlation/technique. They observed that the estimated survival times are not affected by the reduction of predictors from 11 to 4.

In this paper, we have proposed a partial correlation/association 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. So, this research paper aims at

- Estimating the survival time of patients on ART by taking all the prognostic factors by using AFTM;
- Determining the significantly partially correlated/associated predictors in the model by using suitable technique;
- Determining the effect of partially correlated predictors on the survival time of HIV/AIDS patients under ART;
- Estimating the survival times of HIV/AIDS patients on ART in the presence of reduced number of predictors
- Compare the two models (One having all the available prognostic factors and other with the reduced number of prognostic factors) on the basis of their AIC values.
- Comparing the proposed method with the methods discussed in and existing variable reduction methods.

We have used SPSS and R software to fit the models and to compute the associations/correlations among predictors.

**METHODS AND MATERIALS**

**Methods Used**

Assuming that number of HIV/AIDS patients under ART as and survival time of patient as with survival function. Then assuming a linear relationship between and predictors, we have

\[
\log(T_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_p x_{ip} + \sigma_i
\]

Where

- \(\beta_0\) is the intercept
- \(\beta_i\) are the coefficients of “p” explanatory variables for ith patient.
- \(\sigma\) is the scale parameter

**Table 1. Log- Likelihood to compare 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                 |

**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  |
ANCOVA is used to test the association between two categorical predictors with the controlling variable being also categorical predictor.

To test the independence among two categorical predictors while the controlling predictor is also categorical, ANCOVA is used and then on the basis of p-value, we rejected or accepted null hypothesis that these variables are partially independent or not.

**Model Comparison**

Akaike’s Information Criteria (AIC) is used to compare the different models where AIC is given by:

$$AIC = -2k + 2ln(\hat{L}) \quad \ldots (2)$$

where k be the number of estimated parameters in the model and $\hat{L}$ be the maximum value of the likelihood function for the model. The model with smaller value of AIC can be considered as a better model compared to other models under consideration.

**Data Sources and Results**

Data of 767 HIV/AIDS patients undergoing ART in Dr. Ram Manohar Lohia Hospital, New Delhi, India, during the period of January 2004 to December 2014 were collected in this retrospective follow-up study and an Accelerated Failure time model, taking all the possible predictors, is fitted. Different AFT models are fitted to determine the best fitted model. Logistic model is found to be the best model. Results are shown in Table 1 and 2 below:

Two statistical criterion (likelihood ratio test and AIC) are used to compare these models. Gamma model nested exponential model, the Weibull model and the log-normal model (Table 1).

It is observed that the AIC for logistic model in the presence of all predictors is least among all the fitted models. Also, LR test showed that the logistic model fits best as compared to other models (Table1). Also, Cox-Snell residual fits good to the Logistic Accelerated Failure Time model (Figure 1). So, we fit the survival data using Logistic AFT model and results are shown in the table 3 below:

From the results of the table 3, males are supposed to have longer survival time than females and Eunuchs as the time ratio for both of these genders are less than 1 (0.495 for females and 0.333 for eunuchs). Similarly, patients living

---

Fig. 1. Cox-Snell residual plot for Logistic AFT model
in North India are expected to have shorter survival time than the patients living in Central and East India as their time ratios are greater than 1. Again, patients who were smokers or were alcoholic are observed to have shorter survival time than the ones who never smoked or non-alcoholic in their lives. Unmarried patients are expected to have shorter survival patients than the married patients. Patients who have bacterial infection are observed to have shorter survival time than the patients who have viral infection where those having fungal infection have longer survival time than the ones who have viral infection. Patients living in rural areas are expected to have shorter survival time than the ones who live in urban areas.

Then, in the first step, we have tried to determine the predictors which are jointly associated/ correlated with each other. We have used MANOVA to test the joint independence among two or more continuous and two or more categorical predictors and Log-linear models are used to test the joint independence among

### Table 3. Results of Logistic AFTM model for HIV/AIDS patients

| Predictors                | 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              | 0.1433     | 0.16252| 1.154103 (-0.175, 0.461) |
| ALCOHOL- No               |            |        | 1                |
| 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 | 0.6099 | 0.26871| 1.823728 (0.074, 1.127) |
| Urban                     | -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| 1.697122 (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                |

### Table 4. Result of the tests of the joint dependence among significantly associated/ correlated pairs of predictors

| Predictor | Jointly dependent predictors | Model used | p-value |
|-----------|------------------------------|------------|---------|
| Age       | Sex, Alcohol, Occupation and Spouse | MANOVA     | 0.015   |
| Sex       | Opportunistic infection, Marital Status and Spouse | Log-linear | 0.000   |
| Smoking   | Alcohol, living Status, Occupation and Opportunistic infection | Log-linear | 0.000   |
| Drugs     | Occupation and Spouse        | Log-linear | 0.000   |
Table 5. Significant values of Correlation and p-values for partial correlation/association

| Controlling variable | Variables                               | Correlation coefficient or p-value |
|----------------------|-----------------------------------------|-----------------------------------|
| Spouse               | Age, Sex                                | -0.109                            |
| Spouse               | Age, Alcohol                            | 0.045                             |
| Spouse               | Age, Occupation                         | 0.141                             |
| Spouse               | Sex, Alcohol                            | 0.502                             |
| Spouse               | Sex, Occupation                         | -0.152                            |
| Spouse               | Alcohol, Occupation                     | 0.93                              |
| Marital Status       | Sex, Opportunistic Infections           | 0.000                             |
| Sex                  | Age, Alcohol                            | -0.019                            |
| Sex                  | Age, Occupation                         | 0.135                             |
| Sex                  | Age, Spouse                             | -0.170                            |
| Sex                  | Alcohol, Occupation                     | 0.56                              |
| Sex                  | Alcohol, Spouse                         | 0.51                              |
| Sex                  | Spouse, Occupation                      | -0.059                            |
| Alcohol              | Age, Occupation                         | 0.157                             |
| Alcohol              | Age, Spouse                             | -0.216                            |
| Alcohol              | Age, Sex                                | 0.68                              |
| Alcohol              | Spouse, Occupation                      | -0.11                             |
| Alcohol              | Sex, Occupation                         | 0.53                              |
| Alcohol              | Spouse, Sex                             | 0.35                              |
| Occupation           | Age, Spouse                             | -0.209                            |
| Occupation           | Age, Sex                                | -0.158                            |
| Occupation           | Age, Alcohol                            | 0.62                              |
| Occupation           | Spouse, Sex                             | 0.69                              |
| Occupation           | Spouse, Alcohol                         | -0.147                            |
| Occupation           | Sex, alcohol                            | 0.510                             |
| Spouse               | Sex, Opportunistic Infections           | 0.000                             |
| Opportunistic Infections | Sex, Spouse                | 0.000                             |
| Opportunistic Infections | Sex, Marital Status               | 0.51                              |
| Spouse               | Sex, Marital Status                     | 0.000                             |
| Marital Status       | Sex, Spouse                             | 0.000                             |
| Opportunistic Infections | Sex, Spouse                | 0.000                             |
| Opportunistic Infections | Marital Status, Spouse           | 0.66                              |
| Spouse               | Marital Status, Opportunistic Infections| 0.810                             |
| Marital Status       | Spouse, Opportunistic Infections        | 0.207                             |
| Occupation           | Smoking, Living Status                  | 0.000                             |
| Occupation           | Smoking, Alcohol Status                 | 0.000                             |
| Occupation           | Smoking, Opportunistic Infections       | 0.004                             |
| Alcohol              | Smoking, Opportunistic Infections       | 0.000                             |
| Alcohol              | Smoking, Occupations                    | 0.000                             |
| Alcohol              | Smoking, Living status                  | 0.610                             |
| Opportunistic Infections | Smoking, Living status                | 0.000                             |
| Opportunistic Infections | Smoking, Alcohol                        | 0.06                              |
| Opportunistic Infections | Smoking, Living Status                | 0.243                             |
| Opportunistic Infections | Smoking, Occupations                  | 0.000                             |
| Living Status        | Smoking, Occupations                   | 0.083                             |
| Living Status        | Smoking, Opportunistic Infections       | 0.000                             |
| Living Status        | Smoking, Alcohol                        | 0.753                             |
| Living Status        | Alcohol, Opportunistic Infections       | 0.000                             |
| Living Status        | Alcohol, Occupations                    | 0.040                             |
| Occupation           | Living Status, Opportunistic Infections | 0.000                             |
| Spouse               | Drugs, Occupations                      | 0.159                             |
categorical predictors only. Predictors which are found to be jointly significantly correlated/associated are listed below in the Table 4 below:-

Since p-value for each group is less than 0.05, alternative hypothesis can be accepted and can be concluded that these predictors are jointly dependent. After this partial correlation/association between each possible pair are determined using suitable methods. The results are shown in the Table 5 below. The italic values are p-values and bold values are the significant values of correlation (Correlation coefficient > .50 and p-values less than 0.05).

Using this procedure, 4 predictors are selected. In the next step, survival times are estimated using AFTM with only these selected predictors. Again, for this model, Logistic model is found to be the best AFT model by LR test and AIC values. Again, Cox-Snell residual was plotted and it fits well as shown in figure 2. Results are presented below in Table 6:-

| Predictors   | β         | Std. Error | TR        | 95% C.I       |
|--------------|-----------|------------|-----------|---------------|
| Age          | -0.00818  | 0.00796    | 0.9918568 | -0.0242,0.008 |
| DRUGS- Never |           |            | 1         |               |
| DRUGS- Past  | 6.26359   | 0.16436    | 0.0980558 |               |
| DRUGS- Yes   | 0.96137   | 1.05338    | 0.6152831 | -1.240,3.163  |
| Male         |           |            | 1         |               |
| Female       | -1.76615  | 0.1548     | 0.1709898 | -2.0695,-1.462|
| Eunuch       | -1.62039  | 0.0125     | 0.1978218 | -1.620,0.624  |
| Smoking- Yes |           |            | 1         |               |
| Smoking- No  | 0.3776    | 0.19208    | 1.4587837 | -0.077,0.832  |

**Fig. 2.** Cox-Snell residual plot for Logistic AFT model
To compare the proposed method with the existing ones, variables are also selected using LASSO and elastic-net variable selection methods. LASSO method selected six variables, viz, Age, Occupation, Opportunistic Infections, State, Sex and drugs whereas Elastic-net method selects Age, Alcohol status, Smoking Status, Spouse, and Occupation. Also, the standard errors of the coefficients of predictors in the models fitted by different methods 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 7 below:

From table 8, it can be observed that AIC value is least for proposed method when compared with the AIC values for True, LASSO and Elastic-Net models and the methods discussed in 20 and 21.

**CONCLUSION**

Data of 767 patients diagnosed with HIV/AIDS were collected in this analysis. The AFT modelling of the survival rates of patients

---

**Table 7. 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) | S.E (Pairwise Correlation technique) | S.E (Multiple Correlation technique) |
|------------|------------------|-----------------------|--------------------|--------------------------|--------------------------------------|--------------------------------------|
| Age        | 0.00804          | 0.00796               | 0.00799            | 0.00801                  | 0.00796                              | 0.00708                              |
| Male       | 0.34583          | 0.1548                | 0.32554            | 0                        | 0.33493                              |                                     |
| Female     | 0.0125           | 0.0125                | 0                  | 0                        | 0                                    |                                     |
| Eunuch     | 0.77448          | 0.75256               |                    |                          |                                      |                                     |
| North India|                  |                       |                    |                          |                                      |                                     |
| Central India|                |                       | 0                  | 0                        | 0                                    |                                     |
| East India | 0.18308          | 0.18111               | 0.18087            | 0.18728                  | 0                                    |                                     |

---

684  
GROVER & SHARMA, Biomed. & Pharmacol. J, Vol. 13(2), 677-686 (2020)
Table 8. Comparison of proposed and existing variable selection methods on basis of AIC values

| Model                 | No. of predictors | AIC      |
|-----------------------|-------------------|----------|
| True                  | 11                | 1349.4   |
| Proposed Method       | 4                 | 1295.2   |
| LASSO method          | 6                 | 1320.8   |
| Net-Elastic method    | 5                 | 1325.6   |
| Pairwise Correlation technique | 5 | 1316.4   |
| Multiple Correlation technique | 4 | 1313.2   |

from the prognostic factors selected Logistic as the best fitted model based on AIC. The Cox-Snell residual plot further confirmed the Logistic model as well fitted for evaluating the survival time of these 767 HIV/AIDS patients. Then, number of predictors have been tried to reduce by determining partial correlated/associated variables/predictors using partial correlation/ANCOVA. Using this technique, 4 predictors have been selected. It was disclosed that the model fitted with proposed technique is a good fit than the models fitted with all other methods. Therefore, from this study it can be concluded that it is not necessary to take all the predictors are not needed to take in the model to estimate the survival time of the patients as many prognostic factors may be partially correlated/associated with each other and taking only one of them may satisfy the need of all of the remaining correlated/associated predictors.

ACKNOWLEDGMENT

We sincerely acknowledge and thank Dr. Ram Manohar Lohia hospital for permitting us to use the HIV/AIDS database.

REFERENCES

1. Quinn, T.C. Global burden of the HIV pandemic. *Lancet*; **348**: 99-104 (1996).
2. Mann, J.M., Tarantola, D.J.M. HIV—The global picture. *Sci Am*; **279**: 82-83 (1998).
3. Islam, A., Zaffar and Tahir, M. Health sector reform in South Asia: New challenges and constraints. *Health Policy*; **60**: 151-69 (2002).
4. UNAIDS (2014): a GAP report 2014.
5. NACO (2013) ‘India: Annual Report 2012-13.
6. WHO (2013): Global update on HIV treatment 2013: Results, Impact and Opportunities.
method: an application to interval censored data. *American Journal of Mathematics and Statistics*, 3(4): 213-219 (2013).

18. Hafner, R., Inderlied, C. B., Peterson, D. M., Wright, D. J., Standiford, H. C., Drusano, G., & Muth, K. Correlation of quantitative bone marrow and blood cultures in AIDS patients with disseminated *Mycobacterium avium* complex infection. *The Journal of infectious diseases*, 180(2): 438-447 (1999).

19. Rueda, S., Mitra, S., Chen, S., Gogolishvili, D., Globerman, J., Chambers, L., ... & Rourke, S. B. Examining the associations between HIV-related stigma and health outcomes in people living with HIV/AIDS: a series of meta-analyses. *BMJ open*, 6(7): e011453 (2016).

20. Sharma A, Ravi V, Sehgal VK, Grover G. Comparing the Performance of Different High Dimensional Variable Selection Techniques on the Low Dimensional HIV/AIDS Data set. *J Commun Dis* 2020; 52(1): 14-21.

21. Grover G, Sharma A. Examining the Effect of Reduction of Predictors Affecting the Survival Time of HIV/AIDS Patients using a Multiple Correlation/ Association Technique. *J Commun Dis* 2018; 50(3): 15-21.

22. Grover G, Sharma A. Estimating the Survival Time of HIV/AIDS Patients By Reducing The Number Of Patients Using Canonical Correlation Technique, *Int J Biol Med Res*. 2019; 10(2):6714-6719.

23. Sharma A, Ravi V, Sehgal VK, Grover G. Comparing the Performance of Different High Dimensional Variable Selection Techniques on the Low Dimensional HIV/AIDS Data set. *J Commun Dis* 2020; 52(1): 14-21.