Examining the Effect of Reduction of Predictors Affecting the Survival Time of HIV/AIDS Patients using a Multiple Correlation/Association Technique

Gurprit Grover¹, Anurag Sharma²

¹,²Department of Statistics, University of Delhi, Delhi- 110007, India.
DOI: https://doi.org/10.24321/0019.5138.201815

Abstract

The main objective of this paper is to reduce the number of predictors using multiple correlation/association while estimating the survival of HIV/AIDS. MANOVA is used to test the association when we have two or more continuous predictors and two or more categorical predictors. Also, log-linear models are used to test the joint independence of two or more predictors. The survival times are estimated using AFTM. It is observed that the estimated survival times are not affected by the reduction of predictors from 11 to 4. Also, the estimates so obtained are more efficient as they have reduced standard error. Results of the proposed reduction method were also compared with the results of the two existing variable reduction methods, viz, LASSO and Net- Elastic method.

Keywords: HIV, AIDS, AFT, MANOVA, Log-Linear models

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.

Corresponding Author: Anurag Sharma, Department of Statistics, University of Delhi, Delhi- 110007, India.
E-mail Id: anuragsharma532@gmail.com
Orcid Id: https://orcid.org/0000-0002-3482-0774

How to cite this article: 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.
Several studies have been proposed for the estimation of HIV populations and underlying covariate effect on the hazard of death among HIV patients by using Cox proportional model (Ghate et al., 2011 [4]; Rai et al., 2013[3]; Kee et al., 2009[6]; Jerene et al., 2006).[7]

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, log 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 peaks.[8]

Jaya Chakravarty et al. (2013) assessed the factors determining survival of patients on ART under routine programme conditions in an ART centre in north India five years after its inception. The study findings revealed poor survival in the first six months of therapy especially in those with severe immunosuppression. This emphasizes the need for early enrolment into the programme. The high LFU occurring early after initiation of therapy suggests the urgent need to build an efficient patient retrieval system in the programme.[9]

Derek Ngbandor Nawumbeni et al. (2014) compares the performance of these two models viz. Cox Proportional Hazard Model and the Accelerated Failure Time Model using HIV/TB Co-infection Survival data. The study revealed that, the AFT model has the best predictive power compared to the Cox model based on the AIC and BIC values.[10]

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 HIV.[11]

Colins Kingoum Nubed (2016) aimed at assessing the KAPs of senior secondary school students in Fako Division, Cameroon, on HIV/AIDS. This was a cross-sectional study carried out on 464 students aged 13–25 years, selected by systematic quota random sampling from some secondary schools in Fako, from April to June 2014, to evaluate their KAPs regarding HIV/AIDS.[12]

Hanhui Ye et al., (2016) conducted an analysis to determine the correlation between AIDS restriction and metabolic pathway gene expression. The results showed that HIV-1 postentry cellular viral cofactors from AIDS restriction genes are coexpressed in human transcriptome microarray datasets.[13]

Grover and Swain (2013) estimated the survival of HIV/AIDS patients who were undergoing Antiretroviral Therapy in an ART centre, Delhi. Non Parametric Maximum Likelihood Estimation NPMLE (E-M) for interval censoring and KM survival plot for left, right and mid-point imputation had been used to estimate the survival of these patients.[14]

Richard Hafner et. al (1999) studied the relationship between Mycobacterium avium complex (MAC) infection of blood and bone marrow in human immunodeficiency virus—infected patients before and during treatment.[15]

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 2013.16

However, none of these studies aimed at finding out the association among three or more predictors at a single time. In this paper, we have proposed a multi 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 HIV/AIDS patients on ART in the presence of all the prognostic factors by using AFTM
- Determine the significantly multiple correlated/associated predictors in the model
- Determine the effect of 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 with all the predictors and other with reduced number of predictors) on the basis of their AIC values

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

**Methods and Materials**

**Methods used**

Suppose that n number of HIV/AIDS patients are under ART, Let Ti be the survival time of ith patient with survival function $S(t)$. Then assuming a linear relationship between log $(T_i)$ and “p” predictors, we have

$$\log (T_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \ldots + \beta_n x_{ip} + \varepsilon_i$$

Where,

- $\beta_0$ is the intercept
- $\beta$’s are the coefficients of “p” explanatory variables for ith patient.
- $\sigma$ is the scale parameter
- $\varepsilon_i$ is a random variable used to model the deviation of values of log $(T_i)$ from the linear part of model.

It may be possible that two or more predictors are
associated with a single predictor; we can replace those predictors with a single predictor with which they are associated. For example, in the model, \((1)\), \(X_2\) and \(X_3\) may be significantly correlated/associated with \(X_1\), then we can replace \(X_2\) and \(X_3\) in the model with \(X_1\) and it will fulfill the need of all the three variables. So, if we suppose out of these “p” explanatory variables, “r_1” predictor are associated with a single predictor, then, instead of using these “r_1” predictors, we can use only that predictor with which they are associated. There by, reducing the number of predictors in the model to “p-r”. The modified model will then be given by:

\[
\log (T_i) = \beta_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \ldots + \gamma_{p-r} x_{i(p-r)} + \sigma e_i
\]  

where the notations have their usual meaning.

To test the independence among two or more continuous predictors and two or more categorical predictors, MANOVA is used and then on the basis of p-value, we rejected or accepted null hypothesis that these variables are jointly independent or not.

Also, Log-linear models for joint independence are used to test the independence when we have all the predictors as categorical predictors.

Model Comparison

These two models can be compared on the basis of their AIC values where AIC is given by:

\[
AIC = -2LL + 2(a + c)
\]  

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.

Data Sources

A retrospective follow-up study was conducted, involving 767 HIV/AIDS patients who were undergoing Antiretroviral Therapy in the ART centre of Dr. Ram Manohar Lohia Hospital, New Delhi, India, during the period of January 2004 to December 2014 and an Accelerated Failure time model, taking all the possible predictors, is fitted. Results are shown in 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                 |

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  |

All these models are compared using two 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).

Since the AIC for logistic model is least among all the fitted models. Also, according to the LR test, the logistic model fits better (Table1). So, we fit the survival data using Logistic AFT model and results are shown in the table 3 below:

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

| Predictors   | Std. Error | Time Ratio (TR) | 95% C.I          |
|--------------|------------|----------------|------------------|
| Age          | -0.0118    | 0.98827        | (1.30, 3.98)     |
| Male         | 1          |                |                  |
| Female       | -0.7025    | 0.49536        | (-0.027, 0.003)  |
| Eunuch       | -1.0992    | 0.333142       | (-1.099, 0.065)  |
| North India  | 1          |                |                  |
| Central India| 6.8151     | 1.513          | (6.815, 7.086)   |
| East India   | 0.1864     | 1.204925       | (-1.331, 1.704)  |
| Smoking- Yes | 1          |                |                  |
| Smoking- No  | 0.1252     | 1.133381       | (-0.322, 0.572)  |
| ALCOHOL- Yes | 1          |                |                  |
From the results of the table 3, it can be observed that females and Eunuchs are supposed to have shorter survival time than the males as the time ratio for both of these genders are less than 1 (0.495 for females and 0.333 for eunuchs). Also, patients living in central and east India are expected to have longer survival time as compared to the patients living in North India as their time ratios are greater than 1. Again, patients who smoked or were alcoholic are observed to have shorter survival time than the ones who never smoked or non-alcoholic in their lives. Married patients are expected to have longer survival patients than the unmarried 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, we have tried to determine the pairs of predictors which are associated/correlated with each other. We have used One-way Anova to test the dependence between a categorical and numerical variable whereas Chi-Squared test is used to test the correlation/association between two categorical variables. The pairs of predictors which are found to be significantly correlated/associated are listed below in the Table 4 below:

| Predictor          | Significantly correlated predictors                                      |
|--------------------|-------------------------------------------------------------------------|
| Age                | Sex, Alcohol, Occupation, Spouse                                        |
| Sex                | Smoking, Alcohol, Opportunity Infections (OI), Occupation, MR, Spouse    |
| State              | Living Status, Occupation                                               |
| Smoking            | Alcohol, OI, Living Status, Occupation                                   |
| Alcohol            | Drugs, Occupation, MR, Spouse                                            |
| Drugs              | Occupation, Spouse                                                      |
| Spouse             | Occupation                                                              |

After finding these pairs of associated/correlated predictors, we have tried to find the joint association among the predictors which are associated/correlated with a single predictor. For eg, since Sex, alcohol, occupation and Spouse status of the patients are found to be associated/correlated with Age, we have tried to find that whether these 4 predictors are jointly associated with Age or not. 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 categorical predictors only.
So, it can be seen that p-values for each group is less than 0.05, we can reject the null hypothesis that these predictors are not significantly jointly independent. As a result, we can jointly replace dependent predictors with the independent predictor. Then the survival times of the patients are estimated using AFTM with only 4 selected predictors. Again, for this model, Logistic model is found to be appropriate AFT model according to LR test and AIC values. Results are presented below in Table 6.

### Table 5. 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   |

Then the variables are reduced by two existing variable reduction methods namely, LASSO method and Net- Elastic method. LASSO method selected 5 predictors viz, Age, Occupation, Opportunistic Infections, State, Sex and drugs and Net- Elastic methods selected Age, Alcohol status, Smoking Status, Spouse, and Occupation. Also, it can be observed that the standard error of the coefficients of the predictors chose by the proposed method is less than the standard error of the coefficients of the predictors in the true model, LASSO method and Net- Elastic methods as shown in the table 7 below:

### Table 6. Results of modified Logistic AFTM model for HIV/AIDS patients

| Predictors       | $\beta$    | Std. Error | TR     | 95% C.I                |
|------------------|------------|------------|--------|------------------------|
| Age              | -0.00768   | 0.00708    | 0.9923536 | (-0.024,0.008)         |
| DRUGS- Never     |            |            | 1.0000  |                        |
| DRUGS- Past      | 6.1882     | 0.16728    | 1.9667244 | (5.86,6.52)            |
| DRUGS- Yes       | 0.99829    | 1.11051    | 2.7136364 | (-1.178,3.175)         |
| Male             |            |            | 1.0000  |                        |
| Female           | -1.66416   | 0.33493    | 0.18935 | (-2.320,-1.007)        |
| Eunuch           | -1.88715   | 0.33493    | 0.18935 | (-2.320,-1.007)        |
| Smoking- Yes     |            |            | 1.0000  |                        |
| Smoking- No      | 0.38687    | 0.21578    | 1.472366 | (-0.067,0.841)         |
| Government Employee |          |            | 1.0000  |                        |
| Non- Working     | 0.55018    | 0.16578    | 1.7335653 | (-0.161,1.261)         |
| Agricultural Labor | 0.0901    | 0.17885    | 1.0942814 | (-0.289,0.469)         |
| Regular Employee | 0.53221    | 0.30557    | 1.7026931 | (-0.107,0.957)         |
| Business Man     | -0.61725   | 0.30557    | 0.5394283 | (-1.298,0.063)         |
Also, the AIC value of the model fitted by the proposed method is least among the all models fitted by the different methods as shown in table 8.

Table 7. Comparison of the Standard Error of the coefficients of the True model and model fitted by proposed method

| Predictors                              | S.E (True model) | S.E (Proposed method) | S.E (LASSO Method) | S.E (Net- Elastic Method) |
|-----------------------------------------|------------------|-----------------------|-------------------|--------------------------|
| Age                                     | 0.00804          | 0.00708               | 0.00799           | 0.00801                  |
| Male                                    |                  |                       |                   |                          |
| Female                                  | 0.34583          | 0.33493               | 0.32554           |                          |
| Eunuch                                  | 0                | 0                     | 0                 |                          |
| North India                             |                  |                       |                   |                          |
| Central India                           | 0                | -0.001                |                   |                          |
| East India                              | 0.77448          | 0.75256               |                   |                          |
| Smoking- Yes                            |                  |                       |                   |                          |
| Smoking- No                             | 0.22841          | 0.21578               |                   | 0.20144                  |
| ALCOHOL- Yes                            |                  |                       |                   |                          |
| ALCOHOL- No                             | 0.16252          | 0.12564               |                   |                          |
| DRUGS- Never                            |                  |                       |                   |                          |
| DRUGS- Past                             | 0.18161          | 0.16728               | 0.17512           |                          |
| DRUGS- Yes                              | 1.0675           | 1.03011               | 1.06114           |                          |
| Opportunistic Infection- Viral          |                  |                       |                   |                          |
| Opportunistic Infection- Bacterial      | 0.14163          |                       | 0.14256           |                          |
| Opportunistic Infection- Fungal         | 0.26871          |                       | 0.16852           |                          |
| Urban                                   |                  |                       |                   |                          |
| Rural                                   | 0.18308          |                       |                   |                          |
| Government Employee                     |                  |                       |                   |                          |
| Non- Working                            | 0.35621          | 0.16578               | 0.1725            | 0.25879                  |
| Agricultural Labor                      | 0.1855           | 0.17885               | 0.1810            | 0.1829                   |
| Regular Employee                        | 0.21158          | 0.20152               | 0.22147           | 0.22346                  |
| Business Man                            | 0.32049          | 0.30557               | 0.32247           | 0.31862                  |
| Un Married                              |                  |                       |                   |                          |
| Married                                 | 0.54588          |                       |                   |                          |
| Spouse- Positive                        |                  |                       |                   |                          |
| Spouse- Negative                        | 0.14234          |                       |                   | 0.13596                  |

Table 8. Comparison of True and AIC models

| Model         | No. of predictors | AIC   |
|---------------|-------------------|-------|
| True          | 11                | 1349.4|
| Proposed method | 04                | 1313.2|
| Lasso method  | 06                | 1320.8|
| Net- Elastic  | 05                | 1325.60|

Conclusions

In this study we have tried to estimate the survival time of HIV/AIDS patients after reducing the number of predictors. 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 we tried to find the multiple/ associated predictors using MANOVA and Log-Linear models. Finally, only 4 predictors are selected. The
survival times are then again estimated using AFTM. Then the true model (with all possible predictors) and model fitted with proposed method (with reduced number of predictors) are compared with respect to their AIC values. It is found that proposed method has the model with the minimum AIC value. So, the model which has 4 independent predictors is a good fit as compared to the model which has 11 correlated/associated predictors. The proposed method was also compared with two different existing methods and it also gave better results than the existing methods. Therefore, from this study we can conclude that it is not necessary to take all the predictors which are effecting the survival time of the patients as many prognostic factors may be correlated/associated with each other and taking only one of them may satisfy the need of all of the remaining correlated/associated predictors.

Conflict of Interest: None

References

1. AIDS info Fact Sheet, http://aidsinfo.nih.gov/2012.
2. WHO (2013): Global update on HIV treatment 2013: Results, Impact and Opportunities.
3. UNAIDS (2014): a GAP report 2014.
4. Ghate M, Deshpande S, Tripathy S et al. Mortality in HIV infected individuals in Pune, India. *Ind J Med Res* 2011; 133(4): 414-420.
5. Rai S, Mahapatra B, Sircar 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. 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. DOI: 10.1186/1471-2334-9-128.
7. Jerene D, Endale A, Hailu Y et al. Predictors of early death in a cohort of Ethiopian patients treated with HAART. *BMC Infect Dis* 2006; 6: 136-136. DOI: 10.1186/1471-2334-6-136.
8. Erango MA, Goshu AT. Prediction of Survival of HIV/AIDS Patients from Various Sources of Data Using AFT Models. *Science Journal of Applied Mathematics and Statistics* 2017; 5(4): 127-133.
9. Chakravarty J, Tiwary NK, Prasad SR et al. Determinants of survival in adult HIV patients on antiretroviral therapy in Eastern Uttar Pradesh: A prospective study. *The Indian Journal of Medical Research* 2014; 140(4): 491-500.
10. Nawumbeni ND, 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.
11. Orisakwe EE, Ross AJ, Ocholla P.O. Correlation between knowledge of HIV, attitudes and perceptions of HIV and a willingness to test for HIV at a regional hospital in KwaZulu-Natal, South Africa. *African Journal of Primary Health Care and Family Medicine* (2012): 4(1), 1-8.
12. Nubed CK, Akoachere JFTK. Knowledge, attitudes and practices regarding HIV/AIDS among senior secondary school students in Fako Division, South West Region, Cameroon. *BMC public health* 2016; 16(1), 847.
13. Ye H, Yuan J, Wang Z et al. A Canonical Correlation Analysis of AIDS Restriction Genes and Metabolic Pathways Identifies Purine Metabolism as a Key Cooperator. *Computational and mathematical methods in medicine* 2016.
14. Grover G, Das R, Swain PK et al. On the estimation of survival of HIV/AIDS patients on antiretroviral therapy using NPMLE method: an application to interval censored data. *American Journal of Mathematics and Statistics* 2013; 3(4): 213-219.
15. Hafner R, Inderlied CB, Peterson DM et al. Correlation of quantitative bone marrow and blood cultures in AIDS patients with disseminated Mycobacterium avium complex infection. *The Journal of infectious diseases* 1999; 180(2): 438-447.
16. Rueda S, Mitra S, Chen S et al. Examining the associations between HIV-related stigma and health outcomes in people living with HIV/AIDS: a series of meta-analyses. *BMJ open* 2016; 6(7): e011453.