

Research Article

Comparing the Performance of Different High Dimensional Variable Selection Techniques on the Low Dimensional HIV/AIDS Data set

Anurag Sharma¹, Vajala Ravi², Vijay Kumar Sehgal³, Gurprit Grover⁴

^{1,2,4}Department of Statistics, University of Delhi, Delhi, India.

³Department of Mathematical Sciences & Computer Applications, Bundelkhand University, Jhansi, Uttar Pradesh, India.

DOI: <https://doi.org/10.24321/0019.5138.202003>

I N F O

Corresponding Author:

Anurag Sharma, Department of Statistics,
University of Delhi, Delhi, India.

E-mail Id:

anuragsharma532@gmail.com

Orcid Id:

<https://orcid.org/0000-0002-3482-0774>

How to cite this article:

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.

Date of Submission: 2020-03-17

Date of Acceptance: 2020-04-08

A B S T R A C T

Heavy censoring and high dimensionality have caused a great deal of difficulties for fitting and selection of model. This paper focuses on the performance of these four variable reduction techniques proposed by Khan and Shaw (2013) to select the variables to estimate the survival time of low dimensional HIV/ AIDS patients' data. The techniques used are adaptive elastic net, weighted elastic net, adaptive elastic net with censoring constraints and weighted elastic net with censoring constraints. The performance of these approaches is compared among themselves along with the full model (model with all the predictors). It is observed that Adaptive Elastic Net with Censoring Constraints performed best among all the methods. Moreover, these four techniques can also be used for future prediction of survival time under AFT model.

Keywords: Elastic Net, HIV, AIDS, AFT, Censoring

Introduction

Variable selection is of practical importance and well recognized in the survival analysis. Various techniques for variable selection have been discussed for linear regression models. Some of these methods are stepwise selection,¹ bridge regression,² Smoothly Clipped Absolute Deviation (SCAD),³ least absolute shrinkage and selection operator (LASSO),⁴ MM algorithms,⁵ which is based on extensions of the well-known class of EM algorithms, least angle regression selection (LARS),⁶ the Dantzig selector,⁷ the elastic net⁸ and penalized likelihood based techniques, such as Akaike's information criterion (AIC).⁹ Another variable selection procedure is Meinshausen N, Bühlmann P.¹⁰ This procedure is based on subsampling in combination with high-dimensional selection algorithms.

Also, many variable selection studies have been focused on the Cox model.¹¹ A regularized Cox regression was developed by Tibshirani R¹² which minimized a lasso penalty to the partial likelihood, a Bayesian variable selection method by Faraggi and Simon (1998),¹³ a non-concave penalized likelihood approach was developed by Fan J and Li R³, a threshold gradient descent regularization estimation method was introduced by Gui and Li (2005).¹⁴ A variable selection approach based on Dantzig selector for the Cox model based on the Dantzig selector was developed by Antoniadis A et al.¹⁵

For AFT models also, some variable selection techniques have been proposed. For e.g. a lasso regularization for estimation and variable selection in the AFT model was developed by Huang J and Harrington D,¹⁶ which was

based on the inverse probability of censoring. Huang J and Harrington D¹⁶ and Datta S et al.¹⁷ investigated the lasso regularized Buckley - James method for the AFT model. A Bayesian variable selection approach was developed by Sha N et al.¹⁸ Wang S et al.¹⁹ used elastic net for variable selection. A lasso regularization rank based estimator for variable selection was proposed by Engler D et al.²⁰ and Cai T et al.²¹ A bridge method was also used for variable selection by Huang J et al.¹⁶ A sparse penalization technique with censoring constraint was proposed by Hu S and Rao JS.²²

Khan and Shaw (2013a)²³ considered variable selection methods for the AFT models of high dimensional data sets. They also proposed new regularized Stute's Weighted Least Squares (SWLS) approaches. They introduced different classes of elastic net type regularized variable selection techniques which were based on SWLS. These classes were adaptive elastic net, a weighted elastic net and two extended versions which introduced censoring constraints into the optimization function.

In this paper, we have tried to apply the variable selection techniques proposed by Khan and Shaw (2013a)²³ on a low dimensional data set and have compared the results with other existing variable selection techniques. The rest of the paper is structured as follows: the introduction and regularized framework of SWLS, variable selection criteria and prediction formula given by the Khan and Shaw (2013b),²⁴ demonstration with a low-dimensional data set and the performance of the proposed approaches is compared with three other variable selection approaches, namely, elastic net which is implemented for weighted data, the adaptive elastic net approach²⁵ and a Bayesian approach.¹⁸

Methodology

Stute's Weighted least squares for an AFT model has an objective function:

$$\text{argmin} \left[\frac{1}{2} \sum_{i=1}^n w_i \{Y_{(i)} - \beta_0 - X_{(i)}' \beta\}^2 \right] \quad (1)$$

Where,

Y_i = log survival time for the i^{th} observation

$X(1) \leq X(2) \leq X(3) \dots \dots \dots X(n)$ are the associated covariates of $Y(1) \leq Y(2) \leq Y(3), \dots \dots \dots \leq Y(n)$ respectively.

β_0 = Intercept

β_0 = Unknown $p \times 1$ vector of true regression coefficients

w_i are jumps of the Kaplan-Meier estimator and are given as:

$$w_1 = \frac{\delta_{(1)}}{n} \quad (2)$$

$$w_i = \frac{\delta_{(i)}}{n - i + 1} \prod_{j=1}^{i-1} \left(\frac{n - j}{n - j + 1} \right)^{\delta_{(j)}} ; i = 2, \dots, \dots, n \quad (3)$$

w_i are also called Kaplan-Meier weights.

Stute's Weighted Least Squares method²⁶ is more docile to high dimensional covariates than a rank-based estimator²⁷ or a Buckley James estimator.²⁸ The reason being its OLS structure, a regularized method can easily be applied in the AFT model.

The objective function of SWLS can also be written in the matrix notation as:

$$\frac{1}{2} (Y - \alpha - X\beta)' w (Y - \alpha - X\beta) \quad (4)$$

where,

$w = n \times n$ diagonal weight matrix

For notational convenience, the intercept term β_0 can be removed by standardization (weighted) of the response and predictors. The weighted means are defined as:

$$\bar{X}_w = \frac{\sum_{i=1}^n w_i X_{(i)}}{\sum_{i=1}^n w_i}$$

and

$$\bar{Y}_w = \frac{\sum_{i=1}^n w_i Y_{(i)}}{\sum_{i=1}^n w_i}$$

Now, the adjusted covariates and responses can be calculated as:

$$X_{(i)}^w = \sqrt{w_i} (X_{(i)} - \bar{X}_w)$$

and

$$Y_{(i)}^w = \sqrt{w_i} (Y_{(i)} - \bar{Y}_w)$$

respectively.

Objective function of Regularized Weighted Least Squares has a general frame:

$$l(\beta) = \frac{1}{2} (Y - \alpha - X\beta)' (Y - \alpha - X\beta) + \frac{1}{2} \lambda \text{pen}(\beta) \quad (5)$$

where,

λ = Penalty Parameter (Scalar or Vector)

$\text{pen}(\beta)$ = Penalty quantity which is set in such a way that it controls the complexity of the model.

Proposed Model Framework

The Regularized WLS: Adaptive Elastic net (AEnet)

The elastic net⁸ is found to be useful when dealing with a large number of many correlated covariates. However, elastic net is not useful when applied with AFT model because of censoring. This problem has been resolved by the regularized Weighted Least Squares. For censored data, the naïve elastic net estimator $\hat{\beta}$ is given as:

$$\operatorname{argmin} \frac{1}{2} (Y_u - X_u \beta)' (Y_u - X_u \beta) + \lambda_1 \sum_{j=1}^p |\beta_j| + \frac{1}{2} \lambda_2 \beta' \beta \quad (1)$$

However, it is observed that this estimator doesn't have oracle properties which mean that this method doesn't select the correct subset of predictors which tends to one with probability. Therefore, Khan and Shaw (2013c)²⁹ presented a new approach named adaptive elastic net approach which is designed for censored data. They included the adaptive elastic net penalty terms to the regularized Weighted Least Squares objective function:

$$\operatorname{argmin} \frac{1}{2} (Y_u - X_u \beta)' (Y_u - X_u \beta) + \lambda_1 \sum_{j=1}^p \hat{w}_j |\beta_j| + \frac{1}{2} \lambda_2 \beta' \beta \quad (2)$$

where,

$\hat{w} = 1/|\hat{\beta}_0|^\gamma$ = Adaptive weight based on the initial estimator

Khan and Shaw (2013c)²⁹ took $\gamma = 1$ and the estimator given in (1) as the initial estimator $\hat{\beta}_0$. The algorithm for this approach referred to as the AEnet algorithm is given below:

Input: Design X^*_u , response Y^*_u , \hat{w} and a fixed set for λ_2 .

1. Define $X^{*i(u)} = X^*_{i(u)}/\hat{w}_i$ $i = 1, 2, \dots, p$
2. For a fixed λ_2 and all λ_1 , solve the lasso problem,

$$\beta^{**}_{a-nenet} = \operatorname{argmin} \frac{1}{2} (Y^*_u - X_u^{**} \beta)' (Y^*_u - X_u^{**} \beta) + \lambda_1 \sum_{i=1}^p |\beta_j|$$

3. Calculate $\hat{\beta}^*_{ia-net} = (1 + \lambda_2) \hat{\beta}^{**}_{ia-net} / \hat{w}_i$

To determine the optimal values for the tuning parameters (λ_1, λ_2) , typical values of λ_2 is taken in a relatively small grid, for e.g. (0, 0.5, 1, 1.5, , 5). Then, the entire solution is produced using the LARS algorithm for each λ_2 . Using this, the optimal equivalent specification for lasso is obtained in terms of the l_1 norm (t_1). In the last step, using the k -fold cross-validation, the optimal pair of (t_1, λ_2) is obtained.

The Regularized WLS: Adaptive Elastic Net with Censoring Constraints (AEnetCC)

In this technique, Khan and Shaw (2013b)²⁴ presented an extension of the adaptive elastic approach using which the censoring constraints can be implemented into the optimization framework. Adaptive elastic net estimator given by equation 2 can be written with censoring constraints as:

$$\operatorname{argmin} \frac{1}{2} (Y_u - X_u \beta)' (Y_u - X_u \beta) + \lambda_2 \xi' \xi + \frac{1}{2} \lambda_2 \beta' \beta$$

subject to

$$\sum_{j=1}^p \hat{w}_j |\beta_j| \leq t_1$$

and

$$Y_{\tilde{u}} \leq X_{\tilde{u}} \beta + \xi$$

where,

t_1 = lasso tuning parameter.

Khan and Shaw (2013c)²⁹ used quadratic programming approach to solve the minimization problem.

The algorithm for the adaptive elastic net with censoring constraints approach (referred as the AEnetCC algorithm) is given as:

Input:

1. Define $X^{*i(u)} = X^*_{i(u)}/\hat{w}_i$ $i = 1, 2, \dots, p$
2. Solve the elastic problem,

$$\beta^{**}_{a-nenet} = \operatorname{argmin} \frac{1}{2} |Y_u - X_u^{**} (\beta^+ - \beta^-)|' |Y_u - X_u^{**} (\beta^+ - \beta^-)| + \lambda_2 \beta^{+'} \beta^+ + \lambda_2 \beta^{-' } \beta^- + \lambda_0 \xi' \xi$$

Subject to

$$\sum_{i=1}^p (\beta_j^+ - \beta_j^-) \leq t_1$$

and

$$Y_{\tilde{u}} \leq X_{\tilde{u}} (\beta^+ - \beta^-) + \xi, \beta^+ \geq 0, \beta^- \geq 0$$

3. Calculate $\hat{\beta}^*_{ia-net} = (1 + \lambda_2) \hat{\beta}^{**}_{ia-net} / \hat{w}_i$

The Adaptive Elastic net with censoring constraint has three tuning parameters $(\lambda_1, \lambda_2, \lambda_3)$. The same optimal pair of (λ_1, λ_2) as found in Adaptive elastic net had been used by Khan and Shaw (2013a)²³. Then, typical values of λ_0 is taken in a relatively small grid, for e.g. (0, 0.5, 1, 1.5, , 5) and its optimal value was obtained by 5-fold cross-validation.

The value of λ_0 mainly depends on how stringently the user wants the model to satisfy the censoring constraints as compared to how good is the prediction for uncensored data.

The Regularized WLS: Weighted Elastic net (WEnet)

Khan and Shaw (2013b)²⁴ also presented a *weighted elastic net* for censored data. In this technique, for suitable weights w , the ridge penalty term is expressed as $\sum_{i=1}^p (w_i \beta_j)^2$, instead of $\sum_{i=1}^p \beta_j^2$. Therefore, it can also be regarded as a doubly adaptive type model.

Hong D and Zhang F²⁹ first studied this type of regularized technique for uncensored data. The model consistency and its oracle property were established by Hong D and Zhang F²² under regularity conditions. The weighted elastic net for censored data is given by:

$$\operatorname{argmin} \frac{1}{2} (Y_u - X_u \beta)' (Y_u - X_u \beta) + n_u \lambda_1 \sum_{i=1}^p w_i |\beta_j| + \frac{n_u}{2} \lambda_2 \sum_{i=1}^p (w_i \beta_j)^2$$

where, $w_i > 0, i = 1, \dots, p$.

These weights, w_i , are the standard deviations of the associated estimators.²² However, since the standard deviations are hardly known, the standard error of an initial

consistent estimator was used by Khan and Shaw (2013b).²⁴ To estimate the standard error under high dimensional data, they used a bootstrap procedure which is based on the elastic net model on the weighted data defined in (1). When $n > p$, Gehan type rank estimator is chosen as the initial estimator. This rank type estimator³⁰ is defined as the solution to the systematic equations,

$$0 = U_G(\beta),$$

where,

$$U_G(\beta) = n^{-1} \sum_{i=1}^n \sum_{i'=1}^n \delta_i(X_i - X_{i'}) I\{\xi_i(\beta) \leq \xi_{i'}(\beta)\},$$

and,

$$\xi_i(\beta) = Y_i - X_i' \beta$$

After solving these equations, Khan and Shaw (2013c)²⁹ obtained the weighted elastic net estimator as:

$$\hat{\beta}_{jw-net}^* = \hat{\beta}_{jw-net}^* / \sqrt{1 + \lambda_2}$$

which is a naïve elastic net estimator.

The original weighted elastic net estimator $\hat{\beta}_{jw-net}^*$ can be obtained by:

$$\sqrt{1 + \lambda_2} \hat{\beta}_{jw-net}^*$$

The algorithm for the proposed weighted elastic net approach is referred to as the WEnet algorithm.

Input: Design matrix X_u^* response Y_u^* , $\tilde{\lambda}$ and \hat{w}_i for $i = 1, 2, \dots, p$

1. Define,

$$X_{j(u)}^{**} = X_{j(u)}^{**} / \hat{w}_i, i = 1, 2, \dots, p.$$

$$\hat{\beta}_{w-net}^{**} = \arg \min (Y_u^* - X_u^{**} \beta)' (Y_u^* - X_u^{**} \beta) + \tilde{\lambda} \sum_{i=1}^p |\beta_j|.$$

$$3. \hat{\beta}_{w-net}^* = \hat{\beta}_{w-net}^{**} / \hat{w}_i$$

Output: $\hat{\beta}_{w-net}^* = \sqrt{1 + \lambda_2} \hat{\beta}_{w-net}^{**}$

The optimal value for the tuning parameter $(\lambda_1, \lambda_2), \lambda_2$ is typically assumed to be in a relatively small grid which is similar to the grid, they have used for AEnet algorithm. For optimizing the tuning parameters (λ_1, λ_2) , exactly the same procedure as of AEnet was followed by Khan and Shaw (2013a).²³

The Regularized WLS: Weighted Elastic Net with Censoring Constraints (WEnetCC)

The weighted elastic net model with censoring constraints can be defined on the same lines as on the adaptive elastic net model with censoring constraints:

$$\tilde{\beta}_{w-enet}^* = \arg \min \frac{1}{2} [Y_u - X_u]' [Y_u - X_u] + \lambda_2 W \beta' \beta + \lambda_0 \xi' \xi$$

Subject to

$$\sum_{i=1}^p \hat{w}_i |\beta_j| \leq t_1$$

and

$$Y_{ii} \leq X_{ii} \beta + \xi$$

The algorithm for this approach (referred to as WEnetCC algorithm) is given as:

Input: Design matrix X_u^* response Y_u^* , $\tilde{\lambda}$ and $\gamma_u^*, \tilde{\lambda}$ and \hat{w}_i for $i = 1, 2, \dots, p$

1. Define,

$$X_{j(u)}^{**} = X_{j(u)}^{**} / \hat{w}_i, i = 1, 2, \dots, p.$$

2. Solve the lasso problem,

$$\hat{\beta}_{-net}^{**} = \arg \min [Y_u - X_u^{**} (\beta^+ - \beta^-)]' [Y_u - X_u^{**} (\beta^+ - \beta^-)] + \lambda_0 \xi' \xi$$

Subject to

$$\sum_{i=1}^p (\beta_j^+ - \beta_j^-) \leq t_1$$

and

$$Y_{ii} \leq X_{ii} (\beta^+ - \beta^-) + \xi, \beta^+ \geq 0, \beta^- \geq 0$$

3. Calculate $\hat{\beta}_{w-enet}^* = (1 + \lambda_2) \hat{\beta}_{w-enet}^{**} / \hat{w}_i$

Output: $\hat{\beta}_{w-enet}^* = \sqrt{1 + \lambda_2} \hat{\beta}_{w-enet}^{**}$

By optimizing the QP problem, the estimator in the second step is obtained. Tuning parameters $(\lambda_0, \lambda_1, \lambda_2)$ are obtained by following exactly the same procedure as for the AEnetCC algorithm.

Variable Selection Criteria

Khan and Shaw (2013a)²³ Least regression angle selection (LARS) for Adaptive Elastic Net and Weighted Elastic net. This provides exact zero coefficients in solution paths and therefore, it does parsimonious selection. For two methods with censoring constraints, they used an approach which removes several variables simultaneously which have very small coefficients.

An AIC type score which is based on the weighted k -fold cross-validation error (CV-S). CV-S is calculated as the sum of squared residuals of uncensored data multiplied by the K-M weights, i.e. $(Y_u - X_u \hat{\beta})' w_u (Y_u - X_u \hat{\beta})$. AIC score is calculated as:

$$AIC = n_u \log(CV - S) + 2k \left(\frac{n_u}{n_u - k - 1} \right)$$

Variable Selection Algorithm for AEnetCC and WEnetCC

1. By fitting the AEnetCC or WEnetCC, get the optimal pair (λ_1, λ_2) .

2. A set of λ_0 is fixed. Then for each λ_0 :
 - (a) Using the computational procedure of AEnetCC and WEnetCC, fit the AFT model. Then the predictor set is found by using $|\hat{\beta} > \zeta|$.
 - (b) Using the predictor set, the dataset is divided into k parts. Then one part is left out at a time and AFT is fitted by the procedure AEnetCC or WEnetCC.
 - (c) Then the k - fitted models built in second step are combined by averaging their coefficients.
 - (d) CV-S and AIC score are computed.
3. Step 2 is repeated until all λ_0 are exhausted. Then the model with lowest AIC score and corresponding λ_0 are returned.

$\zeta = 1e^{-5}$ is set as the default value for the precision parameter. However, any other suitable value can also be chosen. ζ may be considered as an additional tuning parameter in the variable selection algorithm.

Measures of Fit and Measures of Prediction

$$MSE_{TR} = \frac{1}{n_u} \sum_{i=1}^n \delta_i (\hat{Y}_i - Y_i)^2,$$

where, n_u = number of uncensored observations

This compares the true values with the fitted values corresponding to the uncensored observations. Firstly, a training dataset is generated and then a test dataset Y^{new} (say) of the same size is generated using the same design parameters. Using the training data, all these models are fitted. After that, in order to get the predicted values \hat{Y}_{new} , the X matrix of the test data is used with the fitted model. The prediction accuracy is measured by:

$$MSE_{TE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_{new,i} - Y_{new,i})^2.$$

Khan and Shaw (2013c)²⁹ estimated the variance of regression parameters by using the non-parametric 0.632 bootstrap.

Low-Dimensional Data Example

HIV data: The primary goal of our study is to estimate the survival time of HIV/AIDS patients by selecting the number of predictors using the above variable selection technique proposed by Shaw and Khan (2013). HIV is one of the largest epidemics in the world. Talking, about India only, it is the India's third largest epidemic in the world. For the adults aged 15-49, prevalence was estimated to be 0.2% (UNAIDS data 2018).³¹ In comparison to other middle-income countries, this figure is relatively small. However, taking into account the huge population of India, this figure equates to 2.1 million people.

However, this epidemic is slowing down in India, largely because of the National AIDS Control Programme which has made people aware about this epidemic. Decrease of 27% in new infections has been noticed during 2010 to 2017. Also, AIDS related deaths have fallen down by more than 50% during this period. Since the beginning of 2001, number of new infections has also lowered by more than half. However, in 2017, new infections of HIV were noted to be increased from 80,000 to 88,000 and deaths were found to be 69,000 as compared to 62,000 in 2016 (UNAIDS 2017).³² Manipur, Mizoram and Nagaland are the three Indian states with the highest HIV prevalence.

In 2017, more than 75% of the people living with HIV were aware of their HIV status. HIV positive men are less significantly likely to be diagnosed as compared to HIV positive women (68% to 87%). The main reason behind this is the number of women testing for HIV through preventing mother to child transmission (PMTCT) services. Government of India has also done some appreciable work to lessen the impact of HIV in India. As compared to 67 HIV testing and counseling (HTC) sites 1997, 23,400 facilities centers were opened by 2017. 18.6 million patients used these services which surpassed the India's annual testing target of 14 million between April 2016 and April 2017.

Our data was obtained from a retrospective follow- up study which involved 767 patients who underwent Antiretroviral therapy in the RML hospital, New Delhi from January 2004 to December 2014. The survival time of the patients is taken as the time from the date of registration till the date of last contact. Our data had 11 predictors, namely, age, sex, state, smoking, alcohol, drugs, spouse HIV status, occupation, opportunistic infections, living status and marital status.

We have applied the AFT model with all the methods proposed by Khan and Shaw (2013a).²² All the approaches are employed and the optimal tuning parameters are selected with 5- fold cross- validation. The results are reported in table 2. Among the four proposed methods, AEnetCC selects the lowest number of genes.⁴ The AEnet, WEnet, WEnetCC select 5, 6 and 5 variables respectively.

Table 1. Number of variables selected by different methods

Methods	No. of variables selected
AEnet	6
AEnetCC	4
WEnet	5
WEnetCC	5

In the next step, AFT model is fitted with all the predictors and then by taking the predictors only chosen by the discussed methods one by one. AIC value for each model is then compared to determine which method fits best.

In table 3, the AIC values of different models fitted with variables selected by the discussed methods are shown.

Table 2. Variables selected by different methods

Methods	Variables
AEnet	Age, drugs, sex, opportunistic infections, smoking, alcohol
AEnetCC	Age, smoking, sex, occupation
WEnet	Age, drugs, sex, alcohol, opportunistic infections
WEnetCC	Age, drugs, living status, smoking, sex

The standard errors of the coefficients of the predictors chosen by the 4 methods discussed above are compared among themselves and also standard errors of the model along with the coefficients when all the predictors are

taken to estimate the survival time. The results are shown in Table 4.

Table 3. AIC values of different models fitted with variables selected by the discussed methods

Methods	No. of Predictors	AIC
True (All Predictors)	11	1349.4
AEnet	6	1296.4
AEnetCC	4	1293.8
WEnet	5	1297.6
WEnetCC	5	1297.8

From the above table, it can be observed that the standard errors of the coefficients of the predictor chosen by the Adaptive Elastic Net with Censoring Constraints are least among all the techniques discussed.

Table 4. Standard errors of the coefficients of predictors

Predictors	S.E (True Model)	S.E (AEnet)	S.E (AEnetCC)	S.E (Wenet)	S.E (WenetCC)
Age	0.00804	0.00758	0.00741	0.0075	0.00748
Male					
Female	0.34583	0.30164	0.30001	0.30128	0.30095
Eunuch	0	0.00152	0.00137	0.00144	0.00141
North India					
Central India	0				
East India	0.77448				
Smoking - Yes					
Smoking - No	0.22841	0.19436	0.19264		0.19307
Alcohol - Yes					
Alcohol - No	0.16252	0.12294		0.12086	
Drugs - Never					
Drugs - Past	0.18161	0.19248		0.19631	0.19205
Drugs - Yes	1.0675	1.0174		1.0267	1.0188
Opportunistic Infection - Viral					
Opportunistic Infection - Bacterial	0.14163	0.12405		0.12196	
Opportunistic Infection - Fungal	0.26871	0.25982		0.25875	
Urban					
Rural	0.18308				0.17824
Government Employee					
Non-working	0.35621		0.33577		
Agricultural Labor	0.1855		0.1821		
Regular Employee	0.21158		0.20135		
Business Man	0.32049		0.30123		
Un-married					

Married	0.54588				
Spouse - Positive					
Spouse - Negative	0.14234				

Conclusion

In this study, the four variable selection techniques for the AFT model have been applied to a low-dimensional data set of HIV/AIDS. We have the data of 767 patients diagnosed with HIV/AIDS with 11 prognostic factors. Then, four techniques have been applied to reduce the number of predictors. The four techniques namely, AEnet, AEnetCC, WEnet and WEnetCC select 6, 4, 5, and 5 predictors respectively. Then the AFT model is fitted with the selected number of predictors and with all the predictors and compared using the AIC value. AIC value for all the models fitted by these techniques is least as compared to the AIC value of the model fitted with all the predictors. Also, AEnetCC is found to have the least AIC value. So, in this paper the analysis of low-dimensional HIV/AIDS shows that these approaches can be used for selecting important variables and they can also be used for future prediction of survival time under AFT models.

Conflict of Interest: None

References

- Peduzzi PN, Hardy RJ, Holford TR. A stepwise variable selection procedure for nonlinear regression models. *Biometrics* 1980; 36: 511-6.
- Huang J, Ma S. Variable selection in the accelerated failure time model via the bridge method. *Lifetime Data Analysis* 2010; 16: 176-95.
- Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J Amer Stat Assoc* 2001; 96: 1348-1360.
- Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Statist Soc Ser B* 1996; 58: 267-288.
- Hunter DR, Li R. Variable selection using MM algorithms. *The Annals of Statistics* 2005; 33: 1617-1642.
- Efron B, Hastie T, Johnstone I et al. Least angle regression. *Ann Statist* 2004; 32: 407-499.
- Candes E, Tao T. The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics* 2007; 35: 2313-2351.
- Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Roy Statist Soc Ser B* 2005; 67: 301-320.
- Akaike H. Information theory as an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (Petrov BN, Csaki F, eds.). Akademiai Kiado, Budapest. 1973, 267-281.
- Meinshausen N, Bühlmann P. Stability selection. *J Roy Statist Soc Ser B* 2010; 72: 417-473.
- Cox DR. Regression models and life-tables. *J Roy Statist Soc Ser B* 1972; 34: 187-220.
- Tibshirani R. The lasso method for variable selection in the Cox model. *Statistics in Medicine* 1997; 16: 385-395.
- Faraggi D, Simon R. Bayesian Variable selection method for censored survival data. *Biometrics* 1998; 54: 1475-1485.
- Gui J, Li H. Penalized Cox regression analysis in the high dimensional and low-sample size settings, with applications to microarray gene expression data. *Bioinformatics* 2005; 21: 3001-3008.
- Antoniadis A, Fryzlewicz P, Letue F. The Dantzig selector in cox's proportional hazards model. *Scandinavian Journal of Statistics* 2010; 37: 531-552.
- Huang J, Harrington D. Iterative partial least squares with right censored data analysis: A comparison to other dimension reduction techniques. *Biometrics* 2005; 61: 17-24.
- Datta S, Le-Rademacher J, Datta S. Predicting patient survival from microarray data by accelerated failure time modeling using partial least squares and LASSO. *Biometrics* 2007; 63: 259-271.
- Sha N, Tadesse MG, Vannucci M. Bayesian variable selection for the analysis of microarray data with censored outcome. *Bioinformatics* 2006; 22: 2262-2268.
- Wang S, Nan B, Zhu J et al. Doubly penalized Buckley James method for survival data with high-dimensional covariates. *Biometrics* 2008; 64: 132-140.
- Engler D, Li Y. Survival Analysis with High-Dimensional Covariates: An Application in Microarray Studies. *Statistical Applications in Genetics and Molecular Biology* 8 Article 14. 2009.
- Cai T, Huang J, Tian L. Regularized estimation for the accelerated failure time model. *Biometrics* 2006; 65: 394-404.
- Hu S, Rao JS. Sparse penalization with censoring constraints for estimating high dimensional AFT models with applications to microarray data analysis. Technical Reports, University of Miami. 2010.
- Khan MHR, Shaw JEH. AdapEnetClass: A class of adaptive elastic net methods for censored data R package version 1.0. 2013a.
- Khan MHR, Shaw JEH. On Dealing with Censored Largest Observations under Weighted Least Squares. CRiSM Working Paper, No. 13-07 Department of Statistics, University of Warwick. 2013b.
- Zou H, Zhang HH. On the adaptive elastic-net with a

- diverging number of parameters. *Annals of Statistics* 2009; 37: 1733-1751.
26. Stute W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics* 1996; 23: 461-471.
 27. Ying Z. A large sample study of rank estimation for censored regression data. *Annals of Statistics* 1993; 21: 76-99.
 28. Buckley J, James I. Linear regression with censored data. *Biometrika* 1979; 66: 429-436.
 29. Khan MHR, Shaw JEH. Variable Selection with The Modified Buckley-James Method and The Dantzig Selector for High-dimensional Survival Data. 2013c.
 30. Jin Z, Lin D, Wei LJ et al. Rank-based inference for the accelerated failure time model. *Biometrika* 2003; 90: 341-353.
 31. UNAIDS data (2018): a GAP data report 2018.
 32. UNAIDS (2017): a GAP report 2017.