



## Original Article

# Comparison of Penalized Cox Regression Methods in Low-Dimensional Data with Few-Events: An Application to Dialysis Patients' Data

Shideh Rafati (MSc)<sup>1</sup>, Mohammad Reza Baneshi (PhD)<sup>1,2</sup>, Laleh Hassani (PhD)<sup>3</sup>, and Abbas Bahrapour (PhD)<sup>1,2\*</sup>

<sup>1</sup> Department of Biostatistics and Epidemiology, Kerman University of Medical Sciences, Kerman, Iran

<sup>2</sup> Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

<sup>3</sup> Mother and Child Welfare Research Center, Hormozgan University of Medical Sciences, Bandar Abbas, Iran.

## ARTICLE INFORMATION

### Article history:

Received: 25 April 2019

Revised: 09 June 2019

Accepted: 08 July 2019

Available online: 15 July 2019

### Keywords:

Chronic renal failure

Dialysis

Survival

Cox models

### \* Correspondence:

Abbas Bahrapour (PhD)

Tel: +98 9131404512

Fax: +98 3431325127

E-mail: [abahrapour@yahoo.com](mailto:abahrapour@yahoo.com)

## ABSTRACT

**Background:** Dialysis is a dominant therapeutic method in patients with chronic renal failure. The ratio of those who experienced the event to the predictor variables is expressed as event per variable (EPV). When EPV is low, one of the common techniques which may help to manage the problem is penalized Cox regression model (PCRM). The aim of this study was to determine the survival of dialysis patients using the PCRM in low-dimensional data with few events.

**Study design:** A cross-sectional study.

**Methods:** Information of 252 dialysis patients of Bandar Abbas hospitals, southern Iran, from 2010-16 were used. To deal with few mortality cases in the sample, the PCRM (lasso, ridge and elastic net, adaptive lasso) were applied. Models were compared in terms of calibration and discrimination.

**Results:** Thirty-five (13.9%) mortality cases were observed. Dialysis data simulations revealed that the lasso had higher prediction accuracy than other models. For one unit of increase in the level of education, the risk of mortality was reduced by 0.32 (HR=0.68). The risk of mortality was 0.26 (HR=1.26) higher for the unemployed than the employed cases. Other significant factors were the duration of each dialysis session, number of dialysis sessions per week and age of dialysis onset (HR=0.93, 0.95 and 1.33).

**Conclusion:** The performance of penalized models, especially the lasso, was satisfying in low-dimensional data with low EPV based on dialysis data simulation and real data, therefore these models are the good choice for managing of this type of data.

**Citation:** Rafati S, Baneshi MR, Hassani L, Bahrapour A. Comparison of Penalized Cox Regression Methods in Low-Dimensional Data with Few-Events: An Application to Dialysis Patients' Data. *J Res Health Sci.* 2019; 19(3): e00452.

© 2019 The Author(s); Published by Hamadan University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

The full defect and irreversible reduction in renal function lasting more than three months are called chronic renal failure and its advanced stage, in which survival depends on transplant or dialysis, entitled the end-stage renal disease (ESRD). Chronic renal failure is a disease with an increasing trend in recent years; in the US, the prevalence and incidence of ESRD have been doubled in the past decade<sup>1</sup>. In Iran, the prevalence and incidence of ESRD have increased from 238 in 1,000,000 in 2000 to 357 in 1,000,000 in 2006<sup>2</sup>. Moreover, 48.5% of the patients with ESRD used transplant, 48.5% hemodialysis and 3% peritoneal dialysis. In 2008, approximately 12500 patients with ESRD used hemodialysis in Iran<sup>3</sup>. Based on the statistics of the Center for Transplant and Special Disease Management of the Ministry of Health and Medical Education, the number of registered ESRD patients was 24,000 and 40,000 in 2004 and 2009, respectively<sup>4</sup>. Dialysis is a process for eliminating extra uremic fluids and products from the body, used when the kidneys fail to do so. Of the common kidney replacement therapies, hemodialysis is the prevalent method for many patients with ESRD<sup>5</sup>.

This disease can have a negative effect on the quality of life of patients due to its chronic and debilitating nature and leads to reduced social interactions, depression, frustration, reducing a person's ability to perform independent daily activities of life and ultimately increasing mortality. With regard to the issues and problems that dialysis patients are facing, important to determine the variables that affect survival of these patients.

Usually, the data sets comprised enough sample size ( $n$ ) and limited number of independent variables ( $p$ ), that is  $n > p$ , called low dimensional data and in classic statistical procedures the Cox regression is the common applicable method in such a data sets. In contrast, high dimensional data refers to the situations where  $n < p$ .

Another issue affects performance of regression models is the number of Events (known as effective sample size) Per number of independent Variables (EPV)<sup>6</sup>. The ratio of those experiencing the noted event to the number of predictor variables or, more precisely, the number of parameters, is expressed as EPV and based on the simulated studies, EPV of

10 to 20 has been recommended. When EPV is low, the coefficients of Cox model are not reliable.

In high dimensional data, EPV is always low. Therefore, penalized Cox regression methods (PCRM), including ridge regression<sup>7</sup>, lasso<sup>8</sup>, elastic net<sup>9</sup> and adaptive lasso<sup>10</sup> that shrink some regression coefficients towards zero are applicable<sup>11</sup>.

Even in low dimensional data, EPV might be low. For example, assume a data set when  $n=150$ ,  $p=15$ ,  $n_{\text{event}}=30$ . Here,  $n > p$  and  $\text{EPV}=2$ . Even in the case of low dimensional data with low EPV, standard Cox model might not be applicable<sup>12</sup>. Few studies have used PCRM for low-dimensional data with few events<sup>13,14</sup>.

In this study, we aimed to determine the variables that affect the survival of dialysis patients, especially those added to survival of patients by controlling them, including these variables can be mentioned to treatment duration and number of hemodialysis sessions per week.

## Methods

One function of the models is to predict the risk of a future event. Based on the low number of mortality cases (the outcome of interest) in the sample, standard Cox regression models were not appropriate since the estimated regression coefficients might become invalid and the predictive models might have weak reliability. A strategy used in this situation is the penalized regression method (lasso, ridge, elastic net and adaptive lasso), which can help in the cases where the value of EPV is low.

By maximizing the penalized partial log-likelihood function below, the elastic net regression coefficients are found.

$$\frac{2}{n}l(\beta) - \lambda(\alpha \sum_{j=1}^p \widehat{w}_j |\beta_j| + \frac{1}{2}(1 - \alpha) \sum_{j=1}^p \beta_j^2)$$

Where  $\beta_1, \beta_2, \dots, \beta_p$  correspond to regression coefficients, which are in fact, the weight given to each variable by the model.  $W$  is called known weights vector, which is a vector contains one for all of coefficients the lasso, the ridge, and the elastic net.  $l(\beta)$  is the partial log-likelihood function for the Cox model and  $\lambda > 0$  refers to as the tuning parameter that  $\lambda$  were selected using 10-fold cross-validation. Larger values of  $\lambda$  lead to smaller regression coefficients. The  $\alpha$  parameter varying from zero to one in the elastic net is called a hybrid parameter. The coefficients of the lasso model are obtained based on  $\alpha=1$  and for the ridge model,  $\alpha=0$ <sup>12</sup>.

Adaptive lasso is a version of the lasso, where it assigns different weights to different coefficients for penalizing the coefficients in the lasso. The adaptive lasso can have the oracle properties; namely, it can identify the right subset of true variables and can have optimal estimation rate, if the weights cleverly are chosen. For this method,  $\widehat{w}_j = \frac{1}{|\widehat{\beta}_j^*|}$  where  $\widehat{\beta}^*$  is an initial estimate of the coefficients and usually obtained through the ridge.

Here, the application of the penalized methods is used to a real data example with a low EPV based on the dialysis data. The data were collected through the Dialysis Ward of Bandar Abbas Hospitals, Iran. This project has been approved by the

Ethics Committee of Kerman University of Medical Sciences with No. IR.KMU.REC.1397.599.

Of the total patients admitted from 2010 to 2016, the data of 252 patients were recorded in the dialysis ward. Mortality was considered as the event of interest and censored cases included those who were alive at the end of the study, excluded cases and those treated with kidney transplant. The survival time of the patients was calculated by years from the onset of dialysis to the end of the study in 2016.

The data were collected based on a designed checklist including age, sex, education, job, blood type, marital status, smoking, disease leading to dialysis (diabetes, hypertension, renal stones and obstruction, renal cysts and congenital diseases), dialysis duration(hour per session), number of dialysis sessions per week, history of cardiac-respiratory diseases, history of anemia and familial history of chronic renal failure.

Regarding the low frequency in some categories of independent variables, in the end, 17 variables included sex, job (five indicator variables), blood type (three indicator variables), history of smoking, diabetes and hypertension (all binary), education, dialysis duration (hour per session), number of dialysis sessions per week, body mass index (BMI), age of diagnosis (all continuous) are used in penalized models. For this data, there were 17 regression coefficients and 35 events so the EPV was 2.

Data of 252 patients were randomly divided into training and testing set, and this process was repeated 500 times. Penalized models were used for the training set. The mean of concordance index (C-index), calibration slope (CS) and the root mean square error (RMSE) were used to assess and compare the prediction accuracy of the penalized models in the test set.

The value of 0.5 for C-index shows the inability of the model in differentiating patients and one indicates the full ability of the model in this differentiation<sup>13,15</sup>. The best RMSE has a value close to zero and is calculated as follows.

$$RMSE(t) = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i(t) - \widehat{S}_i(t))^2}$$

Where  $S_i(t)$  and  $\widehat{S}_i(t)$  are true survival probabilities of  $i$  th individual at time  $t$  and survival probabilities estimated by the different models, respectively<sup>12,16</sup>. CS is the slope obtained by fitting a simple linear regression model to  $y = \log(\frac{1-S_i(t)}{S_i(t)})$  and  $x = \log(\frac{1-\widehat{S}_i(t)}{\widehat{S}_i(t)})$ . The ideal value for the CS is one<sup>12</sup>. CS is related to goodness of fit, which relates to the ability of a model to fit a given set of data<sup>17</sup>.

Moreover, to compare the predictive performance of the models using the dataset of dialysis patients, we simulated 500 survival datasets with a sample size of 252 for two EPV scenarios (2 and 5). We generated survival times regarding Cox proportional hazard model and the exponential distribution applied to generate baseline hazard in cox model.

The predictors were independent of each other and the vector of true coefficients was  $\beta = (1.06, -0.298, 0.362, -1.249, 1.195, -0.283, 0.385, -0.397, 0.04, -12.72, -0.21, 0.579, -0.242, -0.253, 0.549, -0.601, 0.099)$ . All continuous predictors were

generated from normal distribution and all categorical variables were produced from binomial distribution.

All the analyses and simulations were carried out using software R version 3.5.3. The glmnet package was used for fitting PCRM (lasso, ridge, elastic net and adaptive lasso). The predicted survival probabilities and C-index were calculated with c060 and Hmisc packages, respectively.

## Results

Overall, 252 hemodialysis patients were studied, of them, 35 (13.9%) cases faced the event of death and 217 (86.1%) cases were censored. The median follow-up was 10 years. The 10 and 20-year survival rate of these patients were 0.86 and 0.69, respectively.

Over 80% of the patients were illiterate or low literacy. Most of the women were housekeeper (87.1%) and most of the men were unemployed or retired (55.9%). About 64% of the patients did not smoke any form of tobacco. All the patients, except for one, had at least one disease leading to dialysis. No case of infection with HIV was observed in the patients. For 194 (77%) patients, each session of dialysis took 4 hours. Moreover, 171(67.9%) patients used dialysis three times per week. Table 1 has described demographic, clinical, and laboratory characteristics of patients in baseline.

**Table 1:** Patients Characteristics

Continuous variables	Censored		Died	
	Mean	SD	Mean	SD
Age (yr)	53.24	17.32	53.39	18.09
Age starting dialysis	42.89	16.71	42.83	19.40
Body Mass Index	23.11	4.28	21.44	3.68
Categorical variables	Number	Percent	Number	Percent
Blood group O	92	42.4	16	45.7
Blood group A	57	26.3	5	14.3
Blood group B	58	26.7	14	40.0
Blood group AB	10	4.6	0	0.0
Illiterate	70	32.3	16	45.7
Low literacy	102	47.0	15	42.9
Diploma	35	16.1	2	5.7
Collegiate	10	4.6	2	5.7
Males	117	53.9	19	54.3
Married	179	82.5	28	80.0
Tobacco use	76	35.0	15	42.9
Diabetes	111	51.2	23	65.7
Hypertension	134	61.8	18	51.4
Urinary stones and Kidney obstruction	22	10.1	1	2.9
Renal cysts	11	5.1	0	0.0
Pulmonary heart disease	45	20.7	5	14.3
Congenital disease	4	1.8	0	0.0
Glomerulonephritis	16	7.4	2	5.7
History of CRF in the family	22	10.1	2	5.7
Anemia	165	76.0	30	85.7
Receive of erythropoietin	204	94.0	35	100
HCV	8	3.7	0	0.0
HBV	3	1.4	0	0.0

In Table 2, we showed the internal and external performance of the methods using cross-validation. The ridge followed by the lasso and adaptive lasso provided the best discrimination in both train and test set. In terms of calibration, the lasso is the best in train and test set (1.224 and 1.332, respectively).

**Table 2:** assessing the prediction accuracy of the models in ten-year based on real dialysis data (500 iterations)

Model	C-index	SE	RMSE <sup>a</sup>	SE	CS <sup>b</sup>	SE
<b>Train set</b>						
Cox-lasso	0.734	0.046	0.133	0.021	1.224	0.655
Cox-ridge	0.760	0.049	0.135	0.016	1.958	0.391
Cox-elastic net	0.730	0.053	0.135	0.007	1.826	0.261
Cox-adaptive lasso	0.732	0.052	0.088	0.013	1.547	0.436
<b>Test set</b>						
Cox-lasso	0.737	0.069	0.158	0.013	1.332	0.286
Cox-ridge	0.762	0.062	0.159	0.018	1.710	0.426
Cox-elastic net	0.731	0.076	0.158	0.014	1.390	0.275
Cox-adaptive lasso	0.734	0.073	0.099	0.020	1.412	0.383

<sup>a</sup> Root Mean Square Error

<sup>b</sup> Calibration Slope

For train set, the best RMSE was attained using the adaptive lasso (0.088), followed by the lasso (0.133), while the elastic net and the ridge had a RMSE of 0.135. The best RMSE was achieved in test set by the adaptive lasso (0.099), followed by the lasso and the elastic net (0.158).

In Table 2, after assessing the performance of the methods based on C-index, RMSE, and CS, these methods have acceptable predictive accuracy in train and test set. No one method dominates the others, and they all seem to have specific strengths.

Table 3 shows the results of assesses and compares the prediction accuracy of PCRM based on mean of C-index, RMSE, and CS of simulated datasets with different EPV. For EPV 2 in Table 3, the highest mean of C-index was attained by the lasso and the ridge (0.672), the lasso and the adaptive lasso performed the best for EPV 5 (0.627).

**Table 3:** comparison of the prediction precision of the models in ten-year based on simulated dialysis data

Model	C-index	SE	RMSE <sup>a</sup>	SE	CS <sup>b</sup>	SE
<b>EPV=2</b>						
Cox-lasso	0.672	0.074	0.075	0.012	1.673	0.852
Cox-ridge	0.672	0.085	0.082	0.018	1.858	0.449
Cox-elastic net	0.669	0.063	0.080	0.016	1.724	0.516
Cox-adaptive lasso	0.670	0.037	0.082	0.018	1.845	0.478
<b>EPV=5</b>						
Cox-lasso	0.627	0.042	0.075	0.016	1.307	0.186
Cox-ridge	0.626	0.044	0.102	0.019	1.710	0.426
Cox-elastic net	0.626	0.044	0.100	0.017	1.429	0.217
Cox-adaptive lasso	0.627	0.047	0.102	0.018	1.488	0.338

<sup>a</sup> Root Mean Square Error

<sup>b</sup> Calibration Slope

Findings in Table 3 show that the lasso had the highest prediction precision based on RMSE for both EPVs (0.075). For both EPVs, the lasso had the highest prediction precision of other three models based on the CS (1.673 and 1.307, respectively). Generally, the performance of the lasso in the simulations was superior to that of the ridge, the elastic net and the adaptive Lasso.

Findings of the lasso (Table 4) indicate that BMI, the level of education, occupation, dialysis duration in each session, number of dialysis sessions per week and age of dialysis onset were the most important variables in predicting the survival time of patients in this study.

Based on this table, for one unit of increase in BMI and the level of education, the risk of mortality is reduced by 0.11 and

0.32, respectively. The risk of mortality was 0.26 higher for unemployed patients than the employed ones. Another significant factor was the duration of each dialysis session. For one unit of increase in the duration of dialysis session, the risk of mortality was shrunk by 0.07. Moreover, by controlling other factors, increasing one unit in the number of dialysis session per week reduce the risk of death by 0.05. Findings revealed that for one unit of increase in age of dialysis onset, the risk of mortality was increased 0.33 by adjusting the effect of other variables.

**Table 4:** the most important variables based on lasso

Variables	Coefficient	95% CI	Hazard Ratio
BMI	-0.107	(-0.182, -0.019)	0.89
Education	-0.376	(-0.424, -0.319)	0.68
Unemployed occupation <sup>a</sup>	0.234	(0.194, 0.264)	1.26
Dialysis Duration	-0.068	(-0.111, -0.021)	0.93
Number of Dialysis	-0.051	(-0.078, -0.019)	0.95
Age of dialysis onset	0.292	(0.230, 0.350)	1.33

<sup>a</sup> Reference is employed

Table 5 shows the most important variables based on the lasso, the elastic net and the adaptive lasso in 500 bootstrap samples drawn from the original dataset.

**Table 5:** The most important variables (500 bootstrap samples)

Variables	Lasso	Elastic net	Adaptive lasso
Body mass index	✓	✓	✓
Occupation	✓	✓	✓
Education	✓	✓	✓
Dialysis duration (hour)	✓	✓	✓
Number of dialyses (per week)	✓	✓	✓
Age of dialysis onset	✓	✓	✓
Blood group			✓

## Discussion

We aimed to determine the survival of dialysis patients using these methods (lasso, ridge, elastic net and adaptive lasso) for low-dimensional data with few events. The application of penalized methods in low-dimensional data with low EPV, although important, has been used less.

Upon assessing and comparing the performance of PCRM (lasso, ridge, elastic net and adaptive lasso) based on RMSE, C-index and CS in simulation data with two low EPV scenarios (EPV 2 or 5), lasso was the better model than the others. This means the lasso had higher prediction precision than the other three models which is consistent the other study. In mentioned study, lasso Cox regression model was used to determine variable selection and constructed a model for predicting mortality in dialysis patients which good accuracy of lasso was confirmed by C-index and CS<sup>18</sup>.

A penalized approach should certainly be considered when survival models are used in low-dimensional, low EPV settings. Particularly, use of either the ridge or the lasso is recommended in situations where the EPV is lower than five<sup>13</sup>.

It is important to avoid classic methods for low dimensional data with few events. The penalized methods can improve calibration and predictive accuracy<sup>14</sup> that present study confirms previous studies.

Ridge shrunk regression coefficients close to zero, but no coefficient was exactly zero. Therefore, the variable selection was not performed. In fact, this method is useful for resolving the problem of multicollinearity and model prediction<sup>14</sup>. The

prediction accuracy of the lasso was higher than the elastic net and the adaptive lasso based simulation data (Table 3). Besides, the lasso is a popular technique for simultaneous estimation and variable selection. Therefore, we focus on and interpret the most important variables in predicting the survival of patients based on the lasso model.

In this study, the significance of the variables of BMI, education level, occupation and dialysis duration (hours) in each session and the number of dialysis sessions per week, age of dialysis onset were confirmed based on lasso and elastic net.

Based on the lasso findings in this study, patients with higher BMI had lower mortality rate (HR=0.89). Past studies like present study have shown that high BMI is protective in these patients<sup>19-21</sup>. Hazard ratio of death, in comparison with younger patients, increases with patient age, consistent with another study<sup>22</sup>.

Results of all the statistical models used here demonstrated the importance of level of education and occupation also were confirmed<sup>23</sup>. For instance, in the present study, for one unit of increase in the level of education the risk of mortality was reduced 0.32 based on the lasso. This finding can be justified by higher adherence of patients with a higher level of education to treatment.

The risk of mortality was higher in unemployed patients, which is justifiable based on undesirable economic situation for these patients. These patients often have poor quality of life. Therefore, by postponing treatment due to its considerable costs, they do not receive treatment in the best possible manner.

In this study, based on the lasso results, for one unit of increase in the duration of dialysis, the risk of death was decreased by 0.07; the importance of this variable was confirmed elsewhere<sup>24</sup>. This finding can be explained as the longer the duration of dialysis, the more enhanced its quality would be. Although healthcare workers in dialysis wards are aware of this point, they do not pay attention to it due to the large number of patients per work shift and inadequate number of dialysis devices per patients.

Based on the results of this study, increasing one unit in the number of dialysis sessions per week decreases the risk of mortality, compatible with some other studies<sup>3,25</sup>. This finding is consistent with our expectations. In other words, if needed, we expect the patient to recover faster by increasing the number of dialysis sessions per week.

No study so far has compared the performance of PCRM to determine the survival of dialysis patients in low-dimensional data with few-events, done in this study for the first time. However, there were limitations too. One limitation is that semi-parametric PCRM were compared here, and no parametric model and the non-parametric were included for the comparisons. It is recommended for future studies to include alternative parametric and non-parametric models for low-dimensional data with few events.

## Conclusion

The performance of penalized models, especially the lasso, was satisfying in low-dimensional data with few-events based on dialysis data simulation and real data, therefore these models are the good choice for managing of low-dimensional data with low EPV.

## Acknowledgements

We thank staff of Dialysis Ward of Shahid Mohammadi, Pediatric, and Persian Gulf Hospitals, Bandar Abbas, Iran for assist and allow us to collect data.

## Conflict of interest

The authors declare that there is no conflict of interests.

## Funding

No funding was received.

### Highlights

- The lowest RMSE and the best calibration slope are achieved by the lasso, followed by the elastic net.
- For EPV 2, the C-index the lasso and the ridge are equal. For EPV 5, the highest C-index is attained using lasso and adaptive lasso.
- The performance of penalized models, especially the lasso, is satisfying in low-dimensional data with low EPV.
- The most important factors in predicting the survival of dialysis patients are BMI, education, occupation, dialysis duration, dialysis sessions and dialysis onset.

## References

- Shiba N, Shimokawa H. Chronic kidney disease and heart failure-Bidirectional close link and common therapeutic goal. *J Cardiol*. 2011; 57(1): 8-17.
- Arefzadeh A, Lessanpezeski M, Seifi S. The cost of hemodialysis in Iran. *Saudi J Kidney Dis Transpl*. 2009; 20(2): 307.
- Aghighi M, Heidary Rouchi A, Zamyadi M, Mahdavi-Mazdeh M, Rajolani H, Ahrabi S, et al. Dialysis in Iran. *Iran J Kidney Dis*. 2008; 2(1): 11-5.
- Tabrizi R, Zolala F, Nasirian M, Baneshi MR, Etminan A, Sekhavati E, et al. Estimation of the prevalence of chronic kidney disease: The results of a model based estimation in Kerman, Iran. *Med J Islam Repub Iran*. 2016; 30: 338.
- Esmaili H, Majlessi F, Montazeri A, Sadeghi R, Nedjat S, Zeinali J. Need assessment of hemodialysis patients to improve dialysis adequacy in Iran: a qualitative study. *Medical Science*. 2018; 22(91): 335-42.
- Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol*. 2016; 76: 175-82.
- Verweij PJ, Van Houwelingen HC. Penalized likelihood in Cox regression. *Stat Med*. 1994; 13(23-24): 2427-36.
- Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997; 16(4): 385-95.
- Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005; 67(2): 301-20.
- Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc*. 2006; 101(476): 1418-29.
- Tibshirani RJ, Taylor J. Degrees of freedom in lasso problems. *Ann Stat*. 2012; 40(2): 1198-232.
- Ojeda FM, Müller C, Börnigen D, Tregouet D-A, Schillert A, Heinig M, et al. Comparison of cox model methods in a low-dimensional setting with few events. *Genomics Proteomics Bioinformatics*. 2016; 14(4): 235-43.
- Ambler G, Seaman S, Omar R. An evaluation of penalised survival methods for developing prognostic models with rare events. *Stat Med*. 2012; 31(11-12): 1150-61.
- Pavlou M, Ambler G, Seaman S, De Iorio M, Omar RZ. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events. *Stat Med*. 2016; 35(7): 1159-77.
- Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biom J*. 2008; 50(4): 457-79.
- Hoseini M, Bahrampour A, Mirzaee M. Comparison of weibull and lognormal cure models with cox in the survival analysis of breast cancer patients in Rafsanjan. *J Res Health Sci*. 2017; 17(1): e00369.
- Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. New York (NY): Springer; 2009.
- Xia X, Zhao C, Luo Q, Zhou Q, Lin Z, Guo X, et al. Nomogram for Predicting Cardiovascular Mortality in Incident Peritoneal Dialysis Patients: An Observational Study. *Sci Rep*. 2017; 7(1): 13889.
- Akhlaghi AA, Najafi I, Mahmoodi M, Shojaee A, Yousefifard M, Hosseini M. Survival analysis of Iranian patients undergoing continuous ambulatory peritoneal dialysis using cure model. *J Res Health Sci*. 2013; 13(1): 32-6.
- Agarwal R. Body mass index-mortality paradox in hemodialysis: can it be explained by blood pressure? *Hypertension*. 2011; 58(6): 1014-20.
- Marcelli D, Usvyat LA, Kotanko P, Bayh I, Canaud B, Etter M, et al. Body composition and survival in dialysis patients: results from an international cohort study. *Clin J Am Soc Nephrol*. 2015; 10(7): 1192-200.
- Villar E, Remontet L, Labeeuw M, Ecochard R. Effect of age, gender, and diabetes on excess death in end-stage renal failure. *J Am Soc Nephrol*. 2007; 18(7): 2125-34.
- Kusztal M, Nowak K, Magott-Procelewaska M, Weyde W, Penar J. Evaluation of health-related quality of life in dialysis patients. Personal experience using questionnaire SF-36. *Pol Merkur Lekarski*. 2003; 14(80): 113-7.
- Goh A, Vathsala A. Native renal cysts and dialysis duration are risk factors for renal cell carcinoma in renal transplant recipients. *Am J Transplant*. 2011; 11(1): 86-92.
- Chandrashekar A, Ramakrishnan S, Rangarajan D. Survival analysis of patients on maintenance hemodialysis. *Indian J Nephrol*. 2014; 24(4): 206-13.