Iran J Public Health, Vol. 45, No.1, Jan 2016, pp.27-33

Original Article

Identifying Important Risk Factors for Survival in Kidney Graft Failure Patients Using Random Survival Forests

Omid HAMIDI [1], ***Jalal POOROLAJAL*** [2], ***Maryam FARHADIAN*** [3], ****Leili TAPAK*** [4]

1. *Dept. of Science, Hamadan University of Technology, Hamadan, Iran*
2. *Modeling of Noncommunicable Diseases Research Center, Department of Epidemiology, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran*
3. *Modeling of Noncommunicable Diseases Research Center, Dept. of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran*
4. *Dept. of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran*

***Corresponding Author:** Email: l.tapak@umsha.ac.ir

(Received 22 Feb 2015; accepted 10 Jul 2015)

Abstract

Background: Kidney transplantation is the best alternative treatment for end-stage renal disease. Several studies have been devoted to investigate predisposing factors of graft rejection. However, there is inconsistency between the results. The objective of the present study was to utilize an intuitive and robust approach for variable selection, random survival forests (RSF), and to identify important risk factors in kidney transplantation patients.

Methods: The data set included 378 patients with kidney transplantation obtained through a historical cohort study in Hamadan, western Iran, from 1994 to 2011. The event of interest was chronic nonreversible graft rejection and the duration between kidney transplantation and rejection was considered as the survival time. RSF method was used to identify important risk factors for survival of the patients among the potential predictors of graft rejection.

Results: The mean survival time was 7.35±4.62 yr. Thirty-seven episodes of rejection were occurred. The most important predictors of survival were cold ischemic time, recipient's age, creatinine level at discharge, donors' age and duration of hospitalization. RSF method predicted survival better than the conventional Cox-proportional hazards model (out-of-bag C-index of 0.965 for RSF vs. 0.766 for Cox model and integrated Brier score of 0.081 for RSF vs. 0.088 for Cox model).

Conclusion: A RSF model in the kidney transplantation patients outperformed traditional Cox-proportional hazard model. RSF is a promising method that may serve as a more intuitive approach to identify important risk factors for graft rejection.

Keywords: Random survival forest, Kidney transplantation, Cox proportional hazards

Introduction

The frequency of chronic kidney disease (CKD) has been progressively increasing over the last two decades (1) and has become a worldwide public health problem. The prevalence of CKD is estimated to be 8–16% worldwide (2). Kidney transplantation is the best alternative treatment for end-stage renal disease and health-related quality of life and survival of the patients are improved compared with dialysis (3, 4). Worldwide, more than 1.4 million patients with CKD receive renal replacement therapy with incidence growing by approximately 8% annually (5). Unfortunately, despite significant improvement in graft function, kidney transplants can still fail due to acute rejec-

tion and chronic allograft nephropathy (1, 3) that can lead to three fold greater risk of death compared to patients with functioning grafts (1, 6). Due to the increasing demand for renal transplants, identifying potential risk factors implicated in graft failure is essential to improve patient survival and quality of life (1).

To achieve this purpose, traditional statistical techniques such as Cox proportional hazards (PH) model has been widely used to analyze survival data and to determine potential risk factors. However, it relies on restrictive assumptions such as proportionality of hazards and linearity of effects on log hazard function (linearity assumption) (7). Besides, the performance of traditional methods like Cox regression is not reliable in the presence of high rate of censoring (8). Potential prognostic factors affecting renal graft have also been investigated by several studies with Cox PH model (3, 9, 10). However, there were inconsistencies among the results. Ideally, it would be important to improve the predictive performance of the models identifying potential prognostic factors affecting renal graft via learning theory and data mining techniques for survival time that require no assumptions.

Machine learning methods such as tree-based approaches have recently been developed to handle right censored survival data and their effective performance has been confirmed in different areas (11). Random survival forests (RSF), is a nonparametric tree-based ensemble learning method that can automatically handle the difficulties of Cox model and can also be used to select and rank variables (7, 11).

Due to the limitations of the Cox model, using RSF to identify effective risk factors for survival has been suggested (7). Although, several studies have confirmed the promising performance of RSF compared to traditional Cox model (8, 12-14) in different disease, there is no attempt to use RSF in renal transplantation and compare its performance with Cox model.

This study aimed to identify prognostic factors affecting renal graft by RSF and compare its performance with Cox proportional hazard model.

Material and Methods

The present study utilized a data set corresponds to a retrospective cohort study which was conducted in Hamadan, western Iran, from 1994 to 2011. The number of 475 patients underwent kidney transplantation in Ekbatan or Besaat hospitals and was eligible to enroll the study. To identify important risk factors, the patients who did not have any information about risk factors were eliminated from the analysis. In this regard, only 378 out of 475 patients were considered in the present study because the information about potential risk factors was not observed for the rest of the patients.

The risk factors were age, sex of donors and recipients, type of donor (living-donor or deceased donor), familial relationship, hemoglobin level, blood groups of donors and recipients, duration of dialysis before transplantation (year), cold ischemic time (min), creatinine level at discharge, body mass index (BMI) of donor (kg/m2), left or right kidney, type of immunosuppressive drugs used (Imuran, prednisolone, cyclosporine vs. cellcept, prednisolone, cyclosporine), duration of hospitalization (day), volume of urine excretion during the first 24 h after transplantation (ml/24 h), and occurrence of acute or hyperacute rejection. In this regard, acute rejection is related to formation of cellular immunity, which usually occurs to some degree in all grafts, except between identical twins and hyperacute rejection is initiated by preexisting humoral immunity and usually manifests within minutes after transplantation (3).

The event of interest was chronic nonreversible graft rejection. The survival time was the time between kidney transplantation and episode of rejection (3).

RSF is an extension of random forest RF to right-censored survival data with the same principles underlying RF, which enjoys all its important properties (7, 15). Random forests consist of several trees based on a random sample with replacement. Each tree consists of nodes (variables) in which classification or split was implemented. In survival settings, tree node splits according to maximizing survival differences between daughter

nodes (new nodes). In this regard, in each tree, survival time and status of the patients were considered as response variables. Then, the ensemble estimate for the cumulative hazard function (CHF) is drawn by calculating the CHF for each sample in a data set, and summing this ensemble over the observed survival times yields the predicted outcome referred to as ensemble mortality (a measure of mortality for a patient that has been shown to be an effective predictor of survival) (15). Each run of RSF was performed for the kidney transplant data set based on 1000 trees under log-rank splitting rule.

The importance of each model covariate was also determined by a rapidly computable internal measure of variable importance (VIMP) that can be used to rank variables. The larger VIMP, the more predictive the variable (the threshold value is 0.002) (11). Moreover, multiple imputation strategy based on RF was utilized for treating missing data (7). Five imputed data set were provided and then combining rules (16, 17) were applied to calculate evaluation criteria and VIMP. In order to compare the performance of RSF and traditional Cox PH, two criteria were used including integrated Brier score (18) and C-index (19) using out-of-bag (OOB) data. A perfect prediction rule would have a concordance of 1 (20).

Analyses were performed by using "randomForestSRC", a freely available package from the Comprehensive R Archive Network (CRAN).

Results

The mean survival time for 378 patients was 7.35±4.62 yr, the median survival time was 6.81 yr. Out of 378 transplantations, 37 (10%) episodes of rejection occurred, and the remaining 341 patients (90%) were censored. Table 1 shows the VIMP of the variables obtained from RSF.

Table 1: Mean and standard deviation of variable importance (VIMP) for kidney transplant data over five imputed data set. Each run based on 1000 trees under log-rank splitting

Variables	VIMP	SD	
Cold ischemic time (min)	0.0153	0.0003	*
Recipient's age (yr)	0.0139	0.0023	*
Creatinine level at discharge (mg/dl)	0.0122	0.0018	*
Donors' age (yr)	0.0113	0.0002	*
Duration of hospitalization (day)	0.0036	0.0003	**
Immunosuppressive drug usage Imuran, Prednisolone, Cyclosporinev.s. CellCept, Prednisolone, Cyclosporine	0.0016	0.0013	
Hemoglobin level (mg/dl)	0.0015	0.0004	
Type of donors Living-donor Deceased-donor	0.0008	0.0001	
Post-transplantation condition No complication, Acute rejection, Hyperacute rejection	-0.0003	0.0003	
Recipient sex	-0.0003	0.0003	
Familial relationship	-0.0004	0.0002	
Donor sex	-0.0005	0.0000	
Urine volume (ml/24 h)	-0.0005	0.0006	
Donor blood group	-0.0006	0.0000	
Side of the kidney	-0.0007	0.0001	
Recipient blood group	-0.0021	0.0001	
Duration of dialysis (yr)	-0.0047	0.0014	

The cold ischemic time, recipient's age, creatinine level at discharge and donors' age are highly predictive, and duration of hospitalization is moderately predictive. However, type of donors, hemoglobin level, donor's sex, immunosuppressive drug usage, post-transplantation condition, recipient sex, familial relationship, donor and recipient blood group, side of the kidney, duration of dialysis and urine volume are unlikely to be predictive.

According to Cox PH model, three variables of recipient age, type of donor (living vs. deceased), and episode of post-transplantation acute and hyperacute rejection were identified as most important variables. Two criteria of evaluation were also computed for Cox PH model. RSF had lower prediction error based on integrated Brier score (0.081) compared to Cox model (0.088). In addition, the C-index of RSF was considerably higher (0.965) than that of the Cox model (0.766).

The effect on survival of the most five influential covariates found in the RSF analysis was displayed with 5-yr partial survival plots in Fig. 1. The estimated partial survival for a covariate indicates estimated survival for different levels of the covariate when the effects of all other covariates are justified. It can be seen from figure that, as cold ischemic time increases up to about 35 minute, the five-year predicted survival increases as well and it tends to decline after 35 minute.

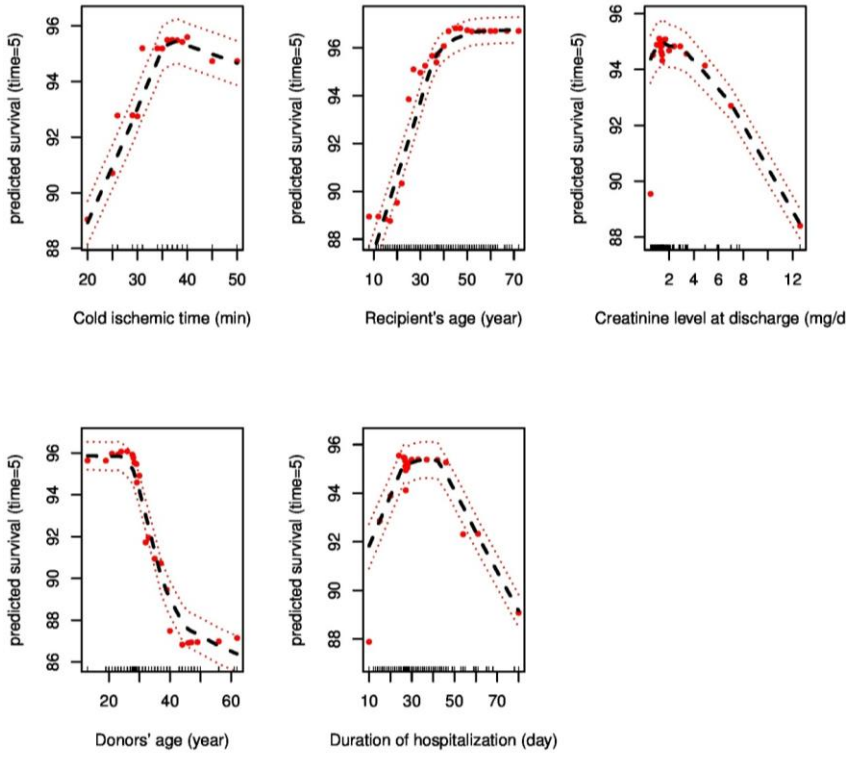

Fig.1: Partial 5-year predicted survival for five most influential covariates on survival in kidney transplant data. Dashed red lines are ± 2 standard error bars

Discussion

RSF identified cold ischemic time, recipient's age, creatinine level at discharge, donor's age and duration of hospitalization as the top five most important predictors of survival for graft failure patients in the present study.

Several authors estimated the survival rate of kidney transplantation and detected the risk factors of graft rejection (21-24). Our results showed that

the cold ischemic time variable was the most important factor in the risk of graft rejection, which is consistent with the results of some other studies (25, 26). Cold ischemic time is one of the risk factors that is involved in immediate anemia in renal transplant recipients (27). The second top risk factor was recipient's age. Based on the results, as recipient's age increases predicted five-year survival time increases as well. This may be a result of stronger and more efficient immune system in younger recipients (3). Previous studies confirm this finding (28-31). The third important variable was creatinine level at discharge. Previous studies have reported creatinine level at discharge as a risk factor in rejection of kidney transplantation (9, 10, 32). Donor age was the fourth top risk factor, which had a negative correlation with graft rejection, i.e. kidney rejection is more likely among those recipients who receive kidney from older donors. This result is also similar to the result of other studies (3, 25, 26, 33, 34). The fifth top most important variable was duration of hospitalization, confirmed eralier (35, 36).

This study focused on the performance of RSF method in identifying potential risk factors for survival of kidney graft failure patients compared to traditional Cox model. The results demonstrated that the RSF model performed significantly better than the conventional Cox-proportional hazard model. Several studies also confirmed the promising performance of RSF compared to Cox PH model in real data sets (8, 12, 14). RSF had better performance compared to Cox PH model based on prediction error criterion (13). Therefore, it can be applied successfully for identifying risk factors of the kidney transplantation survival.

RSF deals with the traditional Cox model issues such as proportionality assumption coherently and automatically (37) and analysts do not require knowing in advance the relationship (i.e. linear, nonlinear) of a variable over time (8). Besides, the performance of Cox regression is not reliable in the presence of high rate of censoring which was the case in the present study (about 90% censor rate). While, RSF is a robust extension of random forest a highly used machine learning method that has gained much interest in a variety of fields of application and generated a vast amount of computational literature in the last decade (8, 38). However, the performance of different methods is data dependent and conducting additional studies is needed to compare RSF to Cox regression to document further its performance in clinical settings (8).

There were some limitations in the present study. Reliable sources of data obtained from prospective design were required for estimation of survival rate and associated prognostic factors, but the present study used a data set of a retrospective cohort study and medical records. Quality and accuracy of estimates depends primarily on the quality of recorded data, but verifying the accuracy of data was not possible in the present study. Besides, quality of the services and technology may vary over time, but we have no document to justify this issue. These issues might bias results. In addition, long-term follow-up duration results in losing some patients, which in turn may lead to biased results (3).

Conclusion

RSF identified a different subset of risk factors in chronic nonreversible renal graft rejection than the Cox PH model. Moreover, RSF model outperformed traditional Cox PH model. The RSF is a promising method for intuitive variable selection and is a way to eliminate the doubt in the "black box" approach to statistical analysis that should be further investigated in survival analysis of other diseases (8).

Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgments

We would like to thank the Vice-Chancellor of Health of Hamadan University of Medical Sci-

ences for their approval and support of this study. No potential conflict of interest relevant to this article was reported.

References

1. Hesketh CC, Knoll GA, Molnar AO, Tsampalieros A, Zimmerman DL (2014). Vitamin D and kidney transplant outcomes: a protocol for a systematic review and meta-analysis. *Systematic Rev,* 3:64.
2. Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, Saran R, Wang AY, Yang CW (2013). Chronic kidney disease: global dimension and perspectives. *Lancet,* 382:260-72.
3. Saatchi M, Poorolajal J, Amirzargar MA, Mahjub H, Esmailnasab N (2013). Long-term survival rate of kidney graft and associated prognostic factors: A retrospective cohort study, 1994–2011. *Ann Transplant,* 18:153-160.
4. van Walraven C, Austin PC, Knoll G (2010). Predicting potential survival benefit of renal transplantation in patients with chronic kidney disease. *Canadian Med Ass J,* 182:666-672.
5. White SL, Chadban SJ, Jan S, Chapman JR, Cass A (2008). How can we achieve global equity in provision of renal replacement therapy? *Bull World Health Organ,* 86:229-37.
6. Tonelli M, Wiebe N, Knoll G, Bello A, Browne S, Jadhav D, Klarenbach S, Gill J (2011). Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes. *Am J Transplant,* 11:2093-109.
7. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS (2008). Random survival forests. *Ann Appl Statis*, 2(3):841-860.
8. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS (2011). Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. *Circulation: Cardiovascular Quality and Outcomes,* 4:39-45.
9. Ghadiani MH, Peyrovi S, Mousavinasab SN, Jalalzadeh M (2012). Delayed graft function, allograft and patient srvival in kidney transplantation. *Arab J Nephrol Transplant,* 5:19-24.
10. Karabicak I, Aytug S, Lewis S, Shah S, Sumrani N, Hayat A, Distant DA, Salifu MO (2011). Long-term kidney transplant outcome in obese patients in a predominantly African American population. *Clin Transplant,* 25:E264-E270.
11. Chen X, Ishwaran H (2012). Random forests for genomic data analysis. *Genomics,* 99:323-329.
12. Friedel G, Fritz P, Goletz S, Kristen R, Brinkmann F, Dierkesmann R, Schwab M, Ott G, Dippon J, Alscher MD (2013). Postoperative survival of lung cancer patients: are there predictors beyond TNM? *Anticancer Res,* 33:1609-1619.
13. Mogensen UB, Ishwaran H, Gerds TA, Afdeling B (2010). *Evaluating random forests for survival analysis using prediction error curves.* ed. Department of Biostatistics, University of Copenhagen.
14. Myte R (2013). Covariate Selection for Colorectal Cancer Survival Data: A comparison case study between Random Survival Forests and the Cox Proportional-Hazards model.
15. Ishwaran H, Kogalur UB, Chen X, Minn AJ (2011). Random survival forests for high-dimensional data. *Statistis Anal Data Mining,* 4:115-132.
16. Rubin DB (1996). Multiple imputation after 18+ years. *J Am Statis Ass,* 91:473-489.
17. Rubin DB (1987). *Multiple imputation for nonresponse in surveys*. New York: J. (ed)^(eds), Wiley & Sons.
18. Gerds TA, Schumacher M (2006). Consistent Estimation of the Expected Brier Score in General Survival Models with Right-Censored Event Times. *Biometr J,* 48:1029-1040.
19. Pencina MJ, D'Agostino RB (2004). Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statis Med,* 23:2109-2123.
20. Farhadian M, Mahjub H, Moghimbeigi A, POOROLA J, Mansoorizadeh M (2014). A Gene Selection Method for Survival Prediction in Diffuse Large B-Cell Lymphomas Patients using 1D Discrete Wavelet Trans¬ form. *Iran J Public Health,* 43:1091-1098.
21. Vavallo A, Lucarelli G, Spilotros M, Bettocchi C, Palazzo S, Selvaggi FP, Battaglia M, Ditonno P (2014). Impact of donor–recipient gender on kidney graft and patient survival: short-and long-term outcomes. *World J Urol,* 32:709-714.
22. Kwon OJ, Kwak JY, Kang CM (2005). The impact of gender and age matching for long-

term graft survival in living donor renal transplantation. *Transplant Proc,* 37:726-8.
23. Einollahi B (2004). Iranian experience with the non-related renal transplantation. *Saudi J Kidney Dis Transpl,* 15:421-8.
24. Feyssa E, Jones-Burton C, Ellison G, Philosophe B, Howell C (2009). Racial/ethnic disparity in kidney transplantation outcomes: influence of donor and recipient characteristics. *J Natl Med Assoc,* 101:111-5.
25. Sert I, Colak H, Tugmen C, Dogan SM, Karaca C (2014). The effect of cold ischemia time on delayed graft function and acute rejection in kidney transplantation. *Saudi J Kidney Dis Transpl,* 25:960-6.
26. Gonzalez-Molina M, Burgos D, Cabello M, Ruiz-Esteban P, Rodriguez MA, Gutierrez C, Lopez V, Baena V, Hernandez D (2014). Impact of immunosuppression treatment on the improvement in graft survival after deceased donor renal transplantation: a long-term cohort study. *Nefrologia,* 34:570-578.
27. Rostami Z, Shafighi N, Baghersad MM, Einollahi B (2011). Risk factors for immediate anemia in renal transplant recipients: a single-center experience. *Transplant Proc,* 43:581-3.
28. Andreoni KA, Forbes R, Andreoni RM, Phillips G, Stewart H, Ferris M (2013). Age-related kidney transplant outcomes: health disparities amplified in adolescence. *JAMA Intern Med,* 173:1524-32.
29. Oetting WS, Guan W, Schladt DP, Wildebush WA, Becker J, Thyagarajan B, Jacobson PA, Matas AJ, Israni AK (2014). Telomere length of recipients and living kidney donors and chronic graft dysfunction in kidney transplants. *Transplantation,* 97:325-9.
30. Gondos A, Dohler B, Brenner H, Opelz G (2013). Kidney graft survival in Europe and the United States: strikingly different long-term outcomes. *Transplantation,* 95:267-74.
31. Smedbråten YV, Sagedal S, Mjøen G, Hartmann A, Fagerland MW, Rollag H, Mollnes TE, Thiel S (2014). High Ficolin-3 Level at the Time of Transplantation Is an Independent Risk Factor for Graft Loss in Kidney Transplant Recipients. *Transplantation,* 99(4): 791-796.
32. Salido E, Martin B, Barrios Y, Linares JD, Hernandez D, Cobos M, Checa MD, Hortal L, Fernandez A, Garcia JJ, Torres A (1999). The PlA2 polymorphism of the platelet glycoprotein IIIA gene as a risk factor for acute renal allograft rejection. *J Am Soc Nephrol,* 10:2599-605.
33. Chen GD, Gu JL, Zhang XD, Qiu J, Wang CX, Chen LZ (2013). Donor factors predictive for poor outcomes of living donor kidney transplantation. *Transplant Proc,* 45:1445-8.
34. Oliaei F, Akbari R, Ghazi Mirsaeid AM (2012). Adding thymoglobuline to the conventional immunosuppressant regimen in kidney transplantation: A cost-benefit analysis. *Caspian J Intern Med,* 3:514-8.
35. Ounissi M, Cherif M, Abdallah T, Bacha M, Hedri H, Abderrahim E, Goucha R, Kheder A, Slama R, Derouiche A (2013). Risk factors and consequences of delayed graft function. *Saudi J Kidney Dis Transplan,* 24:243.
36. Kosmadakis G, Daikos GL, Pavlopoulou ID, Gobou A, Kostakis A, Tzanatou-Exarchou H, Boletis JN (2013). Infectious complications in the first year post renal transplantation. *Transplant Proc,* 45:1579-83.
37. Datema FR, Moya A, Krause P, Back T, Willmes L, Langeveld T, Baatenburg de Jong RJ, Blom HM (2012). Novel head and neck cancer survival analysis approach: random survival forests versus Cox proportional hazards regression. *Head Neck,* 34:50-8.
38. Boulesteix AL, Janitza S, Kruppa J, König IR (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 2:493-507.