Predictors of mortality among hemodialysis patients in Hamadan province using random survival forests

LEILI TAPAK1,2, VIDA SHEIKH3, ENSIYEH JENABI4, SALMAN KHAZAEI5

1 Department of Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran; 2 Modeling of Non-communicable diseases research center, Hamadan University of Medical Sciences, Hamadan, Iran; 3 Clinical Research Development Unit of Shahid Beheshti Hospital, Hamadan University of Medical Sciences, Hamadan, Iran; 4 Autism Spectrum disorders Research Center, Hamadan University of Medical Sciences, Hamadan, Iran; 5 Research Center for Health Sciences, Hamadan University of Medical Sciences, Hamadan, Iran

Keywords
Hemodialysis • Kidney Failure • Survival • Random Survival Forest

Introduction
Globally, chronic kidney disease (CKD) is an important health challenge with an alarming increase in incidence as well as prevalence [1]. CKD patients have a lower survival rate, and no significant improvement has been achieved on their survival over the past two decades [2]. In Iran, unlike some other countries, hemodialysis (HD) is the main way of renal replacement therapy in end stage renal disease (ESRD) patients [3]. Limited studies have been conducted in Iran regarding the survival of HD patients, suggesting a low survival rate for these patients [4, 5]. Common independent predictors of survival in HD patients are: age, race, albumin and hemoglobin levels, etiology of kidney failure and presence of certain comorbidities [3, 6-9]. Apart from these accepted prognostic factors, the role of some factors such as dialysis frequency is controversial on morbidity and mortality among hemodialysis patients [10]. In general, the interaction occurs when the effect of one risk factor is dependent on the presence of another risk factor [11]. When resources are limited, assessing interactions provides insight into the mechanisms for the outcome and identifying subgroups would benefit most from interventions [11, 12]. Interactions between some variables like three-way interaction between protein-energy wasting (PEW), inflammation and cardiovascular diseases (CVD) [13], interaction between albumin and phosphor [14] and three-way interaction between BMI, physical activity and smoking [15] on HD patients survival were shown previously. However, the lack of knowledge in this regard is still high. Identification of survival modifiable factors and interactions between them could help in prioritizing the clinical care of HD patients. Evidences regarding the survival rate of hemodialysis patients in developing countries are rare and most of these studies were conducted in developed countries where patients have more access to health cares, and their results cannot be generalized to developing countries. Therefore, in the present study, we aimed to investigate the survival rate and it’s correlates in HD patients in Hamadan province using random survival forests.

Material and methods

STUDY DESIGN
We performed a retrospective cohort study to investigate predictors of mortality among hemodialysis patients.

STUDY SETTINGS
We examined data on 758 patients who underwent hemodialysis treatment in Hamadan province in western Iran from March 2007 to March 2017. Hamadan province, with an area of 19,493 square kilometers in extent,
is located in western Iran. According to the national census held by the Statistical Center of Iran, Hamadan province had a population of 1,758,268 people in 2011. We used information of patients from eight hospitals in the province with a dialysis wards, including: Alimoradian, Besat, Vali-asr, Ghaem, Imam Hossein, Valtasr, Imam Reza, and Shahid-Beheshti in Nahavand, Hamadan, Tuyserkan, Asadabad, Malayer, Razan, Kabudarahang and Hamadan city, respectively.

Eligibility criteria
All ESRD patients who initiated chronic hemodialysis programs at the dialysis units of the above mentioned hospitals during 2007 to 2017 were considered as inclusion criteria. Patients with acute renal failure or under treatment with peritoneal dialysis, patients on transient hemodialysis and patients with incomplete medical records were excluded from study.

Clinical and demographic measures
Data were gathered by a checklist on hospital records of all HD patients in the province. The checklist used in this study included characteristics related to demographic profiles (age (year), gender (male, female), marital status (married, single, divorce, widow), BMI (kg/m²), residence area (urban, rural), educational level (illiterate, primary, guidance, high school, academic), the history of tobacco use (yes, no) and substance abuse (yes, no)), and clinical information (including Hemoglobin (g/dl), BUN (mg/dl), Creatinine (mg/dl), CRP status (positive, negative), Sodium (mEq/L), Calcium (mg/dl), Phosphor (mg/dl), iPTH (pg/ml), Albumin (g/dl) and ESRD cause (Hypertension, Diabetes, Urologic & obstructive diseases, Polycystic Kidney, Glomerulonephritis, Un-Known)). The clinical and laboratory information of patients at the beginning of their treatment and before receiving the first dialysis treatment was gathered. In order to minimize measurement variability, all two baseline measures (two last measurements before the first dialysis treatment) for each patient were averaged. These records were collected by reviewing patients’ medical records.

Outcome
We considered the end point of the patients’ follow up as their death. The survival time was the time between the start of HD treatment to patients death. Patients with renal transplantation, or withdrawal of dialysis, or unable to follow them due to transferred another dialysis facility out of province as well as, patients who died because of injury or accident, or other causes unrelated to renal failure were considered as censored cases.

Statistical analysis
We utilized the random survival forest (RF) method that ensembles binary decision trees and extends the RF regression model to right-censored survival data. In this technique, a random bootstrap sample (containing two thirds of the original data on average) is drawn for every decision tree. Thus, the remaining one third of the data known as out-of-bag (OOB) data is excluded. Decision trees are grown based on the bootstrap samples by applying a random node splitting process which works as follows: At each node random candidate variables (mtry determined by the square root of the total number of variables) are selected for random node splitting. Then the variable that maximizes the survival differences between two daughter nodes (determined by a splitting rule like the log-rank statistic) for a special split point is selected for node splitting. The growth of a decision tree is stopped when all the terminal nodes contain only a predefined minimal number of unique events. To comparison, the stepwise Cox proportional hazards model was used. So, the Harrell’s concordance index (C-index) criterion was utilized [16] using out-of-bag (OOB) data. The minimal depths of the covariates were obtained to select predictive variables. Minimal depth is a dimensionless order statistic that measures the predictiveness of a variable in a tree. It can be used to select variables in high-dimensional problems. It assesses the predictiveness of a variable by a depth calculation relative to the root node of a tree. The smaller the minimal depth, the more predictive the variable.

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

Results
Descriptive statistics of the characteristics of the patients were presented in Tab. I. According to the table, the majority of the patients were male (54.9%), married (79.1%), non-smoker (76.7%), non-substance abused, illiterate (54.9%) and lived in rural area (65.6%). About 47.8% of the patients experienced death. The median survival time was 613 days. The mean and standard deviation of other variables were reported in Table I.

Figure 1 shows the minimal depth of the variables obtained from RSF. According to the figure, hemoglobin, CRP, albumin, age at diagnosis and iron were highly predictive and URR, uric acid, dialysis weekly time, PLT, Na, P, Ca, K, HCT, ALK and vascular access were moderately predictive. Moreover, other used variables were unlikely to be predictive. According to the stepwise Cox PH model, the variables of age at diagnosis, marriage status, BMI, addiction, hemoglobin, iron, albumin and CRP were selected as the most important variables (Tab. II). The performance of the RSF was assessed using Harrel’s C-index and compared with the stepwise Cox PH. According to the results, the RSF had a higher c-index (0.808) compared with the Cox model (0.727). The effects of the most four influential variables found in the RSF with 5-year partial survival plots analysis were demonstrated in Figure 2. The estimated partial survival for a variable shows the estimated survival for different levels of the variable when the effects of all other variables are justified. For example, patients with negative CRP shower a higher 5-year predicted survival.
compared with those with positive CRP. The non-linear nature of the selected variables is evident from the figure. For example, as hemoglobin increases up to about 10 (g/dl (the five-year predicted survival increases very slightly and it tends to increase dramatically after that point up to.

Figure 3 displays interaction between the three important variables of CRP, hemoglobin, and Albumin on 5 year predicted survival. Patients with positive CRP and hemoglobin values lower than 11 have the worst survival (see the first row, second column) and most had low albumin. Survival was best for patients
Tab. II. The results of stepwise Cox proportional Hazards in selecting predictors affecting survival time of HD patients in Hamadan Province from 2007 to 2017.

<table>
<thead>
<tr>
<th>Variables</th>
<th>B</th>
<th>SE</th>
<th>P.Value</th>
<th>Exp(B)</th>
<th>95% CI for Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at diagnosis</td>
<td>-0.023</td>
<td>0.006</td>
<td>&lt; 0.001</td>
<td>0.977</td>
<td>0.966 - 0.988</td>
</tr>
<tr>
<td>Marriage status</td>
<td>Married</td>
<td>1.000</td>
<td></td>
<td>1.000</td>
<td>1.000 - 1.000</td>
</tr>
<tr>
<td></td>
<td>Single</td>
<td>-1.425</td>
<td>0.615</td>
<td>0.021</td>
<td>0.241 - 0.804</td>
</tr>
<tr>
<td></td>
<td>Divorce</td>
<td>-0.174</td>
<td>0.464</td>
<td>0.707</td>
<td>0.840 - 2.085</td>
</tr>
<tr>
<td></td>
<td>Widow</td>
<td>0.308</td>
<td>0.486</td>
<td>0.326</td>
<td>1.361 - 3.525</td>
</tr>
<tr>
<td>Location</td>
<td>Urban</td>
<td>1.000</td>
<td></td>
<td>1.000</td>
<td>1.000 - 1.000</td>
</tr>
<tr>
<td></td>
<td>Rural</td>
<td>-0.220</td>
<td>0.110</td>
<td>0.046</td>
<td>0.803 - 0.996</td>
</tr>
<tr>
<td>BMI status</td>
<td>&lt; 18.5 kg/m²</td>
<td>1.000</td>
<td></td>
<td>0.010</td>
<td>0.647 - 0.996</td>
</tr>
<tr>
<td></td>
<td>18.5-25 kg/m²</td>
<td>1.000</td>
<td>0.305</td>
<td>0.001</td>
<td>2.719 - 4.926</td>
</tr>
<tr>
<td></td>
<td>25-30 kg/m²</td>
<td>0.663</td>
<td>0.250</td>
<td>0.008</td>
<td>1.941 - 3.165</td>
</tr>
<tr>
<td></td>
<td>&gt; 30 kg/m²</td>
<td>0.740</td>
<td>0.265</td>
<td>0.005</td>
<td>2.096 - 3.523</td>
</tr>
<tr>
<td>Addiction</td>
<td>Yes</td>
<td>1.000</td>
<td></td>
<td>1.000</td>
<td>1.000 - 1.000</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>-0.526</td>
<td>0.156</td>
<td>0.017</td>
<td>0.722 - 0.942</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>-0.203</td>
<td>0.030</td>
<td>&lt; 0.001</td>
<td>0.816</td>
<td>0.769 - 0.866</td>
</tr>
<tr>
<td>Iron</td>
<td>-0.003</td>
<td>0.001</td>
<td>&lt; 0.001</td>
<td>0.997</td>
<td>0.996 - 0.999</td>
</tr>
<tr>
<td>Albumin</td>
<td>-0.421</td>
<td>0.070</td>
<td>&lt; 0.001</td>
<td>0.656</td>
<td>0.572 - 0.753</td>
</tr>
<tr>
<td>CRP</td>
<td>-1.157</td>
<td>0.119</td>
<td>&lt; 0.001</td>
<td>0.314</td>
<td>0.249 - 0.397</td>
</tr>
</tbody>
</table>

Fig. 2. Partial 5-year predicted survival for four most influential variables on survival in HD data. Dashed red lines are ± 2 standard error bars. Red dots indicate the estimated survival at the observed levels of each correlate and the black hatches are the smoothed curves based on the loess curves for the estimated survival for each individual.
with negative CRP and hemoglobin ≥ 11 (see second row, first column) and further dependent on changes in albumin. In this group the 5 years predicted survival was over 70% for those with albumin values greater than 4, but only about 50% for those with albumin of 1 (g/dl). It is important to note that these interactions and non-linear relationships were identified by the random survival forest, and not prespecified by the analyst. The median survival time in the two categories of low Hemoglobin/positive CRP and high Hemoglobin/negative CRP were 5.38 and 40.53 months, respectively.

Discussion

Hemodialysis is a common treatment modality for ESRD patients in Iran. In this study we investigated the predictors of survival in HD patients. The findings of this study can be used to improve quality of cares provided for the HD patients and better resource allocation. We found that hemoglobin level, albumin level, and CRP status were the top three most important predictors of survival for HD patients in the present study. Along with the results of this study, the role of low hemoglobin level in increasing the risk of mortality in
PREDICTORS OF MORTALITY IN HEMODIALYSIS PATIENTS

Anemia is one of the main risk factors in the development of left ventricular hypertrophy (LVH), and consequence cardiac mortality and morbidity in ESRD patients [19]. Early management of anemia is associated with reduction in the severity of comorbid conditions and can slow the progression to renal failure [20]. Robert et al. showed that normalization of hemoglobin through preventing the development of LV dilation, leads to improved quality of life of HD patients [21]. In agreement with our finding, other studies results indicated that low serum albumin is one of the main predictors of poor survival in HD patients [22, 23]. Hypoalbuminemia usually considered as the proxy of malnutrition. Mafra et al. in a Brazilian cohort of hemodialysis patients found that both low BMI and hypo-albuminemia are strong predictors of death [24]. Combe et al. showed that a decrease in serum albumin over time correlated with increased CVD death [25]. One of the other main predictors of mortality in HD patients in this study was CRP status. Similar findings have been obtained in other studies [26, 27]. Lseki et al. in their study showed that regardless of serum albumin and other possible confounders, CRP is a significant predictor of death in HD patients [28]. Inflammation usually is in relation with insulin resistance, oxidative stress, wasting, infections and endothelial dysfunction [29]. Interactions between some modifiable variables on survival of HD patients were seen previously [13, 15]. We found that the combined effect of CRP status with serum Hemoglobin and CRP status has the significant effect on the survival of HD patients. This finding is important because HD patients with a high mortality risk can be identified through regular screening. More studies are needed to determine multiple pathophysiological pathways may underlie these combined effects. This study had some limitations as well. First, because of the retrospective design of the study, verifying quality control of the data was not possible. Second, the addiction and smoking status of patients was based on their self-report and therefore was prone to information bias and finally quality of the services and technology may vary over time, and also the quality of service provision in the dialysis wards of hospitals is not the same, which could not be considered in this study. Despite these limitations there was strength for this study which was the utilization of the RSF to analyze the data set. RSF can handle the issues with the traditional Cox model like proportionality assumption automatically. This will help analysts to deal with the relationships (i.e. linear, non-linear) between variables over time without any previous knowledge.

Conclusions

We found that higher levels of CRP, low serum albumin and low serum hemoglobin were the top three most important predictors of poor survival for HD patients.

Acknowledgement

This study was approved by the Ethics Committee of Hamadan University of Medical Sciences. The Authors would like to thank all contributors and participants who make this study project possible as well as the personnel of the hemodialysis wards in hospitals of Hamadan province for their kind collaboration. Funding sources: this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest statement

The authors declare no conflict of interest.

Authors’ contributions

SK and LT are responsible for the design of the study. They analyzed the data and wrote the paper. VSH and EJ dealt with the collection of information. All authors reviewed and approved the final version of the manuscript.

References


