**Abstract**

***Background*** Renal dysfunction serves is a complication of heart failure through multiple mechanisms. Studies have shown that the serum level of creatinine, a substance readily filtered out by healthy kidneys, acts as an indicator of kidney function could help predict mortality for patients of HF. However, the potential effect modifications by other demographic and biological markers were not clearly assessed. This project hence sought to investigate the association between serum creatinine level and mortality rates among patients with HF.

***Methods:*** A dataset consisting of 299 patients of HF enrolled from April 2015 to December 2015 was studied. In the main analysis, serum creatinine level was first classified into two different categories (normal vs. abnormal) and then treated as a continuous variable using Cox proportional hazards models used to estimate hazard ratios (HR) and 95% confidence intervals (CI). Lasso were used to help select adjusted variables in the model.

***Results:*** Cox models comparing patients with abnormal serum creatine level to those with normal serum creatine level showed a HR of 2.29 (95% CI: 1.46 to 3.58) in the fully adjusted model. When treating the serum creatine as a continuous variable, the HR associated with 1 mg/dL increment is 1.37 (95% CI: 1.19 to 1.59). We also found a synergy effect between serum creatine and serum sodium. The HR for a with 1 mg/dL increment in serum creatine will be greater among patients with higher serum sodium. No clear nonlinear relationship was found between serum creatine and mortalities.

***Conclusion:*** Our analysis showed that there is a strong association between elevated serum creatinine level and increased risk in post-HF mortality and validated the findings of previous studies.

**Keywords**: Heart Failure; Serum creatinine; Model selection; Cox proportional hazards models.

**1. Introduction**

Heart failure (HF) is a serious medical condition that develops when the heart doesn’t pump enough blood for the whole body’s needs. HF is often caused by morbidities that damage the heart like coronary heart disease, diabetes, and high blood pressure. As a serious condition requiring medical care and treatment, HF affects more than 6 million patients and their families in the United States, which brings a high public health cost and burden.[4]

Renal dysfunction is a common complication of HF, which can lead to mortalities. The reduced cardiac output and the consequently renal under-perfusion is the main pathophysiology cause of renal dysfunction because of the low renal blood flow and increased renal venous pressure.[5] Besides, neurohormonal activation (renin–angiotensin–aldosterone and sympathetic nervous system), inflammatory activation and diuretic treatment are also mechanisms leading to renal dysfunction in patients with HF. [5] Serval study has demonstrated that renal dysfunction can lead to higher mortality rates among patients with cardiovascular diseases. [1,2]

Serum creatinine is an important and commonly-used biomarker to indicator of the presence renal dysfunction. [6] Serum creatinine level is tested through using venous blood is included in routine clinical easy procedures. Serum creatinine greater than 1.5 mg/dL will be regarded as abnormal.[1] Low serum sodium (hyponatremia) is a prognostic marker in heart failure.

In extent literature, Ahmad et al. [1] and Chicco et al. [2] have demonstrated the harmful associations between kidney dysfunction and mortalities among patients with heart diseases. However, the dataset used in their studies are of small sample size, 299 in total. The relatively small sample size would lead to great estimator variance if all the covariates in the dataset were adjusted for. Variable selection process is essential in this situation. Also, these studies concerned less on the modification of other biomarkers such as eject fraction and serum sodium, which are also related to HF through diverse mechanisms.[8] To fill the current research gap, this study aimed to re-analyze the same dataset with an additional variable selection process. We aimed to check the difference between their findings with ours in the aspects of point-estimate as well as interval estimates. Also, we explored potential effect modifications on the multiplicative scale measurement between serum creatinine and mortalities among patients with heart diseases.

**2. Review of literature and domain expertise**

The association between serum creatinine and mortalities among patients with heart diseases has been studied by Ahmad et al utilizing survival analysis[1]. In their work, Ahmad mainly identified potential risk factors for mortalities among patients with heart failure. Renal dysfunction (measured by serum creatinine) is reported as an important risk factor. A 1-unit increment in serum creatinine is associated with 2.24 hazard ratio with p-value less than 0.05. Chicco et al. re-analyzed the dataset using machine learning methods and reported that the serum creatinine level is the most important predictor for these patients’ mortalities [2], Analogous research topic has been studied by Van Domburg et al., using estimated glomerular filtration rate (eGFR) as the indicator of renal dysfunction in patients with known or suspected coronary artery disease. [3] Van Domburg et al. employed a multivariable adjusted Cox model and found the hazard ratios were 1.33, 1.67, and 3.38 among patients with mild, moderate, and severe renal impairment compared to their peers with normal renal function. In summary, although extent literature employed different biomarkers for renal dysfunction, increasing hazard ratios were reported for impaired renal function among patients with cardiovascular diseases.

***3. Research and Analysis Methods s***

3.1. Material sources and data description

The dataset was retrieved from UCI Machine Learning Repository. It was first collected and analyzed by Ahmad et al. [1] based on medical records of 299 patients with HF at NYHA class III and IV, the most two severe HF stages. A total of thirteen variables were collected in the dataset, including age (years), sex (binary), anaemia (binary), high blood pressure (binary), creatinine phosphokinase level (mcg/L), diabetes (binary), ejection fraction (percentage), platelets (kiloplatelets/mL), serum creatinine (mg/dL), serum sodium (mEq/L), smoking (binary), follow-up period (days) and death event (binary).

3.2. Exposures

In this study, the main exposure is the serum creatinine. We first treat it as a continuous variable as it is in the original dataset. Then, to make our results comparable with the previous studies, we categorized the serum creatinine into normal level (≤1.5 mg/dL) and abnormal level (>1.5 mg/dL).

3.3. Outcome

In this survival analysis project, the primary outcome is the survival time with the end point status for the patients with heart failure during the follow-up period. The survival time is on the daily scale. We also conducted a Poisson regression with the outcome as the incidence of time in our sensitivity analysis.

3.4. Statistical analysis

We first conducted a descriptive analysis to demonstrate the characteristics of the study population at baseline. T-tests and Chi-squared were used for the continuous and categorical variables respectively. Then, Kaplan Meier curves was used to visualize the difference of survival probabilities between normal serum creatine group and abnormal serum creatine group. A Log-rank test was performed. Last, we performed Cox Proportional Hazards regression model to compare the hazard ratio (HR) as well as confidence intervals comparing participants with different levels of serum creatinine.

When constructing Cox models, we first conducted a crude model by only including serum creatinine as the independent variables. Then adjusted models were performed. When adding adjusted covariates into the model, the small sample size was a main concern to us. Ideally, we should adjust as many as confounders whenever it was reasonable in causality meaning. However, the relatively small sample size forced us to just include only important variable that could help increase the model performance to get relatively precise estimates. Thus, we used LASSO to select variables added into cox model. Deviance was chosen as the metric of the model performance. WE did Lasso for binary and continuous serum creatinine separately, and the results were very similar (figure 1). According to the results of Lasso, serum creatinine, age, anemia, ejection fraction, serum sodium was deemed as the most appropriate dataset (table 1). We also slightly changed the tunning parameter (λ) to add and delete another variable to see if there is significant influence on the estimates for the Cox regressions. After adding sex into the Cox model, the estimated HR didn’t change significantly for the serum creatine. Thus, to make it more reasonable in the biological mechanism, sex was added into the model.

We also explored the effect modification by adding interaction terms into the model. Also, nonlinear relationships between continuous serum creatine and mortalities were conducted by replacing the linear term with a natural cubic spline with 3 degrees of freedom.

The significance level was set to 0.05. All the analyses were conducted using R 4.1.0. Packages “glmnet” and “survival” were used.

**4. Results**

A total of 299 patients were included in this study. Of them, 232 (77.6%) had serum creatinine lower than 15 mg/dL and were categorized as normal, 67 (22.4%) had serum creatinine greater than 15 mg/dL and were categorized as abnormal. The normal group were averagely younger than the abnormal group (59.64 years vs. 64.97 years0, with a p-value of 0.0022. In both groups, the proportion of males was slightly higher than of the females, roughly six to four in ten patients. About three in ten patients smoked and four in then ten patients had diabetes at baseline in two groups with no significant difference. The serum sodium in serum creatinine normal group (137.32 mEq/L, SD = 3.61 mEq/L) is averagely greater than in the abnormal group (134.21 mEq/L, SD = 5.89 mEq/L). Detailed difference is shown in table 1.

A total of 96 deaths from 38,948 person-days were observed during the follow-up. The median follow-up time for the patients is 115 days. However, since less than half of patients died, KM-estimates for the median survival time cannot be obtained. In the group with normal serum creatinine level, 53 of 232 patients died (Figure 2). In the group with abnormal serum creatinine level, 43 of 67 patients died, with the KM-estimated median survival time of 113 days (95% CI 82 to 196). The log-rank test demonstrated significant difference of survival rates between normal group and abnormal group (p < 0.0001).

We applied Cox regression to analyze the association between serum creatinine and mortalities among patients with heart disease. Firstly, we treated serum creatinine as binary (normal vs. abnormal). In the crude model, compared to normal group, the HR for the abnormal group is 3.39 (95% CI 2.26 to 5.08). Then we adjusted variable selected by Lasso, and the adjusted HR decreased to 2.27 (95% CI 1.45 to 3.55). Last, we further adjusted for sex, the fully adjusted HR became 2.29 (95% CI 1.46 to 3.58) not very much different from the Lasso selected model. Then we treated the serum creatinine as a continuous variable. In both Lasso adjusted and fully adjusted model, the HR associated with 1 mg/dL increment in serum creatinine is 1.37 (95% 1.19 to 1.59), suggesting that serum creatinine is harmful for the survival of patients with heart diseases (Table 3).

Survival analysis is the main component of our project. The outcome is the survival time with the end point status for the patients with heart failure during the follow-up period. The exposure of interest is serum creatinine level and, according to our analysis above, possible covariates that we could adjust on include age, ejection fraction, platelets, serum sodium level, and anemia, with the first 4 covariates being continuous and the last one being binary. However, out of the rigor of our analysis, we will conduct LASSO to select the variables to put in the model.

Results from LASSO demonstrated that a set of 5 parameters may be the best for fitting a Cox PH model. We identified age, anemia, ejection fraction, serum creatinine level, and serum sodium level as variables whose β coefficients are not shrunk to 0 by LASSO, and thus fitted a Cox PH model using these variables and obtained the model as follows:

Then we explored the potential effect modification between serum creatinine and selection. When treating serum creatinine as a binary variable, the estimated hazard ratios are homogenous across all covariate groups. However, when treating it as a continuous variable, a positive effect modification was found by serum sodium. Compared to patients with lower serum sodium concentration, the harmful effects of serum creatinine becomes much stronger among patients with higher serum sodium concentration, with a p-value of 0.017 (Table 4). Finally, we explored the potential nonlinear relationship between serum creatinine and mortalities. The relationship is found quite monotonic. No clear non-monotonic trend was demonstrated (Figure 3).

5.**Discussion**

In this study, we found that the increment of serum creatinine is positively associated with higher hazards of mortalities among patients post heart failure. The positive association is robust no matter binary or continuous variables were used. The associations are homogeneous when employed as a binary variable. Using the serum creatinine. The homogeneity roughly holds except for the serum sodium. Higher serum sodium value strengthens the harmful effects of serum creatine. When add a spline into the model, the findings show the relationship between serum creatinine and mortalities is quite linear.

Serum creatinine level serves as a powerful indicator of renal function and is shown to be associated strongly with mortality post heart failure. In the original study, Ahmad et al. found the HR with an 1 mg/dL increment in serum creatinine is 2.24 greater than our findings 1.37. The estimated standard error for the linear coefficient of serum creatinine is 0.267 greater than ours estimates, 0.267 in the fully adjusted model. Thus, the discrepancy of the estimated HR is very likely attributed to the high variability in their model. Also, this comparison confirms our assumption. Including all the variables in the dataset will decrease the precision of the estimates. In their model, Ahmad et al. adjusted for 13 variables (including four dummy variables) whereas, according to the results of Lasso and background knowledge, we only adjusted for six variables. Estimation precision can benefit from this parsimonious model. Also, we compared our study to Chicco and Jurman’s work, which used machine learning method on the same dataset. They found that the most 6 important variables are serum creatine, ejection fraction, age, serum sodium, high blood pressure, and anemia. Except the high blood pressure which was not included in our analysis because of the unclear definition, all the remain five variables are automatically selected by Lasso as the most appropriate variable set. Overall, our studies are consistent with findings in previous research using the same dataset.

Our study showed that the effects of serum creatinine are quite homogenous across all subgroup except serum sodium. The effect modification by serum sodium indicated :

In summary, we found the higher serum creatinine value, an indicator of renal dysfunction is associated higher hazards of mortalities among patients post heart failure. Serum sodium concentration which can also indicate renal dysfunction, can worsen the harmful effects of the serum creatinine value. These findings are of clinical importance. Additional care should be paid to patients with renal dysfunction after heart failures.

**6. Limitations**

There are several limitations in this study. First, the sample size is really small, which didn’t allow us to include enough confounders. We had to make a trade-off between bias and precision. We chose Lasso to get an efficient estimate, but we admit that the estimated HR can be somehow biased. Second, we chose the cut-off of 1.5 mg/dL for the serum creatinine, which is a general standard for adults and is mentioned in the original study by Ahmad et al. However, other literature recommends different cut-offs for men and women separately. Also, age is also an important factor influencing the cut-off value. To simplify the analysis in this study, we just chose the 1.5 mg/dL cut-off. This simplification prohibits us from exploring the finer association by sex and age groups. Last, several other confounders such as SES and heart failure stage were not included in the data. Omitting those confounders not only can bias our estimates but also decrease the precision of the estimation.

7. Future scope

Although some studies have mentioned the correlation between kidney dysfunction and HF mortality, we still believe that our findings will potentially help explain the complex biological mechanisms of death due to HF, not just in the manner that HF causes serious kidney dysfunction, but also in the mechanism that kidney dysfunction leads to a more serious HF in acute and chronic backgrounds.

The interaction between serum creatinine and serum sodium suggests that we may be able to generate a new variable representing the value of kidney dysfunction based on some indicators used to evaluate kidney dysfunction (serum creatinine, serum sodium also include glomerular filtration rates, serum cysC and other variables), which will effectively Eliminate collinearity, and it is likely to be able to predict mortality in HF patients more accurately.

**Reference**

1. Anand, I., McMurray, J. J., Whitmore, J., Warren, M., Pham, A., McCamish, M. A., & Burton, P. B. (2004). Anemia and its relationship to clinical outcome in heart failure. Circulation, 110(2), 149–154. https://doi.org/10.1161/01.CIR.0000134279.79571.73
2. Dunlay, S. M., Weston, S. A., Redfield, M. M., Killian, J. M., & Roger, V. L. (2008). Anemia and heart failure: a community study. The American journal of medicine, 121(8), 726–732. https://doi.org/10.1016/j.amjmed.2008.03.039
3. Go, A. S., Yang, J., Ackerson, L. M., Lepper, K., Robbins, S., Massie, B. M., & Shlipak, M. G. (2006). Hemoglobin level, chronic kidney disease, and the risks of death and hospitalization in adults with chronic heart failure: the Anemia in Chronic Heart Failure: Outcomes and Resource Utilization (ANCHOR) Study. Circulation, 113(23), 2713–2723. <https://doi.org/10.1161/CIRCULATIONAHA.105.577577>

图形用户界面, 图表, 直方图

Description automatically generated

Figure 1. Partial Likelihood Deviance in Lasso for variable selection.

Notes: A) selecting variables in Cox regression by setting the main exposure as serum creatinine group (normal vs abnormal); b) same approach was applied only replacing the binary serum creatinine group with continuous serum creatinine concentration.

Table 1. Variable selection based on Lasso

|  |  |  |  |
| --- | --- | --- | --- |
|  | Best variable set by Lasso | Add one covariate | Delete one covariate |
| For binary serum creatinine | | | |
| Log Tunning parameter (log λ) | -3.59 | -3.96 | -2.67 |
| Numbers | 5 | 6 | 4 |
| Variables | Age (continuous);  Anemia (binary);  Ejection fraction; (continuous);  Serum creatinine (binary);  Serum sodium (continuous) | + Sex (binary) | -Anemia (binary) |
| For binary serum creatinine | | | |
| Tunning parameter (λ) | 0.028 | 0.021 | 0.058 |
| Numbers | 5 | 6 | 4 |
| Variables | Age (continuous);  Anemia (binary);  Ejection fraction (continuous);  Serum creatinine (continuous);  Serum sodium (continuous) | + Sex (binary) | -Anemia (binary) |

Table 2. Characteristics of the 299 patients at baseline by serum creatinine levela

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Normal serum creatinine (N = 232)** | **Abnormal serum creatinine (N = 67)** | **p-value** |
| Average serum creatinineb (SD) | 1.03 (0.21) | 2.67 (1.60) | < 0.001 |
| Average age (SD) | 59.64 (11.49) | 64.97 (12.41) | 0.0022 |
| Sex |  |  | 0.7651 |
| Male | 149 (64.22%) | 45 (67.16%) |  |
| Female | 83 (35.78%) | 22 (32.84%) |  |
| Smokers, Yes | 78 (33.62%) | 18 (26.87%) | 0.3710 |
| Anemia, Yes | 102 (43.97%) | 27 (40.30%) | 0.6937 |
| Diabetes, Yes | 98 (42.24%) | 27 (40.30%) | 0.8860 |
| Ejection fraction |  |  | 0.0025 |
| ≤ 30 | 61 (26.29%) | 32 (47.76%) |  |
| 31-44 | 102 (43.97%) | 24 (35.82%) |  |
| ≥ 45 | 69 (29.74%) | 11 (16.42%) |  |
| Average plateletsc (SD) | 265270.8 (96762.6) | 256734.7 (101796.5) | 0.5424 |
| Average serum sodiumd (SD) | 137.32 (3.61) | 134.21 (5.89) | <0.001 |

a. Number and proportion are reported for categorical variables. Average and standard deviation (SD) are reported for continuous variables. Chi-squared tests and t-tests were applied to categorical and continuous variables respectively.

b. The unit of serum creatinine is mg/dL.

c. The unit of platelets is kilo platelets/mL.

d. The unit of serum sodium is mEq/L.

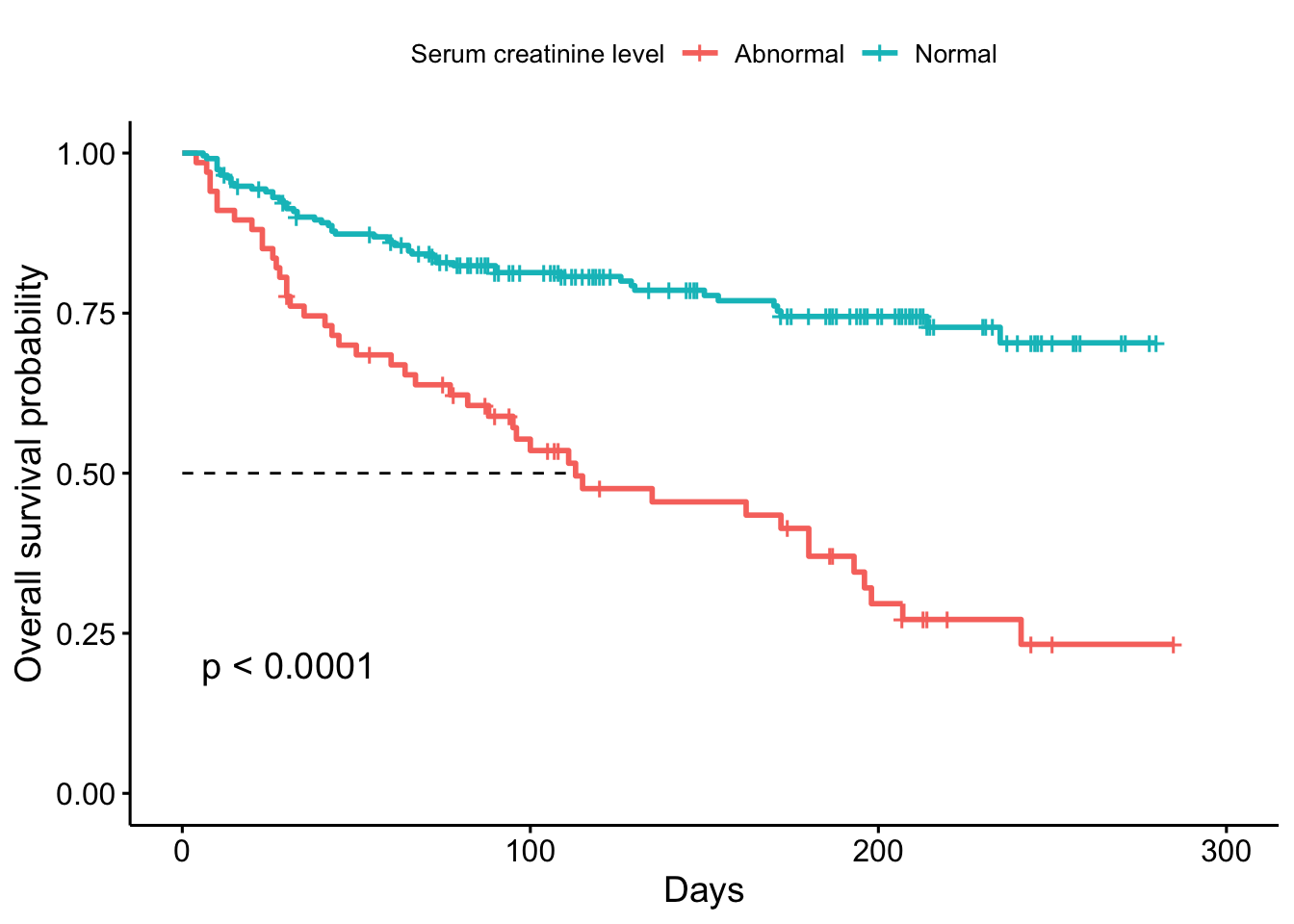


Fig 2. Kaplan-Meier Survival curves for patients with normal (≤1.5 mg/dL) and abnormal serum (> 1.5 mg/dL) creatinine level.

Table 3 Association between serum creatine and mortalities among patients with heart diseases, using Cox regression (N = 299)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Binary serum creatinine | | Continuous serum creatinine | |
|  | Hazard ratio | 95% CI | Hazard ratio | 95% CI |
| Crude modela | 3.39 | 2.26 to 5.08 | 1.34 | 1.20 to 1.49 |
| Lasso selection modelb | 2.27 | 1.45 to 3.55 | 1.38 | 1.19 to 1.59 |
| Fully adjusted modelc | 2.29 | 1.46 to 3.58 | 1.37 | 1.19 to 1.59 |

Notes:

a. In the crude model, only serum creatine (binary or continuous) was included in the model.

b. In the Lasso selection model, age (continuous), anemia (binary), ejection fraction (continuous), serum sodium (continuous) were adjusted for.

c. In the fully adjusted model, sex was further adjusted for on the basis of Lasso selection model.

Table 4. Effect measure modifications between serum creatine and mortalities among patients with heart diseases, using an interaction term in Cox regression (N = 299)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Binary serum creatinine | | Continuous serum creatinine | |
| Interaction with | Log HR for the interaction term | p-value | Log HR for the interaction term | p-value |
| Age | -0.005 | 0.760 | -0.001 | 0.880 |
| Sex | -0.730 | 0.094 | -0.048 | 0.734 |
| Anemia | 0.189 | 0.654 | 0.006 | 0.152 |
| Ejection fraction | 0.040 | 0.064 | 0.006 | 0.152 |
| Serum Sodium | 0.024 | 0.618 | 0.038 | 0.017 |

Notes:

The models were all fully adjusted for age, sex, anemia, ejection fraction, serum sodium.

图表

Description automatically generated

Fig 2. Nonlinear curves between serum creatinine and mortalities among patients (N = 299).

Notes:

We chose serum creatine equal to 0 as the reference group to calculate the estimated HR.