Heart failure(HF) is a serious condition that develops when hearts don’t pump enough blood for the whole body’s needs. Heart failure is often caused by conditions that damage the heart like coronary heart disease, diabetes, high blood pressure.

As a serious condition requiring medical care and treatment, HF is affecting more than 6 million patients and their families in the United States, which brings a high public health cost and burden.[1]

A common complication of heart failure is renal dysfunction. Generally speaking, the reduced cardiac output and the consequently renal under-perfusion is the main pathophysiology cause of renal dysfunction because it leads to low renal blood flow and increased renal venous pressure.[1] 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 heart failure. [2]On the contrary, renal dysfunction may cause heart failure through mechanisms such as inflammatory activation and anemia (caused by a depression of renal erythropoietin production). [3]

The primary exposure, serum creatinine, is a chemical compound that can be filtered out of the blood by healthy kidneys, therefore serum creatinine could be used to calculate the glomerular filtration rate, which is an important indicator of kidney function. [4] Serum creatinine level is tested through venous blood, and has become routine clinical practice because of efficiency and easy procedures. Serum creatinine greater than its normal level (1.5) is an indicator of renal dysfunction.[6]

Based on this biological knowledge, we want to use serum creatinine as a biomarker to explore the relationship between renal dysfunction and mortality rates among patients with heart failure. Some previous studies are also related to this issue. Ahmad et al mainly identified potential risk factors for mortalities among HF patients and reported renal dysfunction (determined by serum creatinine levels) as a key factor contributing to increased risk of mortality. [6] 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. [7] However, these studies concerned less on the modification of other biomarkers such as eject fraction and serum sodium, which are also related to heart failure through diverse mechanisms.[8] Therefore, we aimed to study potential effect modification by other biomarkers and fill the vacancy of existing studies.

1. <https://www.nhlbi.nih.gov/health-topics/heart-failure>

2. Núñez, J., et al., *Early serum creatinine changes and outcomes in patients admitted for acute heart failure: the cardio-renal syndrome revisited.* European Heart Journal. Acute Cardiovascular Care, 2017. **6**(5): p. 430-440.

3. Metra, M., et al., *The role of the kidney in heart failure.* European Heart Journal, 2012. **33**(17): p. 2135-2142.

4. Silverberg, D.S., *The role of erythropoiesis stimulating agents and intravenous (IV) iron in the cardio renal anemia syndrome.* Heart Fail Rev, 2011. **16**(6): p. 609-14.

5. Butler, J., et al., *Renal function, health outcomes, and resource utilization in acute heart failure: a systematic review.* Circ Heart Fail, 2010. **3**(6): p. 726-45.

6. Ahmad, T., et al., *Survival analysis of heart failure patients: A case study.* PLoS One, 2017. **12**(7): p. e0181001.

7. Chicco, D. and G. Jurman, *Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone.* BMC Medical Informatics and Decision Making, 2020. **20**(1): p. 16.

8. Patel, Y.R., et al., *Prognostic Significance of Baseline Serum Sodium in Heart Failure With Preserved Ejection Fraction.* J Am Heart Assoc, 2018. **7**(12).

Data description

The dataset we use comes from UCI Machine Learning Repository. The data was first collected and analyzed by Ahmad et al. [5] They collected data based on medical records of 299 patients with heart failure at NYHA class III and IV, the most two severe HF stages. 13 clinical features are shown in the dataset, including age(years), anemia (boolean), high blood pressure (boolean), creatinine phosphokinase level (mcg/L), diabetes (boolean), ejection fraction (percentage), platelets (kiloplatelets/mL), sex (binary), serum creatinine (mg/dL), serum sodium (mEq/L), smoking (boolean) ,follow-up period (days) and death event (boolean).

In our project, Serum creatinine level was first treated as a continuous variable and then categorized into two different levels (≤ 1.5 mg/dL for the normal level vs, and > 1.5 mg/dL for the abnormal level).

Outlines

Materials:

The study population of this project is heart failure patients who were admitted to the Institute of Cardiology and Allied hospital Faisalabad-Pakistan during April-December in 2015. As described in the data description, their original information and data were collected by Tanvir Ahmad et al at Government College University, Faisalabad, Pakistan. We downloaded this dataset from the UCI Machine Learning Repository and will use it under the same Attribution 4.0 International (CC BY 4.0) copyright. The dataset can be accessed through <https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records.>

Methods:

*Since survival analysis is the main part of our project, the project is not over yet. There are some methods that have not yet been implemented. But in order to maintain writing consistency, we assumed that all methods in this part were used.*

Data cleaning:

We checked potential missing data in the outcome, exposure, and covariates, and there was no missing in this dataset.

Checking the exposure:

We used serum creatinine as the primary exposure, which was provided in a continuous form in the dataset. In this project, we first treated the serum creatinine as a continuous variable and then categorized it into two different levels (≤ 1.5 mg/dL for the normal level vs, and > 1.5 mg/dL for the abnormal level). The sample mean, standard deviation, median, range and the proportion of two different levels were calculated.

Checking the outcomes:

The average person-time until death was calculated in the normal serum creatinine group and the abnormal serum creatinine group. We counted the survival/death status of patients within 30 days of follow-up. This data provided a new outcome for our project and helped us build logistic models. Survival plots were made to visualize the mortality rates in serum creatinine groups. Chi-squared tests were applied to test the difference of Death 30-day in these groups.

Checking other covariates.

In this project, age (continuous), sex (male vs. female), anemia (yes vs. no), diabetes (yes vs. no), ejection fraction (≤ 30, 31-44, and ≥ 45), smoking (yes vs. no), platelets (continuous, kilo platelets/mL), and serum sodium (continuous, mEq/L) were considered as covariates.

The proportions for categorical covariates and mean (standard deviations) for continuous covariates in the normal serum creatinine group and abnormal serum creatinine group were calculated and compared using chi-squared tests and t-tests respectively.

Modeling analysis

First, the simple linear regression was used to evaluate the association between survival time and serum creatinine level in patients who died by the end of the. For the linear regression, we counted the death time for people who died by the end and treat it as a linear outcome. The potential modification effects of other covariates with multivariate linear models. Lasso was used to determine the covariates that should be adjusted in multivariate linear models.

Second, the probability of Death 30-day will be modeled by logistic models. Lasso was used to determine the most appropriate covariates sets.

Third, we combined the incidence of deaths and time at risk among patients with identical covariate patterns. Poisson regression models were employed to model the association between serum creatinine level and incidence rate of deaths.

Last, in our main analysis, we performed survival analysis. A Kaplan-Merrier plot was made for patients stratified by serum creatinine (normal vs. abnormal). Then Cox proportional-hazards model was performed, with the outcome to be the survival time with the event (0 for censored and 1 for death). Appropriate covariates were adjusted for these models.

Model selection

We utilized lasso regressions to determine the optimal covariate sets in the models mentioned above, except the Poisson regression. To justify the selection by the lasso regression, we decided to include and delete a covariate based on the covariates automatically calculated by the software. Thus, in each model we had three sets of covariates. We performed models using these three covariates sets and compared the model performance based on AIC.

Subgroup analysis for potential effect modification.

To check whether effect modification exists, we included an interaction term between serum creatinine and selected potential covariates in Cox models. The potential effect modifiers were determined by p-value of the interaction term as well as variance-deviance analysis.