***1. Background***

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.[1]

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.[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 HF. [2]On the contrary, renal dysfunction may cause HF through mechanisms such as inflammatory activation and anemia (caused by a depression of renal erythropoietin production). [3] Serval study has demonstrated that renal dysfunction can lead to higher mortality rates among patients with cardiovascular diseases.

Serum creatinine is an important and commonly-used biomarker to indicator of the presence kidney dysfunction. [4] 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.[6]

In this study, we aimed to employ serum creatinine as an indicator for renal dysfunction and explore its association with mortality rates among patients with HF. Ahmad et al. [6] and Chicco et al. have demonstrated the harmful associations in their previous studies. [7] However, 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] Therefore, this study aimed to fill the current research gap by studying potential effect modifications by other important factors and fill the vacancy of existing studies.

**References**

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).

***2. Method***

**2.1. Material sources and data description**

The dataset was retrieved from UCI Machine Learning Repository. It was first collected and analyzed by Ahmad et al. [5] based on medical records of 299 patients with HF at NYHA class III and IV, the most two severe HF stages. 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.   
A total of 13 clinical features are 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).

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, > 1.5 mg/dL for the abnormal level).

***2.2. Exposure***

The serum creatinine was the exposure of interest in this project. It originally assessed as a continuous variable 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.

***2.3. Outcome***

Checking the outcomes:

In this survival analysis project, the primary outcome is the time to event (i.e., survival time and the end status) for each patient. To make our outcome compatible with the linear regression, logistic regression, and Poisson regression framework, we generated several secondary outcomes and did exploratory analysis. First, among patients who died by the end of the study, their survival time until death was calculate. Second, the status of each patient by the 30 days from the start of the follow-up was identified and transformed into a binary variable (alive vs., deceased). Last, we collapsed the individual-based dataset into group-based dataset and calculated the cases of death and the total time at risk. The collapse was conducted by groups with the identical covariate patterns

***2.4. 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 potential covariates.

***2.5. Statistical analysis***

We first checked potential missing data in this dataset. After careful check, there were no missing values in exposure, outcome, or covariates. Then descriptive analysis was conducted exploring the baseline characteristics for patients with normal and abnormal serum creatine. T-tests and Chi-squared tests were conducted for continuous and categorical variables respectively.

In the modeling stage, 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 study period. Second, the probability of deaths by the 30 days was modeled using logistic models. 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.

***2.6. Model selection***

Since the dataset is of relatively small size (289 samples in total), 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.

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 by deviance analysis.