1. Group 23 (Group LGL)

Jinglun Li

Fuyu Guo

Xinhao Li

2. Background knowledge, feedback, and eyes on the field:

***a.***

In this study, we aim to explore the relationship between renal dysfunction and mortality rates among patients with heart failure. The question was firstly studied by Ahmad et al utilizing survival analysis. 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. 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. 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.

After reviewing existing literature, we aimed to use a multivariable adjusted Cox model as the main model as it is commonly used in this research area. However, previous studies reported less on the modification of other biomarkers such as eject fraction. Potential effect modification by other biomarkers may reveal heterogenous effects of renal dysfunction among different subgroups. Thus, we aim to fill this gap in the current project.

***b.***

We appreciate very much the teaching stuff and out peers’ comments. These comments point out several very important questions which we will incorporate in our analysis. We aggregate the comments and show our response and analysis plan below (***for question i and ii***).

* In the primary question, you want to explore the association between serum creatine and mortality rates for patients with heart failure. I think the question is not specific enough. What kind of association you expect they should have? Which aspects you aim to explore?

Response: It is a very helpful question. We will make it more clear that there are three sperate models and the association definitions are different: 1) The association between the expected survival time and serum creatine: the association is on average how many days will increase if the patient’s serum creatinine level decreased by 10% in a linear model; 2) the association between death by 30 days and serum creatine: the association is the average odds ratio of death by 30 days if the patient’s serum creatinine level decreased by 10% in a logistic model; and 3) the association between mortality rates and serum creatine: the association is the average hazard ratio if the patient’s serum creatinine level decreased by 10% in a Cox model. In model fitting part, how about add a model selection process that is to fit different models that we mentioned in class, can compare which models fit better?

* The dataset is small. Overfitting is easier to happen. Simper model may be better.

Response: We appreciate this comment. As a solution to limited sample size, we aim to utilized a lasso regression in the linear model to help us identify key covariates influencing patients’ morality. Based on the sparse results of the lasso regression, we will decide which covariate to be included in our main analysis. In this study, we prefer lasso regression as a prediction selection method to stepwise regression, since stepwise regression is somehow time-consuming and may get inconsistent predictor sets when utilizing different step directions. To determine the optimal penalty parameter in lasso, , we will utilize a cross-validation process, or let the function in nlmet package to automatically decide the optimal parameter. After getting the most suitable covariate sets, we will delete and add a covariate according to the lasso results, and assess the goodness-of-fit in the model with these three different covariate sets. Detailed analysis please see the methods section below.

* My only advice is about your data set. Even though your dataset has only 299 cases, the dataset seems clean and complete, so after you check if there is any missing data and if it’s not too much, maybe using complete cases with all data entries would be both easier and more accurate, Or at least the variable you are interested in is complete. If the missing data reduce the number of cases a lot, the mice function in R could also be helpful.

Response: We appreciate this comment. Actually, after checking the dataset, luckily, we found there were no missing values. Thus, we think it is no need to make imputations. Thanks for the comment, and thanks for sharing this important package *MICE.*

***iii.***

As the research problem has been widely discussed in previous medicine literature, though the quantitative results are relatively limited, its clinical meaning is relatively clear. Given the limited time of this project and our limited resources, we admit we fail to contact a clinician expert during the past 1 month. We will keep reading related literature and try our best to seek domain experts’ help.

3. Analysis Plan: