Project Proposal – MA678

Annika Balk

Motivations:

The data set aligns well with my goals for PhD research. I would like to understand cardiovascular and renal diseases better. After I finish my PhD and I am an expert in biostatistics I will pursue a career in pharmaceuticals. I would like to use this project to practice using advanced statistical techniques, conducting effective literature review, understanding the data well and becoming better at independent research. My aim is to have over ten sources that will help provide understanding and context to my analysis.

Data Source and Data Background:

The data I found is from Harvard Dataverse and uses data that was published in the paper titled, “The Relationship of Fetuin-A with Coronary Calcification, Carotid Atherosclerosis, and Mortality Risk in Non-Dialysis Chronic Kidney Disease”. It contains 33 variables and has around 160 patients. The link will be provided at the bottom for access to the dataset.

Research Question:

I am not interested in the same relationships the original paper was published about. According to a literature review titled, “Cardiovascular Calcification Heterogeneity in Chronic Kidney Disease” it is stated that there needs to be additional studies on the, “temporal association between declining renal function, the appearance and growth of calcification across various cardiovascular beds, the presence of other confounding risk factors, and cardiovascular morbidity.” The data set that I have contains some critical information that may help understand these associations. I do not have information on specific cardiovascular beds but I do have Coronary Artery Calcification (CAC) scores, comorbidities, and kidney function measures. With this I would like to come up with a model that estimates eGFR (estimated glomerular filtration rate) a measure of kidney function. In my data everyone has some level of chronic kidney disease. I would like to model this using the CAC scores, comorbidities and other attributes of each patient such as sex, age, and BMI. Further research will be needed to identify how to classify patients with higher than normal values in mineral such as calcium that may predispose them to more severe CAC.

Timeline:

EDA – Complete by the 17th

Data Processing – Complete by the 22nd

Modeling – Complete by the 29th

Write-up – Complete by Dec 6th

Data Link:

<https://dataverse.harvard.edu/file.xhtml?fileId=10191788&version=1.1&toolType=PREVIEW>

**This page it to help me with analyses:**

Some key sentences from papers:

Paper/Quote:

The Signaling Pathway of TNF Receptors: Linking Animal Models of Renal Disease to Human CKD

“The etiology of CKD depends on the setting, with diabetes and hypertension being the two major causes of kidney injury in developed countries”

“simply TNF, is a pleiotropic cytokine that can mediate the inflammatory response, regulate immune function by promoting immune cells activation and recruitment, and may trigger cell proliferation, differentiation, apoptosis, and necroptosis [31]”

“higher values of TNF-α, have been associated with autoimmune diseases, such as rheumatoid arthritis, multiple sclerosis, inflammatory bowel diseases [38,39], and chronic inflammatory disease states, such as sepsis, CKD, obesity, and diabetes”

Paper/Quotes: