Chronic Kidney Disease

By Elias Fedai

Bellevue University

Chronic kidney disease is a disease state that normally occurs over a long period of time. It can at times give no signs or symptoms thus given the name 'the silent killer', and is usually detected through blood work. The data I am using originated from Dr.P.Soundarapandian at Apollo Hospitals in India. Data set was created by L. Jerlin Rubini and guided by Dr.P.Eswaran of Alagappa University. Date range in which data was collected was not provided, but the data set was donated July 3, 2015. This data set consists of 400 instances and 25 total attributes. The data was attained in a two month period. The data set appears to be an original data set. In my analysis the questions that I was seeking to answer were; which parameter/s if any correlate well with the diagnosis of chronic kidney disease? Are there any unique relationships among the parameters? Are there any non-significant parameters? Can these variables be used to predict chronic kidney disease?

When I initially took on this project I began with fourteen variables and due to time constraints I needed to reduce the number down to six. The six variables that were chosen where; specific gravity, albumin, hemoglobin, red blood cell count, blood pressure, and diagnosis (chronic kidney disease or no chronic kidney disease). After completing my exploratory data analysis on the data, I came to realize many things. Overall, the model that I used (logistic regression) based on the chosen variables displayed to be significant thus not allowing me to reject the null hypothesis. I did feel like I missed one thing in my analysis in that the two variables (RBC count, hemoglobin) strongly correlated. I wish I had some more time so that I could have removed one of the variables out of the analysis and see what the outcome may have been, kind of a compare and contrast between the two. There were some variables that were not in the dataset that I feel may have been of interest for this research such as serum phosphorus. Also, extending the data collection time to more than two months and also attaining more data possible from other populations to increase diversity in the data (possible different countries worldwide). There were some assumptions in the data that I did not agree with. The biggest being that it was never stated exactly what source the results were from (i.e. peripheral blood, urine, body fluid, etc...) this is significant in that the normal ranges are different for each source and also the methodology of data result was not noted. Also, the blood pressure variable never stated if it was systolic or diastolic (fortunately did not correlate well with any other variable). This caused some challenges for me because based on this lack of information, assumptions had to be made which can truly cause impact on how the results are interpreted, which led me to not come to a solid conclusion of the analysis even though the model displayed a strong predictive outcome.

References:

1) Tonelli, M., Wiebe, N., Culleton, B., House, A., Rabbat, C., Fok, M., ... & Garg, A. X. (2006). Chronic kidney disease and mortality risk: a systematic review. *Journal of the American Society of Nephrology*, *17*(7), 2034-2047.

2) Levey, A. S., & Coresh, J. (2012). Chronic kidney disease. *The lancet*, *379*(9811), 165-180.
3) Jha, V., Garcia-Garcia, G., Iseki, K., Li, Z., Naicker, S., Plattner, B., ... & Yang, C. W. (2013). Chronic kidney disease: global dimension and perspectives. *The Lancet*, *382*(9888), 260-272.
4) Go, A. S., Chertow, G. M., Fan, D., McCulloch, C. E., & Hsu, C. Y. (2004). Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *New England Journal of Medicine*, *351*(13), 1296-1305.
5) Kestenbaum, B., Sampson, J. N., Rudser, K. D., Patterson, D. J., Seliger, S. L., Young, B., ... & Andress, D. L. (2005). Serum phosphate levels and mortality risk among people with chronic kidney disease. *Journal of the American Society of Nephrology*, *16*(2), 520-528.