NRSG 741: Predicting 30-day Inpatient Readmission in Children With Sickle Cell Disease

Alexander Maillis

2/2/2020

Basic Information

Project title: 30-day inpatient readmission prediction among a cohort of children with sickle cell anemia.

Name: Alexander Maillis.

Email: alexander.nicholas.maillis@emory.edu

Overview and Motivation

This particular topic was mostly inspired by the work I currently do for Children's Healthcare of Atlanta. At the moment I work for a sickle cell epidemiology team that is hoping to implement more advanced analytical techniques in effort to change the way we can investigate the epidemiology of sickle cell anemia. Within the past month Dr. Peter Lane, the director of the sickle cell program, has brought up the idea of looking into 30-day readmissions. Understanding how important 30-day readmissions are as a parameter of care for hospitals, combined with how arduous these constant trips can be on the patients and their families, this topic seemed like one that would be interesting to investigate further. The topic of 30-day readmissions within research have become more and more popular over the years, but yet the sickle cell disease population has only a limited amount of papers on this topic despite their typically high utilization of hospital services compared to other diseases considered to be common. My personal interests in this paper are focused on the hope that we can form some sort of predictive model for these patients that may be able to predict the probability of, to some degree of accuracy, that a patient is readmitted to an inpatient service at Children's Healthcare of Atlanta within 30-days after their index admission.

Primary Objectives

The primary focus of this paper is to attempt and quantify the likelihood that a patient, between 9 months and 18 years old, will be readmitted to an inpatient service at Children's within 30-days of discharge from an inpatient admission. I hope to use prior research within the field of hematology to identify factors that have been commonly associated with 30-day readmissions to then build a score that can capture the percentage chance of that readmission. Since I have access to nearly 10 years worth of utilization data through the Children's Healthcare Sickle Cell Registry, I am hopeful that this dataset will provide adequate information for risk prediction.

The outcome of interest will be broken into three categories. **No 30-day readmission**, **Planned 30-day readmission**, and **Unplanned 30-day readmission** will be our three outcomes of interest. Since we are capable of using Qlikview to differentiate planned vs. unplanned readmissions, it may be useful to use the prediction score in the scope of an unplanned readmission.

Data Source

The data used in this study will be provided by the Children's Healthcare of Atlanta Sickle Cell Disease Registry and Utilization databases. These databases are maintained by myself and Kristina Lai, under Dr. Peter Lane, and track the hospital visits, days, diagnoses, and medical treatments that all confirmed patients with Sickle Cell disease undergoe while seen at a Children's facility. These databases have been maintained since 2010, allowing me to compile a dataset with approximately 6,000 unique sickle cell patients totalling just over 115,000 hospital visits. The databases mentioned currently have IRB approval and will be

de-identified for the purpose of this project. Additional data include genotype of disease, primary hospital of use, BMT date (Y/N), deceased (Y/N), and patient classification (IP/OP/ED). Due to the confidential nature of this data no link to the dataset will be provided.

Data Wrangling

I predict that a moderate amount of data wrangling will be expected. At the moment, there are a few variables that exist as character variables in our file but are more easily managed when in a numerical/categorized format (i.e. gender and genotype of disease). Additionally, my preliminary idea is that certain diagnoses may be more useful for predicting readmissions if they are treated as a counted score, rather than an individual occurence; An example of this is Acute Chest Syndrome (ACS) which is a pneumonia-like disease specific to the sickle cell disease population. Some variables, like baseline hemoglobin, will require computing an average value of x amount of events. At the moment this is the wrangling I for see to be doing but, depending on how these variables contribute to the project, this could change.

Exploratory Analysis

During exploratory analysis I think it will be useful to view data through a couple different methods. For non-binary factors such as baseline hemoglobin, admissions within the last 12 months, length of stay of index admission, and number of medications at discharge, it will be useful to see how the readmitted and individuals without readmissions compare via boxplots. Other factors such as genotype of sickle cell disease and history of recurrent ACS may be more easily visualized within a table that notes percentages of each group.

Analysis

Analysis of this question will require several steps. Prior research on a similar topic (Donze et al) was successful in building a risk prediction score using univariable logistic regression to identify factors of interest, backwards elimination of a multivariable model to remove variables not statistically significant, and finally a regression-coefficient based scoring method to construct the prediction score off the final multivariable model. Additionally, it is suggested that the data is randomly assigned to either a derivation set or a validation set, and that we use a C statistic and goodness-of-fit test to ensure that the two sets are appropriately able to differentiate between the two outcomes of interest. An idea that this study did not use, would be to stratify this prediction among 30-day readmissions, 7-day readmission, and 72-hour readmissions in effort to identify factors that may more strongly influence the odds of unplanned readmission.

Tentative Schedule:

2/5: Submit proposal for project idea

2/19: Finish constructing new variables and cleaning data

2/26: Finish univariable regression models and have variables of interest for score construction

3/11: Finish constructing multivariable model, conduct backwards elimination, **finalize variables within risk score.**

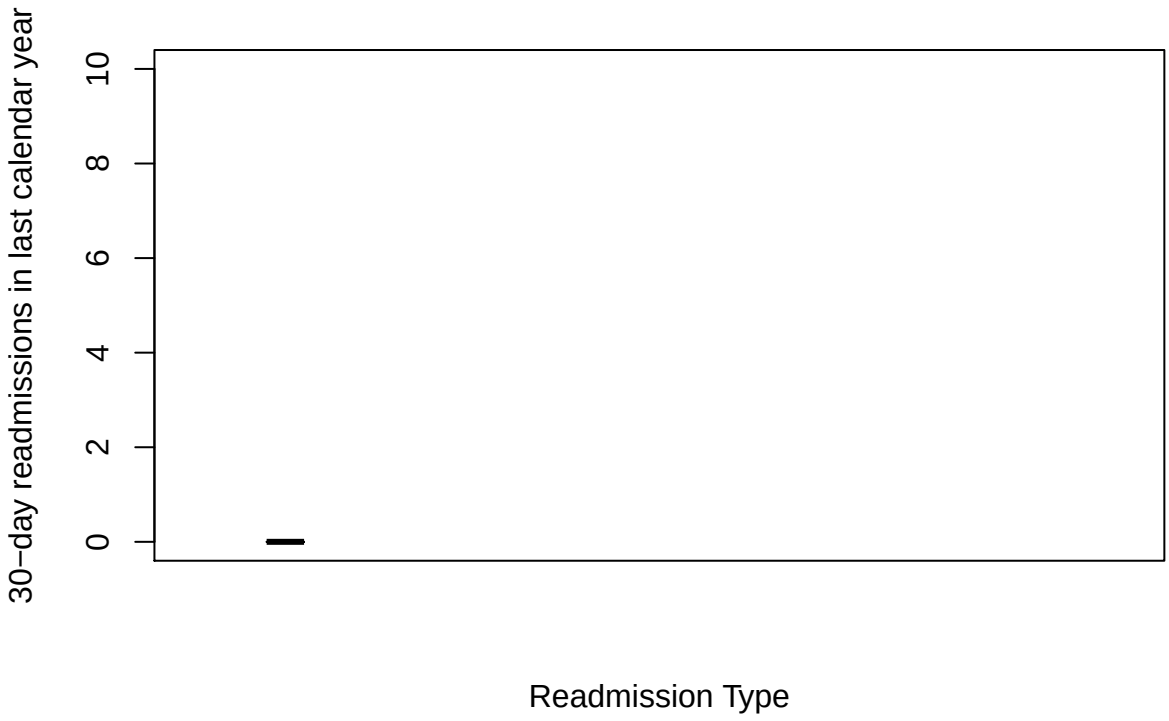
3/25: Finish prediction score construction, begin process of validating dataset.

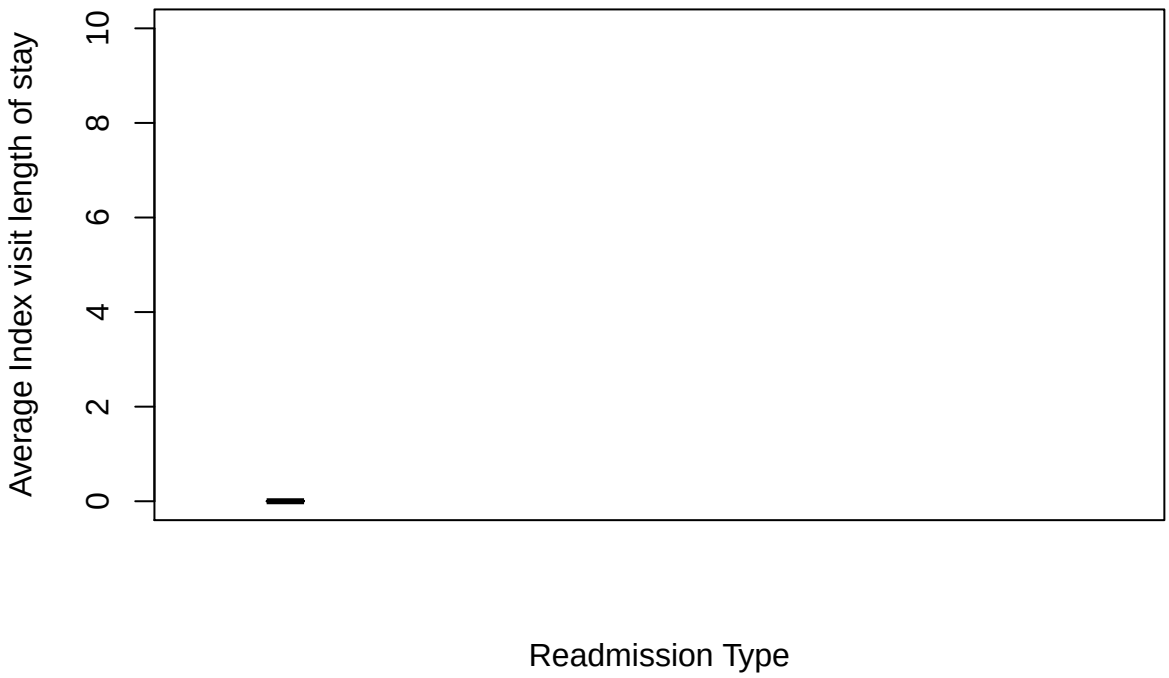
3/27: Submit a working prototype.

4/15: Finish validating derivation set using validation set. Construct figures and tables.

4/22: Finalize manuscript.

Using baseline statistics to compare variables of interest




Average Index visit length of stay
0
2
4
6
8
10
Readmission Type