A Predictive Modeling of Tracheostomy Readmissions

Aabha Latkar BA, Febriany Lete SST, Maysen Pagán BA

Master of Science in Statistical Practice, Boston University.

Abstract

Hospital readmissions following the completion of a procedure poses significant challenges for both patients and hospitals. Being readmitted could increase the risk of complications for patients and introduce states of distress. For hospitals, readmissions present a strain on resources and their reputation. This project aims to develop a model that can predict whether a patient will be readmitted within 30 days of being discharged. After thorough data preprocessing and feature extraction, three models were trained on patient data from 2018 and tested on patient data from 2019 to determine predictive performance: a null model, a support vector machine (SVM), and a random forest (RF). Using Matthew's Correlation Coefficient to compare models, the HERE model had the best performance with a coefficient of HERE while the null model has a coefficient of HERE, and the HERE model has a coefficient of HERE.

1 Introduction

In the medical field of Otolaryngology, preventing hospital readmissions following procedures such as tracheostomies, total laryngectomies, or mastoidectomies is significant both medically for patients as well as financially for healthcare institutions. Medically, avoiding readmissions can benefit patients' well-being as it reduces the possible distress and suffering experienced from complications from new or returning medical conditions. Financially, preventing readmissions is crucial for hospitals who are paid by capitation. Capitation is a payment system that pays hospitals a fixed amount per patient for a prescribed period, therefore incentivizing hospitals to conduct less procedures and treat patients as efficiently as possible. As a result, hospitals paid by capitation incur the costs that are associated with providing care to patients who are readmitted. Knowing if a patient might be at higher risk of a readmission

would allow doctors to increase the effectiveness of their initial interventions and promote a smoother recovery process while maintaining their reputation and quality of care. Therefore, developing predictive models that can predict whether a patient will be readmitted is essential for ensuring the efficiency of healthcare.

This project aims to build a model that predicts if a patient who underwent a tracheostomy procedure is going to be readmitted within 30 days of being discharged from the hospital. For those patients who are readmitted within 30 days, this project also analyzes the number of days until they will be readmitted as well as the most common diagnoses that the patients will be readmitted with.

2 Data Preprocessing

The data for this project was provided to us from the Healthcare Cost and Utilization Project (HCUP) Nationwide Readmissions Database (NRD). This database provides information on admissions and discharges for patients with and without repeat hospital visits within a given year. To address the research questions of this project, we used data from the years of 2018 and 2019, each of which had three data sets. For each year, there was a `core` data set which contained main information such as the age and sex of the patient as well as the diagnoses and procedures the patient received for this given admission. The `severity` data set provides information on the severity of the condition for a given admission and the `hospital` data set contains characteristics of the hospital the patient was admitted to such as bed size.

2.1 Empty Diagnoses and Procedures

The `core` data set contains 40 ICD-10-CM diagnosis code columns representing 40 possible diagnoses a patient could receive at one admission. Similarly, the data includes 25 ICD-10-PCS procedure code columns representing 25 possible procedures a patient could receive at one admission. One patient may not have 40 total diagnoses or 25 total procedures and as a result, many of the cells in these columns are empty. The first step in preparing the data for analysis was replacing these empty cells with `NA`s which will make future filtering and one-hot-encoding steps easier.

2.2 Creating Binary Readmitted Response Column

The focus of this project is predicting readmissions for tracheostomy procedures. Therefore, our next step was to filter the `core` data set to include only those patients who received a tracheostomy during their admission. The ICD-10-PCS codes for tracheostomies all begin with "0B11" with other numbers and letters following for different tracheostomy approaches such as an open approach or percutaneous approach. We applied the filter function to the 25 procedure columns to obtain all patients who had an ICD-10-PCS code that began with "0B11" in any one of the 25 possible columns. Extracting the unique IDs or Visit Links from this filtered set and filtering `core` to get patients only with those Visit Links allowed us to create a new data frame that contained all patients who were admitted for a tracheostomy and all other

admissions by that patient. A binary column was then added to this new data frame which was a 1 if the patient received a tracheostomy at the corresponding visit and 0 otherwise. For those tracheostomy admissions, we added the length of stay (LOS) variable to the days to event variable which gives us a sequence of numbers representing the "date" the patient was discharged. This data frame contains all patient admissions both before and after they received the tracheostomy. As a result, for each patient, we then filtered the data frame again to include only the tracheostomy admission and all admissions that occurred after that visit.

This new data frame now allows us to create our binary response variable that is 1 if the patient was readmitted to a hospital within 30 days of being discharged from a tracheostomy procedure. If a patient had no admissions following their tracheostomy admission, the readmitted response variable was 0. If subtracting the days to event value of the next admission after the tracheostomy from the days to event value of the tracheostomy admission results in a value less than or equal to 30, the readmitted response variable was 1. If the difference of the two days to event values was greater than 30, the readmitted response variable was 0.

We then removed all rows that represented hospital admissions after the tracheostomy to obtain a data frame where each row represented a unique patient's tracheostomy admission along with the binary response variable indicating if they were readmitted within 30 days of the procedure.

2.3 Right Censoring

Right censored data is used to describe data where subjects leave the study before an event occurring or the study ends before the event has occurred. In the context of this project, right censoring can occur if a patient received a tracheostomy in December of one year and was readmitted in January of the following year. The data from the NRD is constructed in such a way that it is difficult to identify patients. The patient IDs or Visit Links change from one year to the next for the same patient. As a result, if a patient were to receive a tracheostomy in December of 2018, we would be unable to determine and therefore predict if the patient was readmitted within 30 days since it is unknown if they were readmitted in January of 2019. To solve this issue, we make our window of inclusion for patients between January and November of a given year by filtering out those patients who were admitted for a tracheostomy in December.

2.4 One-Hot-Encoding

When creating the predictive model for readmissions, we wish to conduct comorbidity analysis which will allow us to include the presence of certain diagnoses as predictor variables. For that reason, we will need to one-hot-encode all ICD-10-CM diagnosis codes. For each unique diagnosis code in the 40 diagnosis columns, we added a column for that code which equaled 1 if the patient received the diagnosis on their tracheostomy admission and equaled 0 otherwise.

2.5 Joining Data Sets

The last step in preparing the data for modeling and analysis was to join the three data sets. Each data set had a NRD record identifier variable which gave a unique code to each hospital admission. Therefore, to our cleaned `core` data set, we joined the `severity` and `hospital` data sets by the record identifier.

3 Exploratory Data Analysis

4 Methods and Analysis

5 Conclusion

Acknowledgments. Acknowledgments are not compulsory. Where included they should be brief. Grant or contribution numbers may be acknowledged.

Please refer to Journal-level guidance for any specific requirements.

Appendix A

An appendix contains supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem or it is information that is too cumbersome to be included in the body of the paper.