Preventing Hospital Readmission in Neonates

Ashwati Nair

HI 2454 Final Project

University of Pittsburgh

17 April 2024

Introduction

Hospital readmission rates are a key indicator of the quality of care a healthcare organization provides. They can also point to what comorbidities require a higher level of intervention and resources. Most importantly, increased hospital readmission rates are associated with poor quality of life and prognosis (Morrison et al, 2022). This is especially the case for neonatal patients. Newborns are one of the most fragile patients, and complications at birth can develop into chronic conditions. Repeat hospital admissions can be detrimental to children's development, especially for those with chronic conditions. It is also an emotional and financial strain on parents who must take time off work and pay for their child's care.

Neonatal comorbidities are one of the top reasons for readmissions. According to a systemic review done in 2023, newborns who either had a medical condition at birth or developed medical conditions following their birth were associated with an increased risk of 31-day unplanned readmission. These conditions include congenital abnormalities, chronic conditions, and surgical intervention (Della et al). A study in the 2022 issue of *Clinics* journal found that the leading cause of readmission is respiratory tract infections, followed by jaundice and feeding difficulties (Kardum et al, 2022). A review conducted in 2020 found that prolonged hospitalization complicated by pulmonary complications, such as bronchopulmonary dysplasia was a significant risk factor. This same review also found that racial disparities among preterm infants and lack of private insurance access further increased the chance of an infant being readmitted to the hospital (Hannan et al, 2020).

Problem Description and Data

Thus, it is important to determine which diagnoses have an increased risk of hospital readmission for neonatal patients, as well as other potential socioeconomic and demographic factors that may contribute. This includes acquiring data on diagnoses associated with frequent newborn readmissions and demographic data. Establishing a relationship between these factors and hospital readmissions is the first step in coming up with interventions and education for parents. Proactive measures will reduce the risk of readmission.

Data was acquired from two datasets. The first dataset is the "Pediatric All-Payer Readmissions in Massachusetts". It contains diagnostic, demographic and insurance data of pediatric patients of Massachusetts from the years 2017 to 2022. This data was used to explore how many readmissions are the result of neonatal diagnoses. The second data source is the "Screening Disorders by Race/Ethnicity" table from the California Department of Public Health's Newborn Screening Disorders dataset. This table has total counts of positive screenings of common neonatal diagnoses that are further broken down by race and ethnicity. It was used to see if disorders are more common among certain ethnicities, as well as determine which disorders occur most frequently.

Methods

For both datasets, Python code in the Google Colab environment was used to perform preprocessing, data visualization, and analysis. Pandas, matplotlib, seaborn and sklearn.kit were utilized for these tasks.

Pediatric Readmissions

The "Top Diagnoses" sheet of this dataset was the focus of this project and was assigned as the data frame. As stated above, this data was used to determine what the most common

causes of readmissions for pediatric patients are and where neonatal disorders stand among those. The independent variables being evaluated were common pediatric diagnoses such as Neonate DRGs, Bronchiolitis and RSV pneumonia, and seizure. The dependent variable is the number of readmissions for each diagnosis. Using seaborn, a bar plot was made displaying the diagnoses along the x axis and number of discharges as the y axis.

Due to the mixture of categorical and numerical data, linear and logistic regression analysis, Naive Bayes, and Decision tree models were built. The data was prepared by separating the feature (X) and target (y) variables. One-hot was used to encode the features. The data was then split into training and testing sets and used to train the models. Evaluation of the models were done using classification reports for the logistic regression, Naive Bayes, and Decision Tree models. For the linear regression model, the mean squared error, p score and R^2 were used.

Screening Disorders

The "Screening Disorders" data will be used to determine which specific neonatal disorders have the highest risk for repeat hospitalization, as well as if race plays a role in readmissions. For this dataset the independent variables are Race/Ethnicity and Disorder Type. The dependent variable is the Disorder Count, which is the total number of positive screenings for each disorder. Again, a bar plot was created using seaborn. In this case, each disorder type was sub-divided based on the race/ethnicity categories along the x-axis. Disorder count was along the y-axis.

Since this data is also categorical and numerical, linear regression, Naive Bayes, Decision Tree and Support Vector Machine models were built to evaluate the correlation between race, disorder type, and number of positive screenings. As with the "Pediatric Readmissions" dataset,

the feature (X_cols) and target (y_cols) variables were defined. Categorical variables were encoded using LabelEncoder. The dataset was also split into training and testing sets. Again, the linear regression model was evaluated using the mean squared error and R^2. The Naive Bayes, Decision Tree, and SVM models were evaluated using their individual classification reports.

Results and Discussion

Pediatric Readmissions

The bar graph showed that Neonate DRGs had the highest number of readmissions. In fact, it was the only category where readmission amounts were in the thousands, while other diagnoses were below 500.

For the analysis models, only the decision tree model seemingly performed well in predicting readmissions. It's accuracy, precision, recall, and F1 scores of 1.00. The linear regression model had a high mean squared error and a negative R^2 value, indicating poor fit and large error between actual and predicted values. The model also could not produce an F statistic, coefficients, or a p value for the data. For both the logistic regression and Naive Bayes models, overall accuracy was 40% and precision, recall, and F1 scores were all zero. This also indicates poor performance.

There are a few possible reasons as to why the last three models performed this way. One possibility is that the decision tree was able to handle the complexity of the combination of categorical and numerical data. It appears that even with encoding features, the other models had a hard time establishing a linear relationship between the independent and dependent variables. Another reason is the quality and quantity of the data itself. A small dataset, such as this one, would have caused overfitting and the models to perform poorly.

The decision tree model did demonstrate that certain diagnoses carried a higher risk of readmissions compared to others. This, combined with the results of the bar plot show that neonate DRGs are a major chunk of readmissions in pediatric hospitals. It is important to note that this data comes from Massachusetts. Causes of readmissions will not be the same in other states or countries. Nevertheless, the results highlight the need to investigate which disorders lead to newborns being hospitalized.

Screening Disorders

The bar plot for the "Screening Disorders" data set showed some interesting patterns. The disorder with the highest number of positive screenings regardless of race was hypothyroidism, with about 2,477 total positive screenings. 1,570 of these screenings are from Hispanic newborns. Sickle cell disease, which is the second highest screened disorder, has been positively screened disproportionately amongst Black newborns, with about 809 out of 910 screenings. The third most screened disorder was cystic fibrosis.

The analysis models for the "Screening Disorders" dataset had varying results. The linear regression model had a high mean squared error and negative R^2, which again suggests a poor fit to the data. The p-values were high for the coefficients and F statistic, ranging from 0.498 to 0.834. This suggests that race/ethnicity and disorder type are not statistically significant for predicting positive screenings. The Naive Bayes model had 20% accuracy and scores of zero for the F1, recall and precision metrics, again indicating poor performance. The SVM also had very poor performance, with 7% accuracy and very low F1, recall and precision scores. Again, the Decision Tree model performed the best of them all, with perfect scores across the board.

The same issues that were present with the "Pediatric Readmissions" data occurred with the "Screening Disorders" models. Data size and complexity were factors in the poor performance of the models. Like with the previous scenario, they would have benefitted overall from more training and testing data and model refinement.

Like with the "Pediatric Readmissions" data, the bar plot and decision tree model for "Screening Disorders" does show that there is a relationship between race, disorder type and positive screenings. This is important, as early diagnosis of these conditions is key for the newborns, their parents, and providers. Early diagnosis, intervention, and education is important to prevent repeat hospital admissions. The results also highlight that certain disorders are more common among specific races and ethnicities. This is important for public health tracking and research. Again, this information is vital for developing treatments and interventions for babies and parents. It is important to note that this data is from California. Thus, data will vary between states and countries. Regardless of which disorders are prevalent among which populations, evidence based, targeted healthcare is key for preventing hospital readmissions.

Overall Discussion

The biggest issues with this analysis were small, yet complex data samples that made it difficult for models to establish correlation and be trained on unfamiliar data. To fix this, more diverse data, preferably with larger data samples, must be procured to train the models. As the models are trained with new data, they would be refined as issues arise. A larger pool of data would allow the models to better establish correlation between race/ethnicity, neonatal disorders and increased hospital readmissions.

Even though most of the models were not able to establish correlation between the independent and dependent variables, neonatal disorders clearly carry an increased risk of repeat hospital admissions for babies. It is important to continue using data science techniques to demonstrate this relationship. It is also vital to note that non-neonatal specific illnesses, such as bronchiolitis/RSV and seizures can lead to hospitalization for newborns as well. Thus, these diagnoses should also be factored in when performing further analyses. Besides race and ethnicity, other social determinants such as education level of parents and economic status should be included to see how they impact risk. This will also inform healthcare providers on how to best support their patients. Ultimately, these considerations will improve evidence-based practice for neonatal care.

Conclusion

Hospital readmission rates are crucial for gauging healthcare quality, particularly in neonatal care where newborns face heightened vulnerability. Complications at birth and chronic conditions necessitate repeat hospitalizations that strain both families emotionally and financially. Comorbidities such as congenital abnormalities and surgical interventions significantly elevate the risk of unplanned readmissions within 31 days (about 1 month). Besides chronic illnesses, acute conditions such as respiratory tract infections, jaundice, and feeding difficulties are primary causes of readmission, with prolonged hospitalization and racial disparities exacerbating risks. To mitigate these challenges, it's imperative to identify high-risk diagnoses and demographic factors contributing to readmissions, enabling targeted interventions and parent education to reduce recurrence.

For this project, two datasets from two states in the United States were transformed using data analysis techniques to determine if neonatal disorders were associated with increased risk of

readmission. Data visualizations based on data type was performed. Linear regression, logistic regression, Naive Bayes, Support Vector Machine and Decision Tree models were built and analyzed on their performance. While the models had poor to mixed results, this project was a productive first step in exploring how care for acutely and chronically ill newborns can be improved. This project can be improved by acquiring larger samples of high-quality data from around the world. Different analysis models should be used and fine-tuned based on the data at hand.

Health care for newborns, the most vulnerable members of our society, should be of the highest priority and augmented via extensive study and evidence-based practice. Doing so will improve morbidity and mortality rates for the population overall. A commitment to improving quality of life from the beginning of it is a positive and productive goal for all.

Link to OneDrive Folder: Nair2454 Final Project

References

Della, P.R., Huang, H., Roberts, P.A. et al. Risk factors associated with 31-day unplanned hospital readmission in newborns: a systematic review. Eur J Pediatr 182, 1469–1482 (2023). https://doi.org/10.1007/s00431-023-04819-2

Hannan, K. E., Hwang, S. S., & Bourque, S. L. (2020). Readmissions among NICU graduates: Who, when and why? Seminars in Perinatology, 44(4), 151245. https://doi.org/10.1016/j.semperi.2020.151245

Kardum, D., Serdarušić, I., Biljan, B., Šantić, K., & Živković, V. (2022). Readmission of late preterm and term neonates in the neonatal period. Clinics, 77, 100005. https://doi.org/10.1016/j.clinsp.2022.100005

Morrison, B., Lim, E., Jun Ahn, H., & Chen, J. J. (2022). Factors Related to Pediatric Readmissions of Four Major Diagnostic Categories in Hawai`i. Hawai'i journal of health & social welfare, 81(4), 108–114.