Factors Affecting the Length of Stay and Readmission Rates Among Cystic Fibrosis Patients

Derrine Chia

Wan Roshamiliza A. Rahman

Yee Sern Tan

Fall 2017

Indiana University Bloomington

Authors Note

This paper was prepared as part of the requirement for completion of Data Science for Drug Discovery, Health and Translational Medicine course.

Correspondence concerning this paper should be addressed to:

Derrine Chia: dchia@indiana.edu

Wan Roshamiliza A. Rahman: rosrahma@indiana.edu

Yee Sern Tan: yeestan@indiana.edu

ABSTRACT

Background

A fairly common disease affecting more than 30,000 people in America alone, cystic fibrosis (CF) includes symptoms that causes persistent lung infections and shortness of breath[1]. Steady progress has been made in treating CF patients, but currently there is no cure for CF – and therefore, it is essential to continuously measure and improve the quality of provided hospital services for this group of patients. Measures include length of stay and 30-days readmission often considered as a promising quality indicator and in this paper, we aimed to identify the factors (e.g. age, body mass index, Ivacaftor, Lumacaftor) affecting these hospital quality metrics.

Methods

We obtained a sample of 638 inpatients with CF coded as their primary diagnosis, who were admitted to the hospital of interest. A brief descriptive statistic, which include age, sex and race provides information about our patient sample. Exploratory analysis was conducted on other variables to examine general patterns in our data. We employed regression analysis to identify the relationship between our dependent variables - length of stay and readmission rates, and our independent variables – demographic variables, BMI, severity levels and risk of mortality.

Results

No significant results were found from lognormal and logistic regressions, therefore no firm conclusions can be drawn and this might be due to some limitations in the study which are further discussed in the paper.

Conclusions

Severity levels, risk of mortality and BMI might have an impact on length of stay, however several concerns identified in this paper needs to be addressed – for instance, CF patients being treated for a fixed duration (usually 14 days) and only add length of stay if they are not responding to therapy. As most CF patients develop progressive pulmonary disease, forced expiratory volume (FEV1), which has been used as indicators of disease severity might be an important factor in explaining the length of stay.

INTRODUCTION

Cystic fibrosis (CF) is a common genetic disease that primarily affects the respiratory and digestive systems in the United States. The disease causes the body to make thick, sticky mucus that often contribute to problems in the lungs and the digestive system. The symptoms usually include chronic cough, lung infections, and shortness of breath. Nearly 30,000 people are affected in the United States with about 1000 new cases diagnosed every year. The discovery of cystic fibrosis transmembrane conductance regulator (CFTR) as the gene responsible for the fatal genetic disease called CF almost three decades ago has led many researches to develop effective therapies that could improve the quality of life among CF patients. Drug therapies often focus on specific CFTR gene mutations, and with over 2,000 known mutations, there are still a lot to be done to find the new therapies that can treat all CF patients effectively. Life expectancy has improved from about 10 years in 1962 to 37.5 years with today's health care (Pietrangelo, 2016).

Nonetheless, there is still no cure for CF but advances in research are helping people with CF live longer lives. With the longer life expectancy in CF patients, hospitals and health services research have focused on improving health outcomes through delivery of quality care. The length of stay and readmissions, among other clinical and environmental factors, are the two major quality indicators of outcomes for hospitals providing care to CF patients. There are many factors that could affect the length of stay and readmissions including hospital care and treatment, demographic, disease severity and comorbidity, and environmental factors. Health care providers can focus on improving care and treatment for CF patients since they are less likely to influence other factors.

For an acute care, teaching hospital in a metropolitan area, the length of stay and readmission rates among CF inpatients has always been a concern due to the high costs of care and health outcomes associated to these factors. The mean length of stay over the last 12 months was at 12.5 days. Readmission costs is very high among CF patients as compared to other pulmonary patients. We decided to analyze the hospital's CF data to see if any relationship can be established to identify factors affecting the length of stay and readmission rates. We hope that our findings will provide valuable information to clinical leaders at the hospital in making decision about improvement of care among its CF patients.

Problem Statement

The purpose of this paper is to identify factors affecting the length of stay and readmission rates among adult CF inpatients. The findings will help the hospital to understand their CF patients better and to support decisions on resource allocation and inpatient care. We will examine the effect of age, gender, race, body mass index (BMI), severity of illness, and risk of mortality on the length of stay. We will also attempt to predict readmissions using the available explanatory variables.

Importance

The findings of the project are important for the medical provider to understand how sick their patient is - how the severity of illness and risk of mortality affect the length of stay and the rate of readmission. This information is critical to support decisions on resource allocation and quality improvement efforts that is targeted towards shortening the length of stay and reducing the need for unplanned readmission.

Objectives

The objectives of this clinical analysis are to, 1) examine the relationships between length of stay and explanatory factors, which include demographics, BMI, severity of illness and risk of mortality, and 2) conduct predictive analysis on readmissions based on available explanatory variables.

DATA

The data, obtained from a teaching hospital with 350-500 beds in a metropolitan area, contains 638 entries of patients who admitted for CF in years 2015 and 2016. Each entry contains information on the demographics (age, gender, race and ethnicity), conditions (BMI, severity of illness[1] and risk of mortality[1]), and treatment and outcome (length of stay, readmission, discharge disposition, and whether the drugs Ivacaftor or Lumacaftor are prescribed).

Inclusion and exclusion criteria

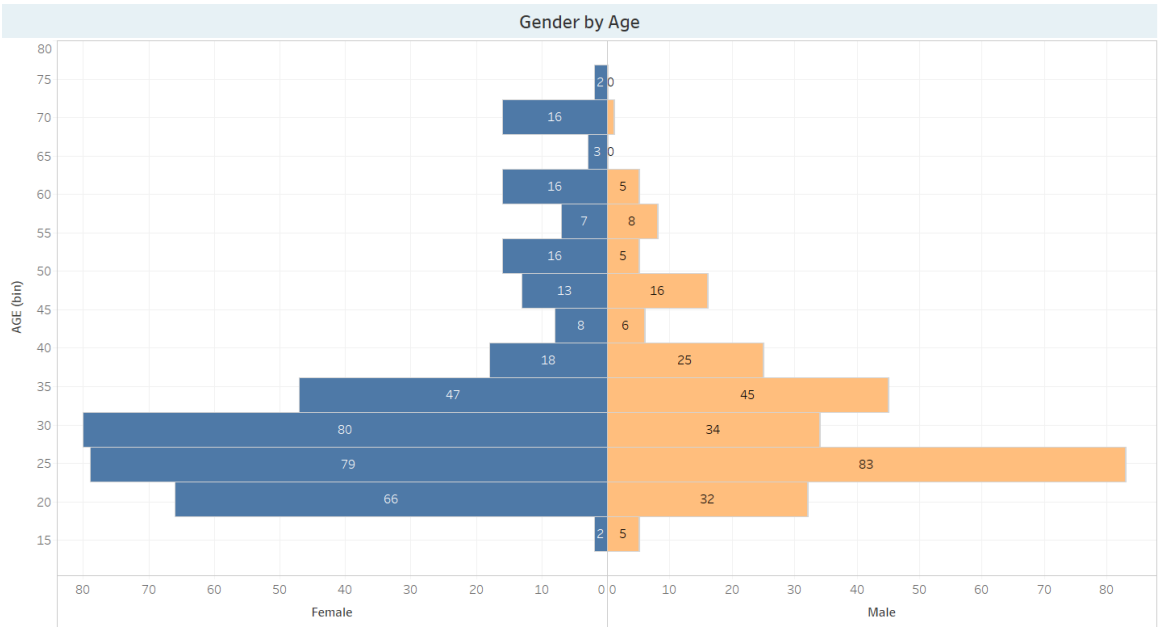
The data includes all patients, 18 years old and older who were admitted to the hospital with any of the CF ICD-9-CM or ICD-10-CM codes (see Appendix I) in the years 2015 and 2016. CF patients who were younger than 18 years old and present at the hospital for observations or short-stays are excluded from the initial sample. Multiple readmissions within the 30-day period are counted as one in order to be consistent with national hospital readmission data.

A routine data cleaning revealed that, out of the 638 samples obtained originally, 2 had problems: one with missing data in BMI, and another has Discharge Disposition filled with value which does not make sense. These 2 entries were removed for subsequent analyses.

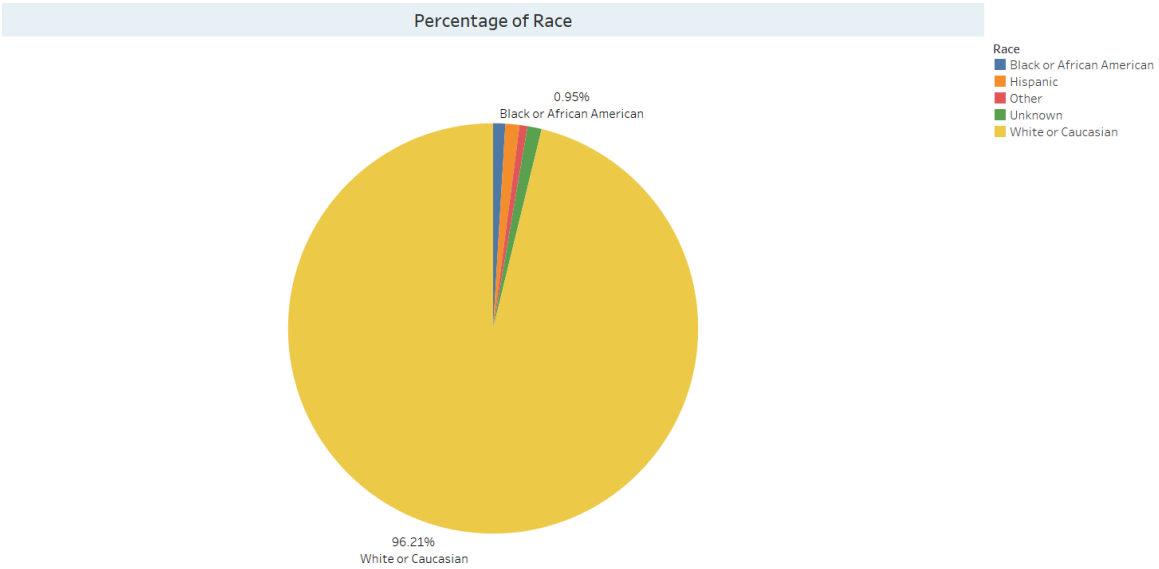
Data exploration

The distribution of age and gender can be summarized with a population pyramid. About half of our CF sample are between 18 and 20 years old. There are about 58 percent females and 42 percent males in the sample. About 60 percent of the females are younger than 30 years old, while about 45 percent of males are younger than 27 years old.

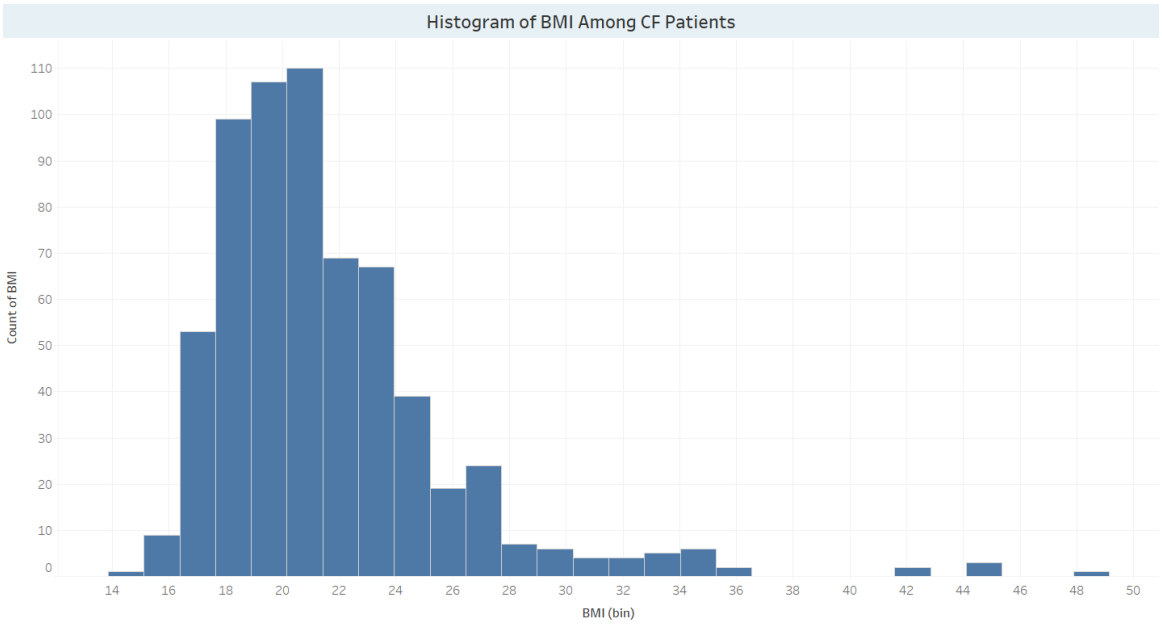
[1] The severity of illness and risk of mortality are as defined by 3M All Patient Refined Diagnosis Related Groups (APR DRGs).



The distribution of race and ethnicity have their vast majorities being “White or Caucasian” (95.6%) and “Non-Hispanic” (95.8%). Such distributions have little information content to add for the prediction of treatment and outcomes.



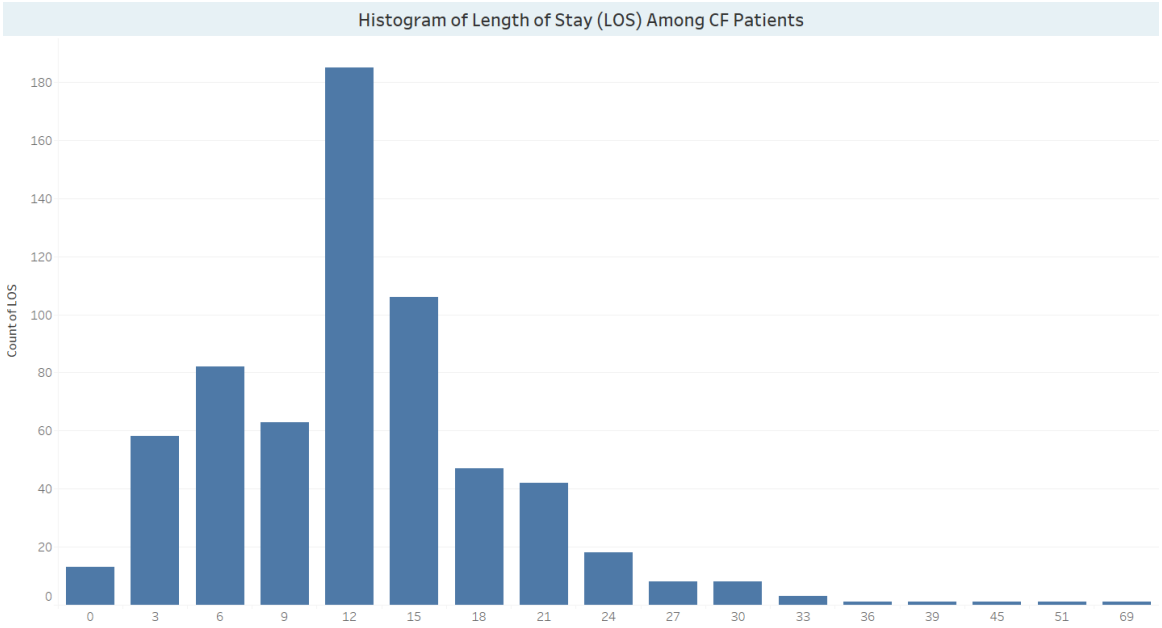
The histogram of patient BMI is generated, displaying a positive skew from normality, with several outlying high values. Low BMI increases in incidence with age and is closely implicated in worsening lung function (McCarthy, O’Carroll, Franciosi & McElvaney, 2015).



The categorical distribution of SOI (severity of illness) and ROM (risk of mortality) are tabulated:

	ROM	Extreme	Major	Moderate	Minor
SOI					
Extreme		8	111	18	1
Major		2	121	212	67
Moderate		0	0	14	76
Minor		0	0	0	6

The length of stay of CF patients follows two regimes. The standard antibiotic treatment is for 14 days. Those who stayed less often go home on intravenous therapy or on continued oral therapy. Those who stay longer are often because of severe illness or of lacking improvement from treatment. As the mode of number of days of length of stay is 14, we extracted the patients with this index larger than 14, and observe a log-normal distribution as follows: log (length of stay – 14) closely resembles a normal distribution.



For the treatments prescribed, there is dependencies among the drugs Ivacaftor and Lumacaftor. Where prescription is done, they are according to the trade names of either Kalydeco or Orkambi. In the former, only Ivacaftor is prescribed; and in the latter, Ivacaftor and Lumacaftor are combined for the treatment.

The interactions between discharge disposition and readmission are as follows:

Readmission	No	Yes
Discharge disposition		
Another hospital with planned readmission	1	0
Another hospital – acute care	2	0
Eloped	0	1
Expired	1	0
Home health care service	122	13
Home or self-care	443	49
Home with home hospice	1	0
Left against medical advice	2	0
Long term acute hospital care	1	0

It can be seen that, where readmission happens, the interesting cases are where the discharge dispositions are: (1) Home health care service, or (2) Home or self-care.

METHODS AND RESULTS

For our analysis, we have identified our targets as length of stay and readmissions, and we wish to predict these values based on the values of other variables. These targets are the main sources of the cost and outcomes of treatment.

Length of stay

The decision to discharge a patient dominantly depends on the standard of care that the hospital guarantees to all its patients. Our work reflects how other variables affect this length of stay. For standard analysis, we apply regression analysis to the subset where length of stay exceeds 14 days, since this subset has a nearly log-normal distribution. Note that, when selecting this subset, the number of entries is reduced to 236, and that there will be no entries having severity of illness level "Minor".

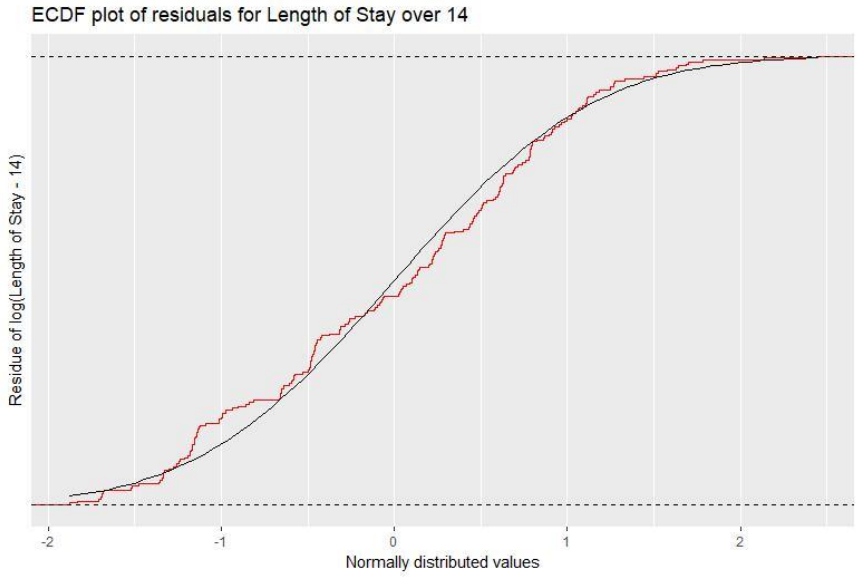
Upon applying the logarithm transformation to the length of stay in excess of 14 days, we obtain our dependent variable, log-LOS. For simplicity of model, we take only the variables that significantly contribute to this prediction, namely BMI, severity of illness, and risk of mortality. Then on, we do linear regression on the model:

Log-LOS = BMI + severity of illness + risk of mortality

The resulting fit is:

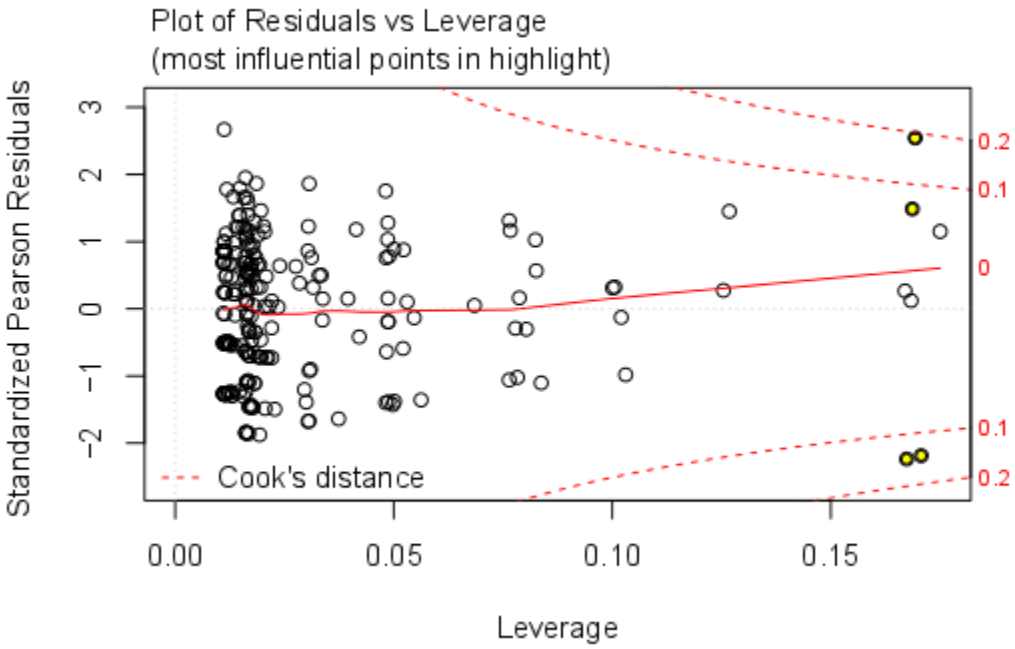
Log-LOS = 2.028 - 0.00797 * BMI
- 0.3367 * I(SOI = Major) - 0.6594 * I(SOI = Moderate)
- 0.1881 * I(ROM = Major) - 0.3608 * I(ROM = Moderate) - 0.2637 * I(ROM = Minor),
where the I() function is the identity function, and the default values for SOI and ROM are both "Extreme". The p-values for the coefficients are only significant for severity of illness, which is also among the strongest effects observed.

To validate on the normality assumptions of this regression, we plot the ECDF of residuals.



From this plot, we conclude that, though not a perfect fit, the ECDF of residuals are close to normal. It also reveals that there are no extreme outliers. Other tests for normality do not pass, as they are stricter. Shapiro-Wilk normality test yields a p-value of 0.0014, while Anderson-Darling normality test yields a p-value of 0.00028, giving evidence that normality is not ideal.

A plot of the Cook's statistics versus leverage of entries reveals that the influences of individual data points are well-contained, apart from a few observations highlighted in the plot. (These 4 observations are in the highest category "Extreme" in severity of illness and risk of mortality).



We have also explored segmentation by age group. In particular, the data for which age is 25 and above, with its 172 size of sample, has an even closer to log-normal distribution of residuals. However, no intuitive explanation is found for this effect, especially since the data only includes patients aged 18 and above, where any biological age cutoff would not make sense.

The R-squared value for this analysis yields a log-likelihood of 0.0257, signaling a far from ideal fit. Therefore, what we have obtained can be considered as an easily interpretable linear prediction that may not be statistically strong.

Readmissions

For readmissions data, we are attempting to predict the occurrence of a situation which is to be avoided, and therefore expect that current practice has already been minimizing it. What we hope to achieve here, is to glean the data for remaining patterns that have not been employed by current practice, but that may increase the likelihood of such an event. To explain further, current practice would suggest patients to stay in a controlled hospital environment, all the way until the relapse of

emergency conditions in the near future is unlikely. This is done until the patient's Forced Expiratory Volume in one second (FEV1) readings approach the baseline level.

To proceed with our analysis, we note that the data on readmissions is binary, and henceforth apply logistic regression in attempts to discover patterns. In this logistic regression, the dependent variable is readmissions, to be predicted with probabilities of the logit function applied to linear combinations of the independent variables.

The independent variables should most prominently feature treatments applied, so that the effects of these treatments can be observed. The above mentioned that the treatment data available can be summarized in one variable with three levels, each corresponding respectively to: no drug treatment, Kalydeco, and Orkambi.

Another interesting predictive factor is the discharge disposition. With readmissions occurring only for 3 levels of this variables, and with the level "elope" signifying factors beyond the scope of this project, the analysis can focus on the 2 remaining levels, namely "Home health care service" (HSERV) and "Home or self-care" (HSELF).

To explore for the effect of all possible variables, they have all been put within the scope of modeling. In addition, another variable indicating whether the length of stay is beyond 14 days has been created.

We express our model as:

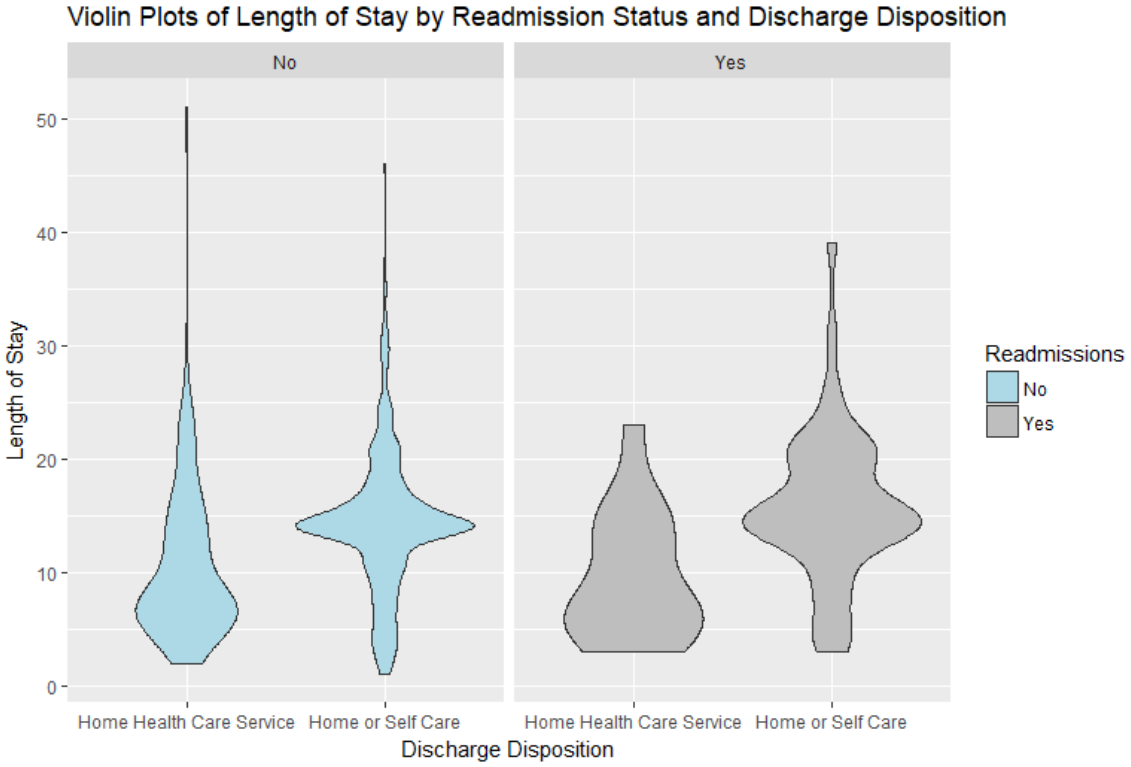
Readmission = logit (I(LOS > 14) + I(discharge disposition = HSERV, HSELF) +
I(prescription = none, Kalydeco, Orkambi) + SOI + ROM + BMI)

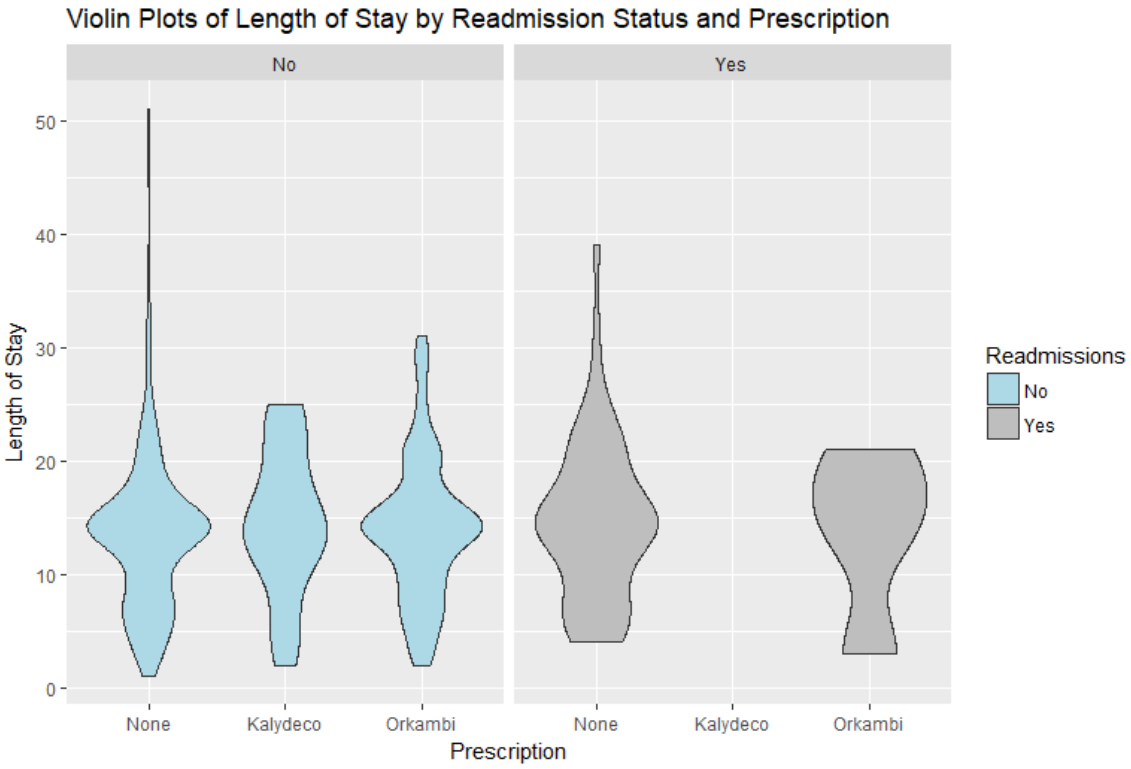
with results:

Readmission =
logit(- 1.574 + 0.09594 * I(LOS > 14) – 0.06884 * I(discharge disposition = HSELF)
- 0.4359 * I(prescription = Kalydeco) – 0.1901 * I(prescription = Orkambi)
- 0.05426 * I(SOI = Major) + 0.02938 * I(SOI = Moderate) – 3.553 * I(SOI = Minor)
+ 0.1726 * I(ROM = Major) – 0.07108 * I(ROM = Moderate) – 0.5268 * I(ROM = Minor)
+ 0.01997 * BMI),

where the default value of discharge disposition is HSERV, that of prescription is none, and that of SOI and ROM are both "Extreme".

It can be suggested, albeit with weak statistical results, that despite staying longer than 14 days, the readmissions of those patients are still higher. On the other hand, the prescription of Kalydeco and Orkambi correlates with lower readmissions, suggesting reduction of readmissions coming from these treatments.


Violin Plots of Length of Stay by Readmission Status and Discharge Disposition
No
Yes
Length of Stay
0
10
20
30
40
50
Home Health Care Service
Home or Self Care
Discharge Disposition
Readmissions
No
Yes


Violin Plots of Length of Stay by Readmission Status and Prescription
No
Yes
Length of Stay
0
10
20
30
40
50
None
Kalydeco
Orkambi
Prescription
Readmissions
No
Yes

Comparing with the null model, where the independent variables are only SOI, ROM, and BMI, McFadden's pseudo R-squared is calculated to have a log-likelihood of 0.0067.

Despite efforts to include all other variables and their interactions, no analysis has found any variable to yield statistically significant results (p-value < 0.1) for which the effect is significant. This shows that the variables we have do not wield much prediction effect on readmissions. There could be other complicating factors not reflected within the limited descriptions of the given variables.

DISCUSSION

To start off with interpreting the results, the source of this data, with the population group treated by this hospital, needs mentioning. Most of the patients who are admitted to the hospital come from the metropolitan area it is situated in. A majority of these patients seek care at a major outpatient respiratory hospital, which frequently refers them to the hospital under study.

We were unable to determine any significant variables in log-normal and logistic regression. Several factors were recognized as important predictors and included in our regression models such as age, body mass index (BMI), severity of illness (SOI), risk of mortality (ROM), types of drug treatment, and discharge disposition. However, none of these were associated with length of stay or readmissions.

It is possible that the readmissions are affected by other factors such chronic anemia and development of acute kidney injury (AKI). Another important outcome for CF patients is lung function, which is measured by FEV_1 and not available in our dataset.

The lack of expert biomedical knowledge prevented us from establishing causal connections from the weak correlations observed in the data. It would be desirable to look at the data in light of these causal, biomedical knowledge.

According to a study published by a group of University of Arkansas for Medical Sciences physicians (K Chatterjee et.al), female gender also appears to be a predictor of increased risk of readmission. We used gender variable as one of the predictors in the initial models, but it did not show a significant influence on readmissions. This could be due to some limitations in our study. As stated in the earlier section, our data was gathered from a teaching hospital and even a smaller sample when we excluded those who had length of stay of 14 days (standard) or lower. Cautions should be exercised in interpreting the results as the sample is not large enough the important predictors and interaction effects, including risk of sample & selection bias.

Steinkamp and Widemann and Gozdzik *et al.* demonstrated that low BMI increases in incidence with age and is closely implicated in worsening lung function, but our efforts are unable to observe this effect, more so without longitudinal data and lung function indicators.

While none of our data are genetic, CF being a genetic disease, indicators of related genes are crucial in advancing the research. It is therefore expected, that progress in precision medicine should improve the study and care of this disease.

CONCLUSION

Given a relatively small sample from the population (i.e. adults CF patients in a teaching hospital with 350-500 beds in a metropolitan area), our data might not be sufficient to conduct the level of complexity that is required by our research objective. Therefore, the results generated could not adequately approximate the true model for our dependent variables. Additional features that affect length of stay and risk of readmission should be considered to improve our models.

APPENDIX

Table 1: ICD-9-CM Codes for Cystic Fibrosis

Disease description	ICD-9-CM Code
Cystic fibrosis without mention of meconium ileus	277.00
Cystic fibrosis with meconium ileus	277.01
Cystic fibrosis with pulmonary manifestations	277.02
Cystic fibrosis with gastrointestinal manifestations	277.03
Cystic fibrosis with other manifestations	277.09

Table 2: ICD-10-CM Codes for Cystic Fibrosis

Disease description	ICD-9-CM Code
Cystic fibrosis with pulmonary manifestations	E84.0
Cystic fibrosis with intestinal manifestations	E84.1
Meconium ileus in cystic fibrosis	E84.11
Cystic fibrosis with other intestinal manifestations	E84.19
Cystic fibrosis with other manifestations	E84.8
Cystic fibrosis, unspecified	E84.9

REFERENCES

25 years later: the impact of the cystic fibrosis gene discovery. (2014). *SickKids Newsroom*. Retrieved from http://www.sickkids.ca/AboutSickKids/Newsroom/Past-News/2014/25-years-later-the-impact-of-the-cystic-fibrosis-gene-discovery.html

3M. (2016). 3M All patient refined diagnosis related group (APR DRGs). Retrieved from https://www.forwardhealth.wi.gov/kw/pdf/handouts/3M_APR_DRG_Presentation.pdf

Chatterjee, K., Goyal, A., Koppurapu, V., Innabi, A., Alzghoul, B., & Jagana, R. (2017). Prevalence and predictors of 30-day readmissions among adults with cystic fibrosis in the United States. doi: http://dx.doi.org/10.1016/j.chest.2017.08.624

Cystic Fibrosis Foundation. Retrieved from https://www.cff.org/Research/About-Our-Research/

McCarthy, C., O'Carroll, O., Franciosi, A. N., & McElvaney, N. G. (2015). Factors affecting prognosis and prediction of outcomes in cystic fibrosis lung disease. In Cystic fibrosis in the light of new research (Chapter 1). Retrieved from https://www.intechopen.com/books/cystic-fibrosis-in-the-light-of-new-research

Pearson, H. (2009, July 8*).* Human Genetics: One gene, twenty years*. Nature*, *460*, 164-169. doi:10.1038/460164a Retrieved from http://www.nature.com/news/2009/090708/full/460164a.html

Pietrangelo, A. (2016). Cystic fibrosis by the numbers: Facts, statistics, and you. *Healthline*. Retrieved from https://www.healthline.com/health/cystic-fibrosis-facts

Steinkamp, G., & Wiedemann, B. (2002). Relationship between nutritional status and lung function in cystic fibrosis: Cross sectional and longitudinal analyses from the German CF quality assurance (CFQA) project. *Thorax, 57*(7), 596-601.

Gozdzik, J., Cofta, S., Piorunek, T., Batura-Gabryel, H., & Kosicki, J. (2008). Relationship between nutritional status and pulmonary function in adult cystic fibrosis patients. *Journal of Physiology and Pharmacology, 59*(Suppl 6), 253-60.