A New Predictive Model for Readmission Among Patients in a California Hospital

Samantha Lee, Ruojie Lu, Clayton Neill[1]

Advisors: Ian Duncan, Jiyoun Myung
Sponsor: Santa Barbara Cottage Hospital

[1] University of California, Santa Barbara

Glossary

Abstract

With the enactment of the Hospital Readmissions Reduction Program under the ACA, hospitals need ways to identify patients who have a high risk of readmission. A popular risk-scoring method is the LACE index, which has limitations. We developed a logistic regression model for Medicare patients using the components of LACE together with other available data. We also utilized third party software, DataRobot, to develop machine learning models (gradient boosted trees were best-performing). Our resulting models identified patients with a high risk of readmission, allowing hospitals to allocate their resources efficiently to reduce readmission risk.

Introduction

Hospital readmission occurs when a patient, who has been discharged from a hospital, is admitted again to the same or different hospital within thirty days. Readmissions are costly and disruptive for both patients and hospitals. For the patient, a readmission increases their possibility of hospital-acquired infections and complications. For the hospital, readmissions lead to higher costs and are logistically inefficient. According to the Agency for Healthcare Research and Quality, hospitals spent $41.3 billion to treat readmitted patients in 2011 [1]. Identifying factors that contribute to readmission risk can lead to improvements in hospital quality and efficiency.

The Hospital Readmissions Reduction Program, established in October 2012, served as a great incentive for hospitals to reduce readmission rates. It requires the Centers for Medicare & Medicaid Services (CMS) to reduce payment to hospitals with readmissions exceeding the national average for certain conditions, such as acute myocardial infarction, heart failure, and pneumonia. In 2015, chronic obstructive pulmonary disease and elective primary total hip and/or total knee replacement were added to the list of penalized conditions. To alleviate this financial pressure, hospitals urgently needed predictive models to identify patients with high risk of readmission.

The LACE index is a widely used tool to calculate readmission risk among hospitals in the United States due to its simplicity and moderate predictive power. When patients are discharged, they are scored by weighting and summing the following components:

- **L**ength of Stay
- **A**cuity of Admission (Emergent admission)
- Charlson **C**omorbidity Index
- **E**mergency department visits within the previous six months

The score is an integer ranging from 0 - 19. A score between 0 - 4 implies a low risk of readmission, 5 - 9 a moderate risk, and 10 - 19 a high risk [5].

In 2015, a group of students at the University of California, Santa Barbara conducted a readmission prediction project to identify factors that contributed to readmission risk [2]. The authors developed a logistic regression model using the LACE index along with other patient data as predictors and yielded better results than using the LACE index alone. In our project, we validate their model on data obtained after their study was completed. We then create our own logistic regression model with additional predictors. Next, we build machine learning models to see if they can perform better than logistic regression. The output of our final model is a score for readmission risk that performs better than the LACE index.

Software packages used: R, Excel, DataRobot.

Data

Description

The data for this study was provided by three regional, not-for-profit hospitals within Santa Barbara County. The dataset consists of admission, readmission, and emergency department visit records from 63,844 unique patients collected between 2010 and 2016. The explanatory variables that were provided for each patient include race, age, sex, source of admission (emergency, observation, etc.), zip code, International Statistical Classification of Diseases and Related Health Problems (ICD) codes, and diagnostic related group (DRG).

ICD codes serve as documentation of a patient's diagnosis and are used to compute risk scores that will later be used as predictors in our model. ICD-9 (ninth revision) codes were used from 2010 to October 2015, after which ICD-10 was used. ICD-10 allows for more detailed diagnosis documentation for the patient. The DRG groups patients into categories for Medicare repayment based on age, sex, ICD codes, and discharge status. Patients in the same DRG are assumed to use similar amounts of hospital resources and have similar "costs".

Additional variables were derived from the raw data such as DRG type, length of stay, Charlson Comorbidity Index, and HCC risk score. DRG type specifies whether the procedure was medical or surgical based on the DRG provided. Length of stay was calculated from the time interval between the given admission and discharge dates. The Charlson Comorbidity Index (CCI) predicts the ten-year mortality for a patient based on diagnosis information. It is computed by mapping the patients ICD codes to a list of comorbidities, assigning weights to each comorbidity, then summing the weights.

The Hierarchical Condition Categories (HCC) risk score identifies high risk individuals based on their demographic data and diagnosis data, and is calibrated for persons over sixty-five years old. Similar to the CCI, it is computed by mapping ICD codes to seventy-nine different categories and assigning additive weights. More than half of the patients in our data do not meet this age requirement for implementing HCC. An alternative, such as the Chronic Illness & Disability Payment System (CDPS) score, would be better for our study. However, CDPS was unavailable at the time of our study, so we resorted to using HCC. Table 1 below gives a summary of the LACE components in our data along with other variables.

Table 1: Data Summary (2010-2016)

Categorical Variables	Factor	Count (%)
Race	Asian	2380 (2.27%)
	Black	2095 (2.00%)
	Hispanic	19,772 (18.83%)
	White	79,213 (75.45%)
	Other	944 (0.90%)
	Unknown	586 (0.56%)
DRG Type	Medical	59,443 (56.62%)
	Surgical	45,551 (43.38%)
Admit Type	Clinic	126 (0.12%)
	Emergency	44,732 (42.60%)
	Pre Admit	40,093 (38.19%)
	Observation	19,844 (18.90%)
	Other	198 (0.19%)
Readmission	No	96,087 (91.52%)
	Readmit	8,907 (8.48%)
Numerical Variables		
Age	Min.	15
	Mean	58.27
	Max.	114
Length of Stay	Min.	1
	Mean	5.05
	Max.	240
ER visits in prev. 6 months	Min.	0
	Mean	0.38
	Max.	4
CCI	Min.	0
	Mean	3
	Max.	17
HCC Score	Min.	0.121
	Mean	0.996
	Max.	9.418

The correlation between readmission and other variables was calculated using two methods: Cramer's V, which measures the strength of association between two categorical variables, and point-biserial correlation, which measures the correlation between one binary variable and one continuous variable. The correlation coefficient for readmission and each predictor is displayed in Table 2 below.

Table 2: Correlation Coefficients (vs. Readmission)

Categorical Variables	**Cramer's V**
Race	0.031
Sex	0.057
Admit Type	0.101
DRG Type	0.109
Numerical Variables	**Point-Biserial Correlation**
Age	0.0781
Length of Stay	0.1031
ER visits in prev. 6 months	0.1488
CCI	0.152
HCC Score	0.187

HCC score appears to be the most correlated with readmission, having a correlation coefficient of 0.187. DRG type, CCI, ER visits in previous 6 months, and length of stay also appear to be significantly correlated. Thus, including these as predictors will likely give the most predictive power to the model.

Data Imputation

The "Admit Type" variable is categorized into clinic, emergency, pre-admit, observation and other, indicating the source a patient is admitted from. Information about admission type serves an important role when predicting readmission risk according to the LACE model. In the dataset, about 5% of individuals have unknown admission type. Although individuals with missing values make up only a small portion of the data, we decided not to remove them and impute their values instead. We used Multiple Imputation by Chained Equations (MICE), which can predict the missing value of one variable from its observed values [4]. To use MICE, we assumed missing values in admission type are missing at random. MICE imputes values for admission type by regressing the observed admission type values on the other covariates using the Classification and Regression Trees (CART) model. This method works for categorical variables and takes into consideration the interaction of nonlinear relationship between admission type and the other covariates. There are several iterations of the imputation step to enhance the accuracy of the imputed values. Lastly, the imputed results are added to the data to replace the missing values.

“Old” Regression Model

This project is an extension of a readmission prediction project in 2015, where the authors built a logistic regression model on similar data. The variables in their model included age, race, sex, DRG type, LACE index, CDPS score, length of stay, and number of ER visits per year (2010-2014). Our study was performed two years later, so we were interested in seeing how the model performed on new data collected between 2015 and 2016.

We attempted to recreate the model, but had to make a few changes. Due to the unavailability of the CDPS score, we used the HCC score. We also used the number of ER visits in the previous six months rather than per year. These differences resulted in a change in all of the model’s coefficients, but performance results on the training set were consistent with the results of the previous study. To test the recreated model on the new data, we chose a cutoff value such that the sensitivity and specificity were equal. Table 3 compares the stated performance of the previous model with the recreated model’s performance on the 2015-2016 data.

Table 3: Model Comparison

Criterion	**Previous Study Results**	**Test on 2015-2016 data**
Cutoff Value	0.086	0.085
Sensitivity	0.700	0.814
Specificity	0.700	0.513
Precision (PPV)	0.150	0.153
AUC	0.780	0.731

We can see that the sensitivity increased with the new data and the precision was higher, even though the area under the curve (AUC) was lower. Metrics such as the F_1 score and balanced accuracy were not provided in the previous study, so we could not compare these values. Although this model has moderate predictive power, we saw areas for potential improvement and proceeded to create a new model.

"New" Regression Model

Data Preprocessing

The provided dataset contains readmissions for the general hospital population. We are also interested in identifying high risk Medicare patients who are readmitted with CMS penalized diagnoses. This way we can potentially reduce readmissions, reducing costs for both the hospital and taxpayers. To do this, we generated a subset of the data containing patients of age 65 or older, then we conditioned the readmission to patients who have penalized diagnoses. The dataset and corresponding models for the general population and all readmissions will be labeled as **General**; the subsetted data with targeted readmissions will be labeled as **Medicare**.

Both the General and Medicare datasets were highly imbalanced. The General dataset contained 8,907 readmissions out of 104,994 observations (8.5%), and the Medicare dataset contained 2,126 readmissions out of 37,199 observations (5.7%). After splitting both datasets to 80% for training and 20% for testing, we used a combination of over and under sampling to balance the training sets so that readmissions made up 50% of observations.

The creators of the LACE index state the limitations of the score in their paper. The index is the result of a logistic regression model trained on a 4,812 person sample in Ontario, Canada with only four predictors, which are the components of LACE [5]. The creators state "the index cannot be used reliably in patient populations not involved in its derivation" and they "recommend that it be used for outcomes research and quality assurance rather than in decision-making for individual patients" (p. 557). For these reasons, we used the components of LACE as predictors rather than the LACE index itself.

We also added socioeconomic data obtained from the U.S. Census Bureau based on the patient's zip code. We grouped patients into their respective counties, and obtained data from the 2010-2016 American Community Survey five-year Estimates for these counties [6]. Three new variables with countywide values were included as predictors: the median household income ("Income"), percentage of population with high-school degree or above ("Education"), and percentage of population below the nation's poverty level (Poverty).

Model Summary

We developed a logistic regression model for each training dataset with readmission as the response. The following table displays the odds ratios of the significant covariates at a 0.05 significance level along with their 95% confidence intervals. A value of "N/A" in Table 4 indicates that the variable was not significant for that particular model.

Table 4: Model Summary

Variable	Odds Ratio (General)	95% Confidence Interval (General)	Odds Ratio (Medicare)	95% Confidence Interval (Medicare)
Intercept	0.424	(0.377,0.478)	0.516	(0.475,0.562)
Race Hispanic (vs. Black)	0.884	(0.797,0.981)	N/A	N/A
Race Other (vs. Black)	0.514	(0.451,0.587)	N/A	N/A
Race White (vs. Black)	0.747	(0.676,0.825)	N/A	N/A
Age	0.997	(0.995,0.999)	1.020	(1.018,1.022)
Sex Male (vs. Female)	1.203	(1.168,1.239)	1.093	(1.043,1.146)
Non-emergent admission	0.744	(0.721,0.767)	0.915	(0.869,0.963)
HCC Score	1.302	(1.277,1.328)	N/A	N/A
DRG Surgical (vs. Medical)	0.544	(0.527,0.563)	0.043	(0.039,0.048)
ER visits in previous 6 months	1.439	(1.417,1.462)	N/A	N/A
Length of Stay (days)	1.213	(1.199,1.227)	N/A	N/A
CCI	1.113	(1.102,1.124)	1.157	(1.141,1.173)
Education	0.943	(0.928,0.958)	0.697	(0.633,0.767)
Income	N/A	N/A	0.714	(0.661,0.771)
Poverty	N/A	N/A	0.665	(0.595,0.744)

For the General model, ER visits in the previous six months, HCC score, length of stay, and CCI all have odds ratios greater than one, showing that an increase in their respective values leads to an increase in readmission risk. Age has an odds ratio of 0.997, likely due to there being more readmissions in lower age groups. Non-emergent hospital admission and surgical DRG both have odds ratios less than one, meaning that they pose less risk of readmission than their respective comparators, emergent admission and medical DRG. For gender, being a female has an odds ratio greater than one, indicating that females are at a higher risk of readmission than males. Among the three socioeconomic variables, only education was significant for the general model. The odds ratio for education was less than one, showing that a patient from a more highly educated county has less risk of readmission. We assumed that education itself may not be directly correlated with readmission, but it can be a proxy for other correlated factors.

For the Medicare model, the patient's race, HCC score, number of ER visits, and length of stay were not significant, which is surprising. The most impactful variable is DRG type, where surgical has an odds ratio of 0.043. This means a patient's odds of readmission decrease significantly if they have surgical DRG. Income and poverty were significant in the Medicare model with odds ratios less than one. The relationship between income and readmission risk makes sense, since we expect patients from wealthier areas to have healthier diets and more time for exercise. However, understanding that patients from impoverished areas are less likely to be readmitted is less intuitive. This is likely due to the location of the hospitals our data came from; they are all located in the same county with a relatively high poverty level. The majority of patients admitted to these hospitals are local, contributing to this phenomenon that patients from poor areas have less risk of readmission.

We then evaluated the models on their respective test sets and compared the results to a LACE model. The LACE model solely uses the LACE index and predicts readmission if the LACE risk is high (greater than or equal to 10). For the logistic regression model, we chose a cutoff value such that the sensitivity was equal to the specificity on the training set. Because the test set is imbalanced, we want to focus on the F_1 score[2] and balanced accuracy[3] for performance measures [7].

Table 5: Model Performance

	Logistic Regression (General)	**LACE (General)**	**Logistic Regression (Medicare)**
Cutoff Value	0.474	N/A	0.596
Error	0.349	0.337	0.322
Sensitivity	0.700	0.620	0.730
Specificity	0.647	0.667	0.675
PPV	0.153	0.145	0.113
F_1 Score	0.251	0.236	0.196
Accuracy	0.651	0.663	0.678
Balanced Accuracy	0.556	0.548	0.546
AUC	0.730	N/A	0.771

The logistic regression model performed slightly better than the LACE model, with higher F_1 score and balanced accuracy. However, these improvements are small and may not give hospitals enough incentive to replace LACE's simplicity and operationality with the more complex logistic regression model.

[2] F_1 score = harmonic mean of sensitivity and precision
[3] Balanced accuracy = average of sensitivity and specificity

Another performance metric considered is lift, which we believe is more convincing for hospitals. If we sort the logistic regression model outputs, which are risk scores for all patients, and split the output list into groups (10 groups in this case), we expect there to be a greater number of readmissions in groups with higher average model scores. For a particular group, the lift is equal to the percentage of actual readmissions in this group divided by the expected percentage of readmissions in this group if patients were chosen randomly (8.4% in this case). Table 6 below shows the lift of each decile for the General logistic regression model on the test set.

Table 6: Lift Table (Logistic Regression - General)

Decile	Number in Decile	Mean Response	Cumulative Mean Response	Cumulative % of Total Responses	Lift	Cumulative Lift	Mean Model Score
10	2099	0.23	0.23	27.4%	2.74	2.74	0.80
20	2100	0.15	0.19	44.7%	1.74	2.24	0.66
30	2100	0.13	0.17	60.7%	1.59	2.02	0.57
40	2100	0.09	0.15	72.0%	1.13	1.80	0.50
50	2101	0.08	0.14	81.4%	.94	1.63	0.43
60	2099	0.05	0.12	87.8%	.64	1.46	0.37
70	2100	0.04	0.11	92.1%	.43	1.32	0.32
80	2100	0.02	0.10	94.9%	.28	1.19	0.28
90	2100	0.02	0.09	97.6%	.27	1.08	0.23
100	2100	0.02	0.08	100.0%	.24	1.00	0.18

The top decile, which consists of the riskiest group of patients according to our model, had a mean response of 23% and a lift = 23%/8.4% = 2.74. The readmissions in this group made up 27.4% of the total readmissions in the test set. If the top two deciles are combined, the group would have a cumulative lift of 2.24 and would contain 44.7% of total readmissions in the test set (compared to an expected 20% by random selection). In this paper, we assign the label **Top 10%** to Cumulative % of Total Responses for the top decile, and **Top 20%** to Cumulative % of Total Responses for the top two deciles combined. These measures were added to the performance metrics for model comparison purposes. The LACE model yielded a Top 10% = 25.5% and a Top 20% = 40.6%, which are around 2% and 4% lower respectively than the logistic regression model. Thus, by implementing our regression model instead of the LACE model, hospitals are able to identify more high-risk patients for intervention. The Top 10%/20% metrics for the Medicare model can be found in Table 7.

The logistic regression model is sufficient at identifying risky patients, and it outperforms the LACE model. Although the logistic regression model is not as easy to operationalize as LACE, it does a better job in stratifying the patients according to their risk, which allows hospitals to allocate extra resources and care to those high-risk patients to reduce their chance of readmission. Next, we discuss the machine learning models we created to compare with the logistic regression model.

Machine Learning Models

We used DataRobot, a third party automated machine learning platform, to deploy several machine learning models. The models in DataRobot were fitted on the same training set and tested on the same testing set as the logistic regression models above, for both the General data and Medicare data. For the General dataset, we included an additional variable "DRG Desc", which specifies the text description of each patient's DRG. For the Medicare data, DRG Desc was excluded, since the target response was conditioned on the patient's DRG.

The capabilities in Datarobot allowed us to execute various models, including a Gradient Boosted Tree (GBM), Light Gradient Boosted Tree (LGBM), TensorFlow Neural Network, Support Vector Machine with Nystroem Kernel, Elastic net, and Random Forest. We also tested an ensemble model, known as a "blender" in DataRobot.

The performances of these models were compared on the basis of F1 score, balanced accuracy, and Top 10%/20% measures. The best performing models were LGBM for the General data and the GBM for the Medicare data. Table 7 shows the performance metrics of these models along with the results from Table 5 to ease comparison between the machine learning models ran in DataRobot and the logistic regression models ran in R. The cutoff value for the DataRobot models was chosen to maximize the F_1 score on the training data.

Table 7: Model Performance

	LGBM (General -DataRobot)	**Logistic Regression (General -R)**	**GBM (Medicare -DataRobot)**	**Logistic Regression (Medicare -R)**
Cutoff Value	0.438	0.474	0.578	0.596
Error	0.325	0.349	0.158	0.322
Sensitivity	0.739	0.700	0.460	0.730
Specificity	0.670	0.647	0.864	0.675
PPV	0.170	0.153	0.161	0.113
F_1 Score	0.276	0.251	0.239	0.196
Accuracy	0.675	0.651	0.842	0.678
Balanced Accuracy	0.568	0.556	0.563	0.546
AUC	0.771	0.730	0.797	0.771
Top 10%	32.4%	27.4%	31.0%	26.8%
Top 20%	50.8%	44.7%	54.8%	46.8%

The machine learning models outperform the logistic regression models in nearly all aspects for both the General data and Medicare data. The GBM for Medicare had a 56.3% balanced accuracy and Top 10%/20% = 31.0%/54.8%, which is a significant improvement on the logistic regression model's 54.6% balanced accuracy and Top 10%/20% = 26.8%/46.8%. The LGBM also outperformed the logistic regression model for the General data, with a 1% increase in balanced accuracy and 5%/6% increase in Top 10%/20%. The sensitivity of the GBM at the chosen cutoff value is 0.460 which is low, but has a higher precision relative to other models. This is due to the tradeoff between sensitivity and precision when changing the cutoff value. As the cutoff value increases, the sensitivity decreases and the precision increases. Thus, we put more focus on the F_1 score, which is the average of these two values.

Conclusion

The LACE index is simple and has moderate predictive power to identify risky patients. However, its limitations in individual-based predictions encouraged us to develop a better model. We generated a logistic regression model using the components of LACE along with other variables as predictors. We yielded better performance compared to the LACE model and gained valuable insights on the association between patient health and their socioeconomic background. We then used DataRobot to create a variety of machine learning models to see if they could perform better than the logistic regression model. The best performing models were the light gradient boosted tree for the General data and the gradient boosted tree for the Medicare data; each outperformed the logistic regression models for their respective datasets by a considerable margin.

Although our models performed well, we believe certain adjustments can be made to increase performance. One improvement would be to use the CDPS score instead of the HCC score as a predictor for the General model. As we stated in the Data section of this paper, HCC is calibrated for patients over the age of 65, while the average age of patients in the General data is 58. Using CDPS can lead to more accurate predictions for working-age patients. Another improvement would be to use a sampling technique such as ROSE[4] or SMOTE[5] to balanced the data instead of over and under sampling. These techniques cannot be used on text variables, so the "DRG Desc" predictor would need to be removed.

Lastly, just as the authors of LACE advised against using LACE on individual-based predictions on patients not used in its derivation, we would like to do the same. Until the model has been validated on patient data outside of the area it was created, we believe our model should be used as an assessment of hospital quality and performance.

[4] Random Over-Sampling Examples
[5] Synthetic Minority Over-sampling Technique

References

1. Hines, A. L., Barrett M. L., Jiang, H. J., & Steiner, C. A. “Conditions With the Largest Number of Adult Hospital Readmissions by Payer 2011.” *HCUP Statistical Brief #172* (2014): 2. Available from: http://www.hcup-us.ahrq.gov/reports/statbriefs/sb172-Conditions-Readmissions-Payer.pdf.
2. Huynh, N., Fallah, H., & Robbins-Kelley, D. “Improving the Prediction of Re-admissions Amongst Medicare Patients in a California Hospital.” *ARCH: Actuarial Research Clearing House,* 2017.1 (2017).
3. Charlson, M. E., et al. “A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation.” *J Chron Dis* 40.5 (1987): 373-383.
4. Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. "Multiple Imputation by Chained Equations: What Is It and How Does It Work?" *International Journal of Methods in Psychiatric Research* 20.1 (2011): 40-49.
5. Van Walraven, C., et al. “Derivation and Validation of an Index to Predict Early Death or Unplanned Readmission after Discharge from Hospital to the Community.” *CMAJ : Canadian Medical Association Journal* 182.6 (2010): 551–557. *PMC*. Web. 15 Mar. 2018.
6. U.S. Census Bureau. "Community Facts." *American FactFinder.* (2017). Web. 15 Mar. 2018.
7. Bekkar, M., Djemaa, H. K., & Alitouche, T. A. “Evaluation Measures for Models Assessment over Imbalanced Data Sets.” *Journal of Information Engineering and Applications* 3.10 (2013): 29.