Title: Impact of Medication Adherence on Hospital Admissions among Medicaid patient populations with Heart Failure.

Abstract

Objective: To examine the impact of adherence to chronic disease medications on hospital admissions among Medicaid enrollees with heart failure.

Subjects: Admissions encounters, diagnosis, medications, and Insurance eligibility datasets were pulled from NYU Langone Health system from 2017 to 2020. The dataset of Medicaid recipients consisted of 18 – 64 ages with Heart Failure and in addition, >= 1 of 7 chronic conditions:1) congestive heart failure 2) hypertension 3) dyslipidemia 4) diabetes 5) asthma/COPD 6) depression 7) schizophrenia/bipolar.

Methods: Poisson regression machine learning model was used to predict the association between medication adherence (based on categorical Proportion of Days Covered – PDC) and 2 dependent variables: number of hospital admissions, total number of comorbidities. In addition, other machine learning models such as Logistic Regression and unsupervised learning methods were also used to do comparative analysis.

Results: Full adherence was associated with fewer hospital admissions among those Medicaid patients with heart failure and any of the 7 comorbidities – congestive heart failure, hypertension, diabetes, asthma/COPD, depression and schizophrenia/bipolar. In addition, Negetive Binomial Regression model showed better results than Poisson Regression model which was used in the model study.

Conclusions: Substantial reduction in hospital admission and therefore, healthcare cost reductions can be achieved with increased medication adherence among Medicaid patient population. Patients at below 0.80 threshold can benefit from medication adherence intervention programs such as chronic care management or pharmacotherapy to achieve medication adherence greater than 0.80 threshold.

Introduction:

The model research paper that I have selected analyzes the impact of medication adherence on Health Services Utilization on Medicaid[1.] This study was based on Medicaid claims, encounter, and eligibility datasets. The author used Poisson regression methods to estimate the impact of medication adherence on 3 measures of health services utilization:1) inpatient hospitalization2) ER visits 3) outpatient visits. In addition, they also used descriptive statistics to analyze the impact of defined 7 comorbidities (based on Charlson Comorbidity Index), age, sex, and race/ethnicity on medication adherence. The access to Medicaid data was not open source due to HIPAA privacy rules and therefore, I have used NYU Langone dataset to replicate similar study which is based on geographic area limited to NYU Langone inpatient admissions data during 2017 to 2020 only.

According to CDC[2] nearly half of US adult population (129 million) suffer from one or more conditions such as hypertension, chronic kidney disease, heart failure and cancer. And, 83 % of total Medicaid recipients (75 million) suffer from at least one or more chronic conditions (Roebuck, Kaestner et al. 2018). So, there are about 62 million Medicaid patients need additional support to manage their chronic conditions. Yet there are relatively few studies have done analytic work which can identify these populations and impact of their medication adherence on hospital admissions and reduce an economic burden through improved prevention and treatment. [3]

My Study examines the impact of medication adherence on hospital admissions among Medicaid recipients with 7 chronic diseases. Three years of encounter data, medication orders, diagnosis, and insurance data of 563 patients were analyzed. I used Poisson regression model, Ordinary Least Squares (OLS) linear regression, Negative Binomial regression model and for supervised method – K means model for comparative analysis.

Methods: Data

This study used data from NYU Langone Health Systems. The data was collected in four different cohort listed below using SQL queries from the Epic EHR data lake (Cogito and Clarity).

This study modeled data collection based on the Roebuck's study I used to replicate. The files below collected health insurance information, hospital encounter records, medication orders, calculated pdc (proportion of days covered) data and finally diagnosis records from problem list for collecting information about comorbidities.

<u>Cohort1: Admissions</u> – This file pulled all patient records with their hospital encounter id, inpatient admission date, inpatient discharge date and diagnosis name and icd10 code at the time of discharge limited to heart failure, period 2017 to 2020 and age 18 to 64. This file contained listed all hospital encounters during defined time frame. Therefore, to calculate the total number of admissions, I used COUNTIF function to calculate

number of admissions based on patient_id and dropped other columns and finalized columns were patient_id and admissionCount with total unique patients 4810.

Cohort2: Diagnosis – This file pulled all above defined patient records with their chronic conditions diagnoses as defined in Roebuck's paper (see attached Appendix 1) which limits to 7 conditions congestive heart failure, hypertension, dyslipidemia, diabetes, asthma/COPD, depression and schizophrenia/bipolar. Data was pulled from the problem list which contains diagnosis history based on all encounters. To calculate total number of comorbidities I used Roebuck's naming conventions which is based on Charlson Comorbidity Index to define 7 classes of diagnosis. Then the 7 classes of diagnoses converted into columns using pivot table function and calculated total number of comorbidities for each patient. So, the finalized spreadsheet contained patient_id, COPD (count), Depression, Diabetes, Dyslipidemia, Congestive Heart Failure, Hypertension, Schizophrenia, and total counts (for all comorbidities). The total unique count of patients 4810.

Cohort3: Medications – This file pulled all defined medication orders for the specified classes in Roebuck's paper (appendix 1) and based on that I also pulled pdc score which is a calculated field in Epic database for each medication order. I added an additional column titled average pdc which calculated average pdc based on total number of defined classes of medication orders for each patient. The final spreadsheet contained columns patient id and PDC avg.

Cohort4: Insurance – This file pulled all the above defined patient cohort with their payer insurance name and insurance plan name. I have applied limitations to payer name only Medicaid and Managed Medicaid. I needed this file to create a match with my final data processing where I wanted to limit above patient cohort who only have Medicaid insurance.

Data Processing in python – After combining and completing data processing, total number of patient count was 563.

1) Combined Admissions data frame and diagnosis data frame on patient ID. For models I dropped all columns except patient_id, admission count and totalcomorb. Total patients count 4810
2) Combined above final dataframe with medications(pdc) dataframe. Total patient count was 4175. So, there were about 635 patients who did not have any medication orders for defined classes for my study.

3) Final merge of above data frame was with insurance file on patient id. Here, my merge found only 569 total count of patients who were on Medicaid insurance which counts to about 13.7 % of total selected cohort of 4175 patients.

Methods: Medication Adherence measurement

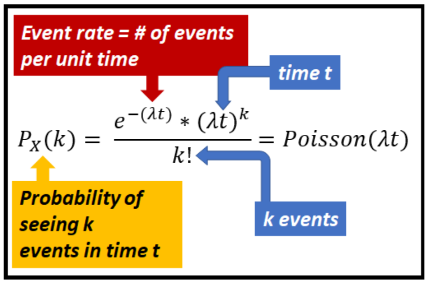
For this study medication adherence measurement was based on pharmacy claims data for patients diagnosed with heart failure and if they have one or more defined comorbidity. I used Epic EHR PDC score that is a calculated variable data in the dataset. Medication Adherence is calculated using Proportion of Days Covered (PDC) which is the metric recommended by the Pharmacy Quality Alliance and Centers for Medicare and Medicaid Services. PDC is the proportion of calendar days in a period that a patient had medication on hand to treat their chronic condition. My sampled data was limited to patients who have at least 1 prescription of defined medication classes for their selected chronic conditions. Patient was considered medication adherent if the score was 0.80 or higher based on average PDC score of total number of medications prescribed. If the score is less than 0.80, patients are considered non- adherent. I have added a column Medication Adherent based on this calculation and identified patient's cohort as adherent (1) or non-adherent (0).

Methods: Models

Roebuck used Poisson Regression model however, I have also added 3 additional models for the comparative analysis: Ordinary Least Squares (OLS) linear regression model, Negative Binomial regression model and unsupervised model K-Means (2 clusters). My goal is to predict the number of hospital admission based on whether patient is medication adherent, Type of Medicaid insurance and Total number of comorbidities.

<u>Ordinary Least Squares (OLS)</u> – As explained earlier, dataset for this study is count based and whole numbers. OLS linear regression models work on the principle of fitting n-dimensional linear function to n-dimensional data, in such a way that the sum of squares of differences between fitted values and actual values are minimized.

<u>Posisson Regression model</u> – I used GLM class of statsmodels and trained the data set. The data set for this study is a whole number of data and distribution is skewed meaning the large number of samples have 2 or less admissions during the defined period. Similarly large number of patients have 2 or more comorbidities and non- adherent. Poisson regression model has following probability Mass function.



[Grab your reader's attention with a great quote from the document or use this space to emphasize a key point. To place this text box anywhere on the page, just drag it.]

Negative Binomial regression model (NB) does not make the variance = mean assumptions about the data. This is another model which can be useful in predicting count-based data.

Supervised method – K Means (2 clusters). – This is a dividing method, and it divides n points to k clusters. I have also used scaled data to avoid outliers. Inputs database scaled data and K is 2 clusters I wanted to divide.

Results: Descriptive Statistics

Initial analysis for total comorbidities as shown in fig1, indicated most patients in this study cohort had at least 2 comorbidities. Fig2 shows most patients are non-medication adherent cohort. Fig 3 showed most patients had at least one or more hospital admissions.

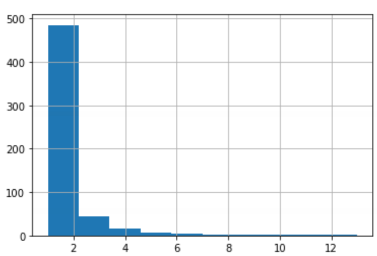


Fig1: Total Comorbidities

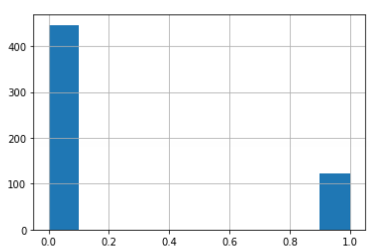


Fig2: Medication

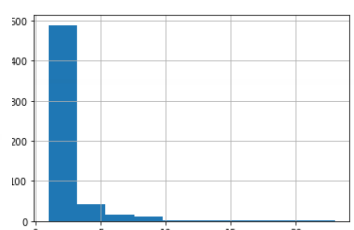


Fig3: Hospital Admissions

Initial data exploration shows correlation between higher number of illnesses with higher number of non-medication adherence and higher number of hospital admissions. Studies have shown that often patients with higher number of illnesses have complex medication regimen and often they tend to become non medication adherent which leads to frequent hospital admissions. Identifying these high-risk patients at the point of care would not only save lives but also reduce huge economic burden.

Results: Which model performed best for predicting hospital admissions?

Below fig4 shows the parameter significance using the two-tailed t-test. The t-test shows that all regression parameters are statistically significant. The absolute value of the t-statistic for each parameter is greater than the specified threshold t-value at the 95% significance level in the two-tailed t-test. All three models show similar result (refer to fig 8 and fig 10). Fig 4also shows the F-statistic's p-value of 1.22 e-153 is much smaller than the 0.025 error threshold of a Chi-squared distribution with 6 degrees of freedom, indicating that all regression parameters are jointly significant. In addition, the adjusted R2 is indicating that the OLSR model is able to explain more than 75% of the variance in the hospital admission counts dependent variable.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         AdmissionCount   R-squared:                       0.798
Model:                            OLS   Adj. R-squared:                  0.797
Method:                 Least Squares   F-statistic:                     585.1
Date:                Sat, 14 May 2022   Prob (F-statistic):          1.22e-153
Time:                        21:58:25   Log-Likelihood:                -459.76
No. Observations:                 447   AIC:                             927.5
Df Residuals:                     443   BIC:                             943.9
Df Model:                           3
Covariance Type:            nonrobust
=================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
Intercept                        0.2027      0.034      6.038      0.000       0.137       0.269
TotalComorb                      0.5970      0.014     41.603      0.000       0.569       0.625
Med_Adh                          0.0548      0.079      0.691      0.490      -0.101       0.211
prod_type_Medicaid               0.0833      0.042      1.982      0.048       0.001       0.166
prod_type_MedicaidManagedCare    0.1194      0.034      3.509      0.000       0.053       0.186
==============================================================================
Omnibus:                      105.463   Durbin-Watson:                   2.203
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1457.875
Skew:                           0.569   Prob(JB):                         0.00
Kurtosis:                      11.774   Cond. No.                     1.62e+16
==============================================================================
```

Fig4: Summary of the fitted OLSR model

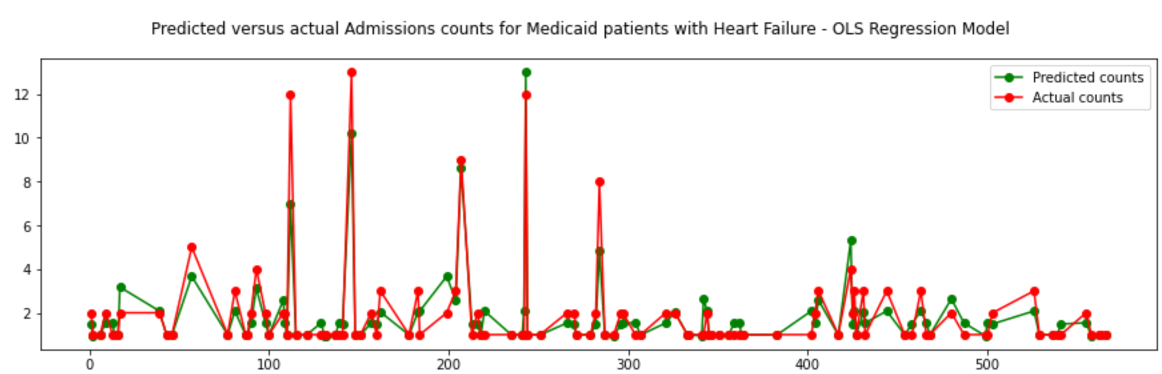
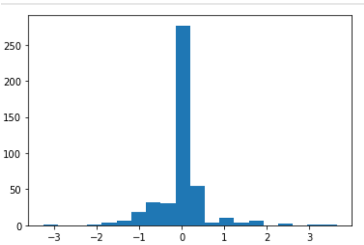


Fig5: Predicted Vs Actual Hospital Admission Counts based on OLSR Model

Fig 6 and fig 7 shows that residual errors of the OLSR model are not exactly normally distributed around the mean of 0 but slightly skewed. Q-Q plot looks mostly linear but flattened at 0.



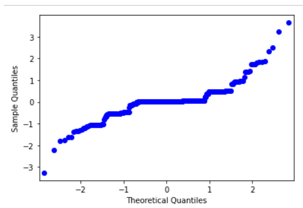


Fig7: Q-Q plot of the residual errors of OLSR

Fig6: Histogram of residual errors

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:           AdmissionCount   No. Observations:                  447
Model:                              GLM   Df Residuals:                      443
Model Family:          NegativeBinomial   Df Model:                            3
Link Function:                      log   Scale:                          1.0000
Method:                            IRLS   Log-Likelihood:                -602.82
Date:                  Sat, 14 May 2022   Deviance:                       60.113
Time:                          21:58:28   Pearson chi2:                     63.8
No. Iterations:                       8
Covariance Type:              nonrobust
===============================================================================================
                                    coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
Intercept                        -0.0415      0.045     -0.924      0.356      -0.130       0.047
TotalComorb                       0.1909      0.013     14.239      0.000       0.165       0.217
Med_Adh                           0.0371      0.109      0.340      0.734      -0.176       0.251
prod_type_Medicaid               -0.0226      0.057     -0.396      0.692      -0.134       0.089
prod_type_MedicaidManagedCare    -0.0189      0.046     -0.407      0.684      -0.110       0.072
===============================================================================================
```

Fig8: Negative Binomial Regression Model

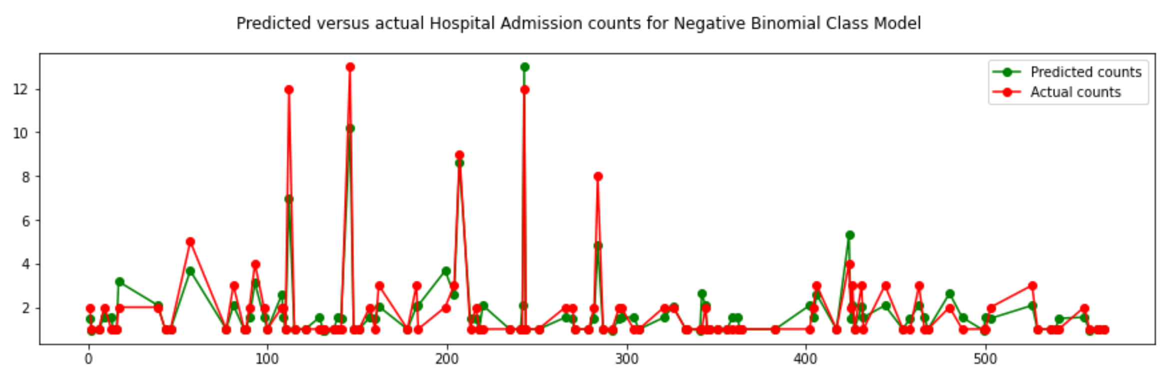


Fig9: Negative Binomial Class Model actual vs Predicted

Poisson Regression Model

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:          AdmissionCount   No. Observations:                  456
Model:                             GLM   Df Residuals:                      452
Model Family:                  Poisson   Df Model:                            3
Link Function:                     log   Scale:                          1.0000
Method:                           IRLS   Log-Likelihood:                -570.48
Date:                 Sun, 15 May 2022   Deviance:                       104.53
Time:                         15:12:11   Pearson chi2:                     116.
No. Iterations:                      5
Covariance Type:             nonrobust
===============================================================================================
                                  coef    std err          z      P>|z|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
Intercept                       0.0293      0.037      0.800      0.424      -0.042       0.101
TotalComorb                     0.1543      0.009     18.016      0.000       0.138       0.171
Med_Adh                        -0.0448      0.091     -0.491      0.623      -0.224       0.134
prod_type_Medicaid             -0.0239      0.051     -0.467      0.641      -0.124       0.076
prod_type_MedicaidManagedCare   0.0532      0.038      1.394      0.163      -0.022       0.128
===============================================================================================
```

Fig10: Poisson Regression Model

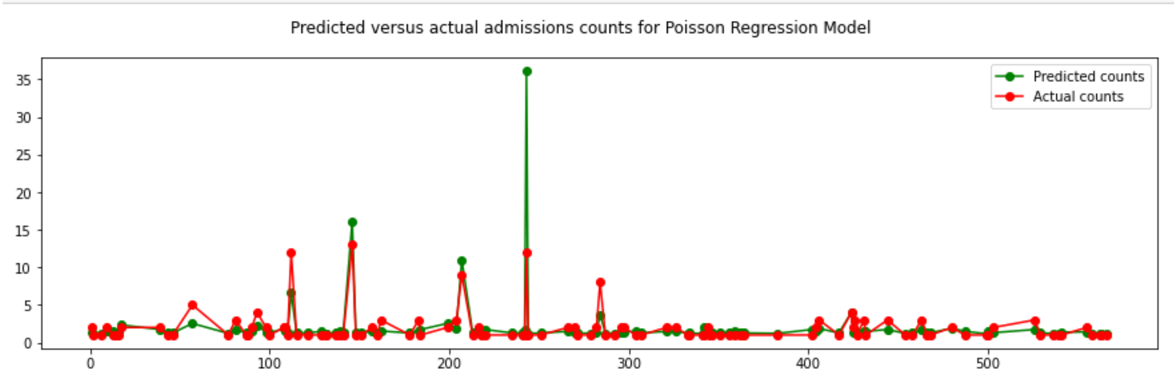


Fig11: Poisson Regression Model actual vs Predicted

Discussion:

In comparison to all three models, NB regression model appears to have fitted the data optimally. Also as show in Fig 11, fig9 and fig 5 plot which shows actual versus predicted value, NB model shows actual and predicted are much closer to each other than other two models. In the model study they have used Poisson regression model. They have used much larger data set and in addition, they also have included social and economic variables in the analysis which might have an impact on how the model is predicting. I also used K- means clustering with 2 clusters of predicted vx actual and it did not perform well. In conclusion, this study can add further data and social and economic data for these patients to perform better analytics.

References

1. Roebuck, M. C., Kaestner, R. J. & Dougherty, J. S. Impact of Medication Adherence on Health Services Utilization in Medicaid. *Med Care* **56**, 266-273, doi:10.1097/MLR.0000000000000870 (2018).
2. cdc
3. Partnership to Fight Chronic Disease National factsheet on the impact of chronic disease in the United
4. States. 2017.
5. Karanasiou, G. S. *et al.* Predicting adherence of patients with HF through machine learning techniques. *Healthc Technol Lett* **3**, 165-170, doi:10.1049/htl.2016.0041 (2016).
6. 2 Gu, Y. *et al.* Predicting medication adherence using ensemble learning and deep learning models with large scale healthcare data. *Sci Rep* **11**, 18961, doi:10.1038/s41598-021-98387-w (2021).
7. 3 Olsen, C. R., Mentz, R. J., Anstrom, K. J., Page, D. & Patel, P. A. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. *Am Heart J* **229**, 1-17, doi:10.1016/j.ahj.2020.07.009 (2020).