# Florida Blue Interview Project

Andrew S. Cistola, MPH

The following document is designed to serve as a comprehensive report for the interview with Florida Blue on Wednesday August18, 2021. The report is organized by the questions provided and corresponding data sources.

**Executive Summary**

Below are brief answers to the requested questions. More information is detailed in the report.

***Question 1:***
ER visits are negatively associated with costs ($r_S$ = -0.67)
Patients with high severity are **significantly** more likely to have higher annual cost, regardless of the number of ER visits (see regression models in Section 1)

***Question 2:***
The raw data was imported into a SQLite database and accessed with:

```
SELECT healthDat.*, fl_zip.County, fl_zip.City,
FROM healthDat
LEFT JOIN fl_zip
ON healthDat.Zip = fl_zip.Zip;
```

***Question 3:***
Chronic conditions by significantly different by age groups.
Costs are concentrated within working age, marketplace eligible beneficiaries.
Data-driven approaches for population specific cost reductions designed to reduce inpatient hospital visits include:

> 30-64: Broad expansion of high-value primary care w/ reduced PCP-copay to assist in chronic disease prevention
> 65+: Introduce focused low cost emergency care w/out copay to prevent hospitalization for acute issues

Initial regression models show that copays do not have a significant effect on overall spending for either group.

***Question 4:***
Crude prevalence: Asthma, Hypertension, and Hyperlipidemia
Area-adjusted prevalence: Asthma, Hypertension, and Hyperlipidemia
Age-adjusted prevalence: Heart Failure, Asthma, and Ischemic Heart Disease

***Question 5:***
See attachments and files.

## Section 1: Emergency Room Visits and Overall Patient Cost

*Question 1: Using the ER.csv file, please describe the relationship between cost and the number of ER visits.*

Emergency care is significantly more expensive than many other forms of inpatient or outpatient care in the U.S(1–4). Because of this cost, patients that frequently utilize the Emergency Room (ER) may be significant drivers of overall system cost(3–7). Understanding the relationship between ER utilization and overall patient cost is a significant question for various stakeholders within the U.S. healthcare system(8–12).

To describe the relationship between the total cost by patient in the previous year (cost) the number of visits to the ER by patient in the past year (ER visits), an ordinary least squares (OLS) regression model was developed. In addition, a number of statistical tests were conducted to verify conditions about the data and a hierarchical linear model was developed in response to those tests. All statistical tests using the data for this population were conducted in R version 3.6.4.

*OLS Assumptions*

A regression model can describe the magnitude and significance of a possible relationship of the predictor variable (ER Visits) on the outcome (cost). For the regression model to provide information appropriate for this interpretation, a number of assumptions need to be met concerning the nature of the data.

Asssumption 1. It can be assumed that the outcome is dependent on the predictor and the predictor is independent of the outcome.

Asssumption 2. The data is randomly sampled among a population and the number of samples exceed the number of predictors.

Asssumption 3. The predictor and the outcome are correlated.

Asssumption 4. There is a linear relationship between the outcome and predictor. The predictors are independent and not correlated with each other

Asssumption 5. The error between the linear model and the actual data follows a normal distribution with a mean of 0.

Asssumption 6. The error between the linear model and the actual data is randomly and evenly distributed across the data[1]

Asssumption 7. The error between the linear model and the actual data is not associated with each other and does not show clear patterns.

*Assumption 1-2*

Since the high cost of care experienced in emergency settings is not likely to induce patients to select to receive emergency care and the average yearly cost occurs after the visits occur, it is reasonable assume that ER visits are independent of cost and cost is

[1] These statements reflect in general assumptions for OLS regression in order to obtain the best linear unbiased estimators (BLUE) but are not directly quoting the original theorem or derivative quotations or explanations of OLS assumptions.

dependent on the number of ER visits. In the sample provided, there are no missing values and the number of observations (N = 500) well exceed the possible predictors variables.[2]

*Assumption 3 - 4*

Spearman's rank correlation coefficient was calculated and a negative correlation in relatively high magnitude was observed ($r_S = -0.67$). While the presence of correlation was expected, the original theoretical framework assumed a positive correlation between ER visits and cost.

The data provided also includes a measure of the most common severity level associated with the past year's ER visits (severity). This data was recategorized into binary variables for each category ("Very Low", "Low", "Moderate", "High", and "Very High"[3]).

Severity and ER were used as independent predictor variables[4] to create a simple OLS regression model. The results of this model are shown below in Table 1:

| **Coefficient** | **Estimate** | **Std. Error** | **t value** | **Pr(>\|t\|)** |
|---|---|---|---|---|
| (Intercept) | 7548.0 | 871.1 | 8.665 | <2e-16 |
| ER_Visits | 1815.2 | 102.8 | 17.655 | <2e-16 |
| Very_High | 27965.6 | 545.5 | 51.262 | <2e-16 |
| High | 19283.5 | 454.6 | 42.422 | <2e-16 |
| Moderate | 14341.1 | 392.7 | 36.518 | <2e-16 |
| Low | 6999.0 | 340.0 | 20.583 | <2e-16 |

Adjusted R-squared: 0.9091
F-statistic: 999.1 on 5 and 494 DF,
p-value: < 2.2e-16

Table 1. Original OLS Model

Utt's rainbow test was used to confirm the presence of a linear relationship (p-value = 0.859). A collinearity matrix was used to evaluate the independence of the predictor variables and low collinearity was observed (Figure 1).

*Assumption 5-7*

While the model above meets assumption 1-4 and has both and F-statistics indicating that it is highly accurate in predicting each observation (high adjusted $R^2$) and accounts for the overall variation within the data (F-statistic)[5], cost outcomes tend to follow non-normal patterns in their distribution leading to the use of generalized linear models.

In order to verify whether these approaches (Poisson regression or Negative Binomial regression) are appropriate, multiple visual and numeric tests were utilized. The Jarque-Bera test (p-value = 0.8152) and "Q-Q" plot (Figure 2) indicated a normal distribution of

---

[2] Since this is an exercise, the sample is assumed to be random or reasonably reflective of the population. However, in a true scenario, the sampling strategy would need to be evaluated for systematic error.
[3] "Very Low" was used as the reference group due to the expected effect of increased severity on overall cost based on the theoretical model.
[4] While the two predictor variables are measured in different ways and are technically independent of each other, patients that are more likely to have high severity visits may also be more likely to have more frequent visits.
[5] These are very untechnical interpretations of these two statistics.

error terms while the Anderson-Darling test (p-value = 1.2e-06) and the "residuals" plot (Figure 3) indicate an even distribution of error terms. Similarly, the Goldfeld-Quandt test (p-value = 0.9091) does not show a significant amount of heteroskedasticity within the data.

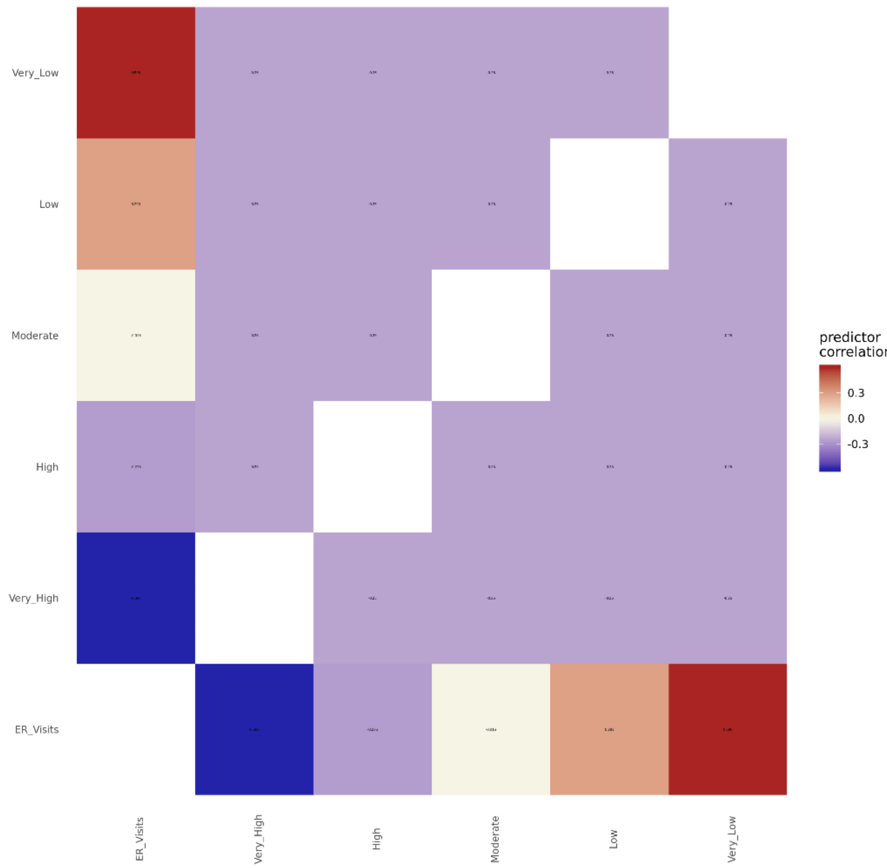
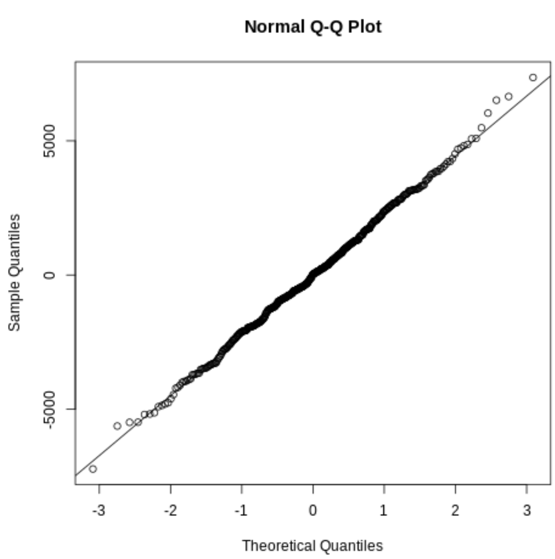



Figure 1. Correlation Matrix for Predictors



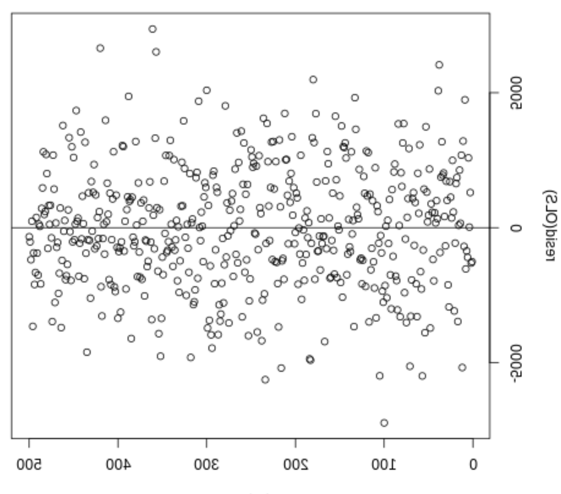




Figure 2. Q-Q Plot and Figure 3. Residuals Plot

The findings above indicate that all OLS assumptions are broadly met and an OLS model is an appropriate way to model the relationship between ER visits and cost. While the outcome is a cost measurement, generalized linear models are not necessary in this case.

However, both the relatively high negative correlation of ER visits with "Very High" severity and positive correlation of ER visits with "Very Low" severity is an observation of interest. Similarly, the coefficients for severity level are much higher in magnitude when compared to the number of ER visits. These initial findings require adjustment in the modeling approach.

*Hierarchical Linear Model*

In order to separate the possible effect of severity on ER visits, a hierarchical linear model (HLM) using severity level as the higher level with intercepts allowed to vary within all observations was developed. This model allows for the calculation of the intra-class correlation coefficient (ICC) that can describe the amount of variation attributed to different levels within the hierarchical model. In this case, the ICC can describe the variation in the data attributed to severity level when modeling the relationship between ER visits and cost. Table 2 shows the results of the final HLM model.

| Fixed effects: | Estimate | Std. | Error | dt value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| (Intercept) | 21305.505 | 4871.102 | 4.124 | 4.374 | 0.0111 |
| ER_Visits | 1808.623 | 102.764 | 494.942 | 17.600 | <2e-16 |

One Way ANOVA p-value < 2.2e-16
Intraclass Correlation Coefficient (Adjusted): 0.958

Table 2. HLM Model

Based on the high ICC, a considerable portion of the variation in the data can be attributed to the level of severity, with ER visits only accounting for less than 5% of the variation in the data. These findings are consistent with the original OLS model.

## Interpretation

It is broadly assumed that patients that frequent the Emergency Room (particularly for non-acute conditions) are significant drivers of overall health care spending(13–16). The passage of the Emergency Medical Treatment and Labor Act of 1986 mandated that emergency services be provided regardless of the ability for patients to pay(17). This had led to the belief that individuals may choose to utilize the ER for non-emergent health needs out of a concern for cost-sharing(2–4). Health systems, local governments, and insurance companies have recently developed ER diversion programs designed to redirect these patients to lower cost sources of care (primary care or urgent care)(9,13,18).

As these programs have become more popular, recent evidence has shown that these types of programs often have little benefit(9,19). While addressing social determinants through health care delivery is a difficult task, the underlying issue may be related to the assumptions around frequent ER visits and high-cost patients(7). Research with this population has indicated that utilizing the ER for non-acute needs is a not a primary driver of avoidable costs, but rather the exacerbation of chronic conditions that are not addressed in primary care (particularly due to social determinants)(3). The number of visits are not necessary the primary issue, but instead the severity of illness that requires emergency care(7).

In this population, a similar pattern can be the observed as the number of ER visits accounts for a small portion of the variation within the data. The initial strong negative association is much smaller in magnitude when accounting for severity level, indicating that the effect of ER visit count is minor when compared to another known predictor.

An appropriate interpretation of the data would state that patients within this population experiencing high severity are significantly more likely to have higher annual cost, regardless of the number of ER visits. For stakeholders interested in controlling yearly costs for this patient population, a more thorough investigation of the causes for high severity visits would be recommended.

**Section 2: Chronic Conditions and Patient Cost**

*Question 2: Using the healthDat.csv and fl_zips.csv files, please bring in the city and county that each patient lives in and show how you did it. Additionally, if you do not have access to a SQL supported tool, please write SQL pseudo-code that will accomplish this as well.*

In order to join the tables “healthDat” and “fl_zips”, the following steps were taken:

1. A local directory was created to house and organize the data, documentation, code, and other files related to the project.

2. A python script was written to import the tables (along with “ER” table) as “.csv” files, create a SQLite database, and export the tables to the database for later access within R, Python, and SAS programs.

3. The following Python script was used to select tables healthDat and fl_zips connect using a left inner join (Figure 4).

```
con_sqlite = sqlite3.connect('_data/project.db')
query = '''
    SELECT healthDat.*, fl_zip.County, fl_zip.City,
    FROM healthDat
    LEFT JOIN fl_zip
    ON healthDat.Zip = fl_zip.Zip
    ;'''
df_Q = pd.read_sql_query(query, con_sqlite)
df_Q.info()
```

Figure 4. SQL query to SQLite database within Python script

*Rationale*

The left inner join as used in order to keep all records from the healthDat table and only records that matched from the fl_zips table. Since booth tables contained the same “Zip” column that was defined and stored identically, this column was used to match records. Since “County” and “City” were the only columns needed from the “fl_zip” table, it was the only two in the “SELECT” statement. This query was used to generate a table that contained all the columns within the “healthDat” table along with the geographic labels for analysis in Questions 3-5.

*Question 3: Using the healthDat.csv and fl_zips.csv files, please provide 2-3 observations or recommendations as they relate to costs. Keep in mind that the goal is to reduce overall costs for this population.*

*Question 4: Using the healthDat.csv file, please list the 3 most prevalent chronic conditions as well as describe any relationships between conditions that you may have found.*

***Observation 1: Differences in prevalence among chronic conditions by age and area.***

In order to identify chronic conditions that are relevant to higher cost in patients, it is necessary to identify the top chronic conditions for the population represented in the data. Table 3 shows the crude prevalence of each of the chronic conditions (CC) among beneficiaries as well as the age and area adjusted prevalence. Table 4 shows the age and areas specific rates along with the adjusted rates for further detail.

***Table III. Crude, Area, and Age Adjusted Rates of Chronic Diseases***

| | **Total Crude** | **Area Adjusted** | **Age** Adjusted** |
|---|---|---|---|
| **Population*** | 15000 | 5000 | 1250 |
| **Chronic Condition***** | | | |
| *Arthritis* | 12.9 | 13.4 | 47.5 |
| *Asthma* | 60.5 | 59.4 | 127.6 |
| *Atrial Fibrillation* | 23.5 | 23.3 | 77.6 |
| *Autism* | 29.1 | 27.7 | 40.2 |
| *Cancer* | 33.5 | 36.0 | 68.9 |
| *COPD* | 45.3 | 45.0 | 93.7 |
| *Dementia* | 10.8 | 10.3 | 44.0 |
| *Depression* | 40.5 | 42.7 | 86.0 |
| *Diabetes* | 27.2 | 24.9 | 69.9 |
| *Heart Failure* | 41.6 | 42.6 | 156.2 |
| *Hepatitis* | 16.6 | 16.7 | 37.9 |
| *HIV AIDS* | 9.7 | 8.8 | 28.2 |
| *Hyperlipidemia* | 50.5 | 51.6 | 115.9 |
| *Hypertension* | 53.6 | 55.4 | 119.4 |
| *Ischemic Heart Disease* | 41.5 | 43.3 | 119.6 |
| *Kidney Disease* | 23.1 | 24.3 | 66.6 |
| *Osteoporosis* | 24.5 | 25.3 | 54.2 |
| *Schizophrenia* | 9.4 | 8.6 | 36.1 |
| *Stroke* | 34.4 | 35.1 | 110.9 |

***Population is represented as counts, not rates**
****Age ranges are determined by Florida Department of Health methods for Age Adjusted Mortality Rates and Areas are defined by Counties Duval, Saint Johns, and Clay**
*****Rates are calculated per 1000 beneficiaries rounded to 1 decimal place**

Table 3. Crude, Age, and Area Adjusted Rates of Chronic Disease

*Table IV. Crude, Age, and Area Adjusted Rates of Chronic Diseases along with Age and Area Specifc Rates*

| | **Total** | **Rates by Area**** | | | | **Rates by Age**** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Crude** | **Clay** | **Duval** | **Saint Johns** | **Adjusted** | **15-19** | **20-24** | **25-34** | **35-44** | **45-54** | **55-64** | **65-74** | **75-84** | **Adjusted** |
| **Popualtion*** | 15000 | 1790 | 10262 | 2948 | 5000 | 1092 | 3434 | 5624 | 2738 | 1218 | 555 | 262 | 77 | 1250 |
| | | | | | | | | | | | | | | |
| **Chronic Condition***** | | | | | | | | | | | | | | |
| *Arthritis* | 12.9 | 14.0 | 12.4 | 13.9 | 13.4 | 0.0 | 0.3 | 7.3 | 14.2 | 26.3 | 66.7 | 122.1 | 142.9 | 47.5 |
| *Asthma* | 60.5 | 52.0 | 60.5 | 65.8 | 59.4 | 0.0 | 9.3 | 36.3 | 104.5 | 167.5 | 189.2 | 202.3 | 311.7 | 127.6 |
| *Atrial Fibrillation* | 23.5 | 25.7 | 24.1 | 20.0 | 23.3 | 0.9 | 2.3 | 14.4 | 32.9 | 44.3 | 104.5 | 148.9 | 272.7 | 77.6 |
| *Autism* | 29.1 | 27.9 | 30.5 | 24.8 | 27.7 | 0.0 | 7.0 | 27.4 | 52.6 | 61.6 | 36.0 | 45.8 | 90.9 | 40.2 |
| *Cancer* | 33.5 | 41.9 | 31.8 | 34.3 | 36.0 | 0.0 | 7.0 | 30.2 | 54.1 | 60.8 | 75.7 | 103.1 | 220.8 | 68.9 |
| *COPD* | 45.3 | 41.9 | 45.1 | 47.8 | 45.0 | 0.0 | 5.8 | 28.8 | 71.2 | 130.5 | 160.4 | 145.0 | 207.8 | 93.7 |
| *Dementia* | 10.8 | 7.8 | 10.9 | 12.2 | 10.3 | 0.0 | 1.5 | 6.2 | 11.3 | 23.8 | 41.4 | 99.2 | 168.8 | 44.0 |
| *Depression* | 40.5 | 49.2 | 39.2 | 39.7 | 42.7 | 0.0 | 7.6 | 28.8 | 69.4 | 95.2 | 106.3 | 133.6 | 246.8 | 86.0 |
| *Diabetes* | 27.2 | 19.6 | 28.8 | 26.5 | 24.9 | 0.9 | 5.2 | 20.5 | 33.6 | 61.6 | 97.3 | 145.0 | 194.8 | 69.9 |
| *Heart Failure* | 41.6 | 46.4 | 41.1 | 40.4 | 42.6 | 1.8 | 3.2 | 19.0 | 44.9 | 114.1 | 183.8 | 389.3 | 493.5 | 156.2 |
| *Hepatitis* | 16.6 | 16.8 | 16.6 | 16.6 | 16.7 | 0.9 | 3.8 | 13.2 | 27.8 | 26.3 | 39.6 | 87.8 | 103.9 | 37.9 |
| *HIV AIDS* | 9.7 | 7.8 | 10.4 | 8.1 | 8.8 | 0.0 | 1.5 | 7.5 | 14.6 | 18.1 | 32.4 | 34.4 | 116.9 | 28.2 |
| *Hyperlipidemia* | 50.5 | 55.3 | 50.0 | 49.5 | 51.6 | 0.9 | 6.1 | 36.6 | 81.8 | 117.4 | 149.6 | 209.9 | 324.7 | 115.9 |
| *Hypertension* | 53.6 | 64.3 | 53.1 | 48.9 | 55.4 | 0.0 | 6.4 | 38.2 | 84.0 | 122.3 | 183.8 | 248.1 | 272.7 | 119.4 |
| *Ischemic Heart Disease* | 41.5 | 46.9 | 40.3 | 42.7 | 43.3 | 0.0 | 4.4 | 26.3 | 56.3 | 88.7 | 182.0 | 274.8 | 324.7 | 119.6 |
| *Kidney Disease* | 23.1 | 26.8 | 22.2 | 23.7 | 24.3 | 0.0 | 4.4 | 10.0 | 30.7 | 66.5 | 97.3 | 167.9 | 155.8 | 66.6 |
| *Osteoporosis* | 24.5 | 26.8 | 23.9 | 25.1 | 25.3 | 0.0 | 5.5 | 20.3 | 29.2 | 59.9 | 90.1 | 72.5 | 155.8 | 54.2 |
| *Schizophrenia* | 9.4 | 6.7 | 9.9 | 9.2 | 8.6 | 0.0 | 1.2 | 7.3 | 12.4 | 17.2 | 30.6 | 38.2 | 181.8 | 36.1 |
| *Stroke* | 34.4 | 35.8 | 33.7 | 36.0 | 35.1 | 0.0 | 3.5 | 10.9 | 46.4 | 112.5 | 169.4 | 232.8 | 311.7 | 110.9 |

***Population is represented as counts, not rates**

****Age ranges are determined by Florida Department of Health methods for Age Adjusted Mortality Rates, Areas are defined by Counties Duval, Saint Johns, and Clay**

*****Rates are calculated per 1000 beneficiaries rounded to 1 decimal place**

When crude rates are considered asthma. hypertension, and hyperlipidemia are the highest among the population. Due to the high prevalence of these conditions among the U.S. population, this is expected(20). This is also the case when adjusting for area.

However, when adjusting for age Heart Failure becomes the condition with the highest prevalence followed by asthma and ischemic heart disease. This indicates that when age is considered, hyperlipidemia and hypertension are less prevalent. This is also consistent with the U.S. population as hypertension and hyperlipidemia effects greater portions of the population over time. Similarly, cardiovascular diseases of any kind are consistently the highest cause of death for the U.S. population(21,22).

Hypertension and hyperlipidemia treatment is relatively commonplace for elderly patients, but ensuring access to care for asthma and COPD patients may be a higher priority for costs within this population. Younger patients may benefit from targeted approaches to prevent cardiovascular disease like access to specialists or early adoption of front-line medications for those individuals at younger ages with higher risk.

The difference in calculating the rates indicate that approaches to improving CC care among the covered population will need to be tailored by age.

***Observation 2: Concentration of cost within working age, marketplace eligible beneficiaries.***

The Patient Protection and Affordable Care Act of 2010 (ACA) introduced many changes to the availability of health insurance to the U.S. population(23). Some of those changes included the expansion of different coverage types based on age and dependency status. To evaluate the opportunities for cost savings within the population, it is necessary to describe the differences in cost experienced by age groups within the covered population that can be offered different types of coverage plans. Table 5 shows the differences in costs as well as inpatient, and ER utilization for enrollees by age range and ACA plan availability group.

Marketplace eligible beneficiaries (not-dependents, not able to qualify for catastrophic plans, and not able to qualify for Medicare) comprise the largest portion of costs ($164,810,102; 67%). Medicare eligible patients experience the largest cost per beneficiary ($77,001) as well as inpatient and ER visit per beneficiary (2.88, 2.91). Each of these rates are more than double the marketplace eligible population.

While Medicare eligible beneficiaries may be the most costly on an individual basis, the overall cost is relatively low when compared to the Marketplace eligible population. This information is particularly helpful when considering where to target interventions that are designed to control costs among various population groups in the context of coverage options.

Medicare eligible beneficiaries have the greatest potential to reduce costs broadly among the population through reductions in per beneficiary utilization. However, the total proportion of the costs generated by that population limit the opportunity for major changes. A 10% reduction in utilization and corresponding spending would be a major achievement, but would only yield a 1% overall reduction in costs.

| *Table V. Cost totals, averages, and percents by ACA coverage availability options** | | | | | |
|---|---|---|---|---|---|
| | **Dependent (0-17)** | **Catastrophic (18-29)** | **Marketplace (30-64)** | **Medicare (65+)** | **All Ages** |
| **Population** | 5208 | 2557 | 6896 | 339 | 15000 |
| **Total** | | | | | |
| *Cost* | $28,876,964 | $26,363,529 | $164,810,102 | $26,103,403 | $246,153,998 |
| *Inpatient Visits* | 1071 | 1203 | 9494 | 978 | 12746 |
| *ER Visits* | 3034 | 2019 | 10244 | 987 | 16284 |
| **Per Beneficiary** | | | | | |
| *Cost* | $5,545 | $10,310 | $23,899 | $77,001 | $29,189 |
| *Inpatient Visits* | 0.21 | 0.47 | 1.38 | 2.88 | 1.23 |
| *ER Visits* | 0.58 | 0.79 | 1.49 | 2.91 | 1.44 |
| **Percent of Total** | | | | | |
| *Cost* | 11.7% | 10.7% | 67.0% | 10.6% | 100.0% |
| *Inpatient Visits* | 8.4% | 9.4% | 74.5% | 7.7% | 100.0% |
| *ER Visits* | 18.6% | 12.4% | 62.9% | 6.1% | 100.0% |
| ***Plan availability is assumed based on age but may not be an accurate option for each beneficiary** | | | | | |

Table 5. Cost totals, averages, and percents by ACA coverage availability options

### *Observation 3: Opportunities for population specific approaches to controlling patient cost*

Based on the first two observations, there is evidence that population specific interventions may be most effective when designing approaches to reduce CC burden and subsequent costs. In order to design these approaches the factors that contribute most to overall cost for each group need to be identified in a robust manner.

Utilizing an approach that combines three open-source, widely utilized machine learning algorithms (Random Forests, Principal Component Analysis, and Recursive Feature Elimination)(24–32), the CC and utilization patterns for the population represented in the data were ranked based on various measures of relevance. These measures compare the relative importance of each CC and utilization variable in predicting the outcome (Random Forests), the amount the variable contributed to the variation in the data (Principal component analysis) and whether the variable would be selected when attempting to build a the most accurate and parsimonious cross-validated regression model (Recursive Feature Elimination). By using multiple machine learning approaches potential biases in algorithmic design can be mitigated and useful information for decision makers can be easily gleaned(33,34). Furthermore, this approach does not require a previous hypothesis to be tested but rather allows for the automated generation of predictive features that can be further tested with appropriate regression models.

Table 6 illustrates the differences among marketplace and Medicare eligible beneficiaries. A number of predictors are present in both groups indicating clear patterns by age and coverage availability. These include clinical conditions such as the number of comorbidities, COPD, stroke, and depression. Inpatient visits were also both present in the top three for each group, indicating that hospital admissions may be the primary driver of costs for both populations. While age was present for both, it was less important for older patients.

***Table VI. Algorithmic ranking of chronic conditions and utilization patterns ability to predict higher overall costs among beneficiaries.****

| **Marketplace Eligible Beneficiaries******** | **Importance Rank** | **Variance Rank** | **Elimination Rank** | **Overall Rank** |
|---|---|---|---|---|
| *Age** | 2 | 1 | 1 | 1 |
| *Inpatient Visits*** | 1 | 2 | 1 | 2 |
| *ER Visits** | 3 | 3 | 1 | 3 |
| *Comorbidities** | 4 | 4 | 1 | 4 |
| *Primary Care Copay* | 7 | 8 | 1 | 5 |
| *Atrial fibrillation* | 6 | 15 | 2 | 6 |
| *COPD** | 11 | 10 | 4 | 7 |
| *Stroke** | 14 | 5 | 8 | 8 |
| *Ischemic Heart Disease* | 9 | 11 | 9 | 9 |
| *Depression** | 10 | 14 | 6 | 10 |

| **Medicare Eligible Beneficiaries******** | **Importance Rank** | **Variance Rank** | **Elimination Rank** | **Overall Rank** |
|---|---|---|---|---|
| *Inpatient Visits*** | 1 | 2 | 1 | 1 |
| *Comorbidities** | 2 | 3 | 1 | 2 |
| *ER Copay* | 9 | 1 | 5 | 3 |
| *Diabetes* | 5 | 9 | 1 | 4 |
| *Stroke* | 10 | 5 | 4 | 5 |
| *COPD** | 6 | 12 | 2 | 6 |
| *Depression** | 8 | 8 | 6 | 7 |
| *Hyperlipidemia* | 16 | 4 | 3 | 8 |
| *Age** | 3 | 24 | 1 | 9 |
| *ER Visits** | 4 | 25 | 1 | 10 |

***Variable is shared among both groups**
****Variable is shared in top 3 among both groups**
*****Only top 10 variables by overall rank are kept. Overall rank is calculated by an average of each of the other three ranks.**
******Marketplace (30-64) and Medicare (65+) regardless of actual plan enrollment**

Table 6. Algorithmic ranking of chronic conditions and utilization patterns ability to predict higher overall costs among beneficiaries

The differences among the lists provide key pieces of information relevant to decision makers. For marketplace beneficiaries, the amount of primary care copay highly ranked but not important for Medicare beneficiaries where the ER copy amount functions in the opposite manner. Combined with the clinical differences, there is evidence that different approaches for each population may be most effective in reducing costs.

*For marketplace beneficiaries:*

Since the primary care copay and number of ER visits are a significant predictor of cost, there may be an opportunity to lower inpatient admissions through expanding low-cost, high value primary care services for chronic disease management. This can be managed through the offering of different plans that offer low to no cost sharing for these types of services, especially when focused on heart disease prevention(35). Non-physician providers (RD, EP, APRN) as well as other wellness sources may be able opportunities for options that have low-marginal cost when applied to a broad population(2,9,36–41).

*For Medicare beneficiaries:*

Just as the primary care setting was more significant for the marketplace population, the emergency setting may be a better place to focus on cost reductions for this smaller but higher average cost population. The ER copay may be discouraging pre-emptively addressing an acute condition, leading to a longer hospitalization. For COPD and Diabetes patients, lapses in regular care can lead to long hospital visits if not addressed quickly. While emergency care in an inpatient setting can be high cost, other emergency providers (EMT, Fire) may be able to provide lower cost high value supports for elderly patients that have higher acuity concerns on a more regular basis. Using plan design and coverage options to expanding high value emergency care that can respond quickly to acute conditions may be best, especially when focusing on a smaller population(9,36–39,42–46).

While these recommendations do their best to interpret the existing data with an objective and quantitative perspective, further analysis is necessary for any kind of implementation or intervention. The methods are designed to showcase how data-driven staring points can be obtained for increasing precision in population health management.

*Initial Regression Models*

In order to provide initial evidence for these recommendations, two regression models were built to test the effect of different co-pays on overall spending for either Marketplace and Medicare beneficiaries. Each model included number of comorbidities, inpatient visits, and age. PCP copay was used as the primary predictor variable for marketplace beneficiaries, and ER copay for Medicare beneficiaries. The results of the model are shown in Figure 5 and 6.

While the regression models show moderate performance (Adj. $r^2$ of 0.39 and 0.34), the variables for PCP and ER copay were not statistically significant. Similarly, other initial tests showed a clear non normal distribution of error consistent with cost outcomes. Further study would be needed to identify more accurate the patters in this population.

However, the initial findings do provide evidence that copays may not have a strong effect on overall costs for covered patients and that other mechanisms may be necessary in order to contain costs and offer more competitively priced products.

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 4 | 1.5648481E12 | 391212029440 | 1122.47 | <.0001 |
| Error | 6891 | 2.4017056E12 | 348527872.09 | | |
| Corrected Total | 6895 | 3.9665537E12 | | | |

| R-Square | Coeff Var | Root MSE | HP_Paid Mean |
|---|---|---|---|
| 0.394511 | 78.11459 | 18668.90 | 23899.38 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Age | 1 | 784120177883 | 784120177883 | 2249.81 | <.0001 |
| Comorbidity | 1 | 375174424318 | 375174424318 | 1076.45 | <.0001 |
| PCP_Copay | 1 | 2854890631.5 | 2854890631.5 | 8.19 | 0.0042 |
| IP_Visits | 1 | 402698624927 | 402698624927 | 1155.43 | <.0001 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Age | 1 | 134666873428 | 134666873428 | 386.39 | <.0001 |
| Comorbidity | 1 | 244845508843 | 244845508843 | 702.51 | <.0001 |
| PCP_Copay | 1 | 2506619521.2 | 2506619521.2 | 7.19 | 0.0073 |
| IP_Visits | 1 | 402698624927 | 402698624927 | 1155.43 | <.0001 |

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| Intercept | -11250.85409 | 1304.654865 | -8.62 | <.0001 |
| Age | 599.31031 | 30.488785 | 19.66 | <.0001 |
| Comorbidity | 6759.03860 | 255.010236 | 26.50 | <.0001 |
| PCP_Copay | -53.41980 | 19.919426 | -2.68 | 0.0073 |
| IP_Visits | 4591.08801 | 135.065460 | 33.99 | <.0001 |

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 4 | 451794556507 | 112948639127 | 44.20 | <.0001 |
| Error | 334 | 853569926937 | 2555598583.6 | | |
| Corrected Total | 338 | 1.3053645E12 | | | |

| R-Square | Coeff Var | Root MSE | HP_Paid Mean |
|---|---|---|---|
| 0.346106 | 65.65214 | 50552.93 | 77001.19 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Age | 1 | 98513073568 | 98513073568 | 38.55 | <.0001 |
| Comorbidity | 1 | 217808151906 | 217808151906 | 85.23 | <.0001 |
| ER_Copay | 1 | 17657308149 | 17657308149 | 6.91 | 0.0090 |
| IP_Visits | 1 | 117816022884 | 117816022884 | 46.10 | <.0001 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Age | 1 | 24450314412 | 24450314412 | 9.57 | 0.0021 |
| Comorbidity | 1 | 193994750140 | 193994750140 | 75.91 | <.0001 |
| ER_Copay | 1 | 15473005059 | 15473005059 | 6.05 | 0.0144 |
| IP_Visits | 1 | 117816022884 | 117816022884 | 46.10 | <.0001 |

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| Intercept | -128859.8052 | 50487.96861 | -2.55 | 0.0111 |
| Age | 2243.3906 | 725.28551 | 3.09 | 0.0021 |
| Comorbidity | 10164.1426 | 1166.60029 | 8.71 | <.0001 |
| ER_Copay | -139.2590 | 56.59555 | -2.46 | 0.0144 |
| IP_Visits | 9004.7949 | 1326.22731 | 6.79 | <.0001 |

Regression model comparing copays and cost for eligible Figure 5. Marketplace and and Figure 6. Medicare Beneficiaries

References


1. Kwack H, Sklar D, Skipper B, Kaufman A, Fingado E, Hauswald M. Effect of managed care on emergency department use in an uninsured population. Ann Emerg Med. 2004 Feb;43(2):166–173.
2. Raven MC, Kushel M, Ko MJ, Penko J, Bindman AB. The effectiveness of emergency department visit reduction programs: A systematic review. Ann Emerg Med. 2016 Jun 8;68(4):467–483.e15.
3. LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. Ann Emerg Med. 2010 Jul;56(1):42–48.
4. Moe J, Kirkland SW, Rawe E, Ospina MB, Vandermeer B, Campbell S, et al. Effectiveness of interventions to decrease emergency department visits by adult frequent users: A systematic review. Acad Emerg Med. 2017;24(1):40–52.
5. Weinick RM, Burns RM, Mehrotra A. Many emergency department visits could be managed at urgent care centers and retail clinics. Health Aff (Millwood). 2010 Sep;29(9):1630–1636.
6. Korczak V, Shanthosh J, Jan S, Dinh M, Lung T. Costs and effects of interventions targeting frequent presenters to the emergency department: a systematic and narrative review. BMC Emerg Med. 2019 Dec 30;19(1):83.
7. Iovan S, Lantz PM, Allan K, Abir M. Interventions to Decrease Use in Prehospital and Emergency Care Settings Among Super-Utilizers in the United States: A Systematic Review. Med Care Res Rev. 2020;77(2):99–111.
8. Ahn S, Basu R, Smith ML, Jiang L, Lorig K, Whitelaw N, et al. The impact of chronic disease self-management programs: healthcare savings through a community-based intervention. BMC Public Health. 2013 Dec 6;13:1141.
9. Cistola AS, Bak AN, Guyer L, Reed A, Rooks B, Chacko L. Approaches for improving health equity: the effect of the gainesville community resource paramedic program on reducing emergency department utilization. 2020 Oct 8;
10. Sheikh A, Sood HS, Bates DW. Leveraging health information technology to achieve the “triple aim” of healthcare reform. J Am Med Inform Assoc. 2015 Jul;22(4):849–856.
11. Davies S, Schultz E, Raven M, Wang NE, Stocks CL, Delgado MK, et al. Development and validation of the agency for healthcare research and quality measures of potentially preventable emergency department (ED) visits: the ED prevention quality indicators for general health conditions. Health Serv Res. 2017 Oct;52(5):1667–1684.
12. Sentell TL, Seto TB, Young MM, Vawer M, Quensell ML, Braun KL, et al. Pathways to potentially preventable hospitalizations for diabetes and heart failure: a qualitative analysis of patient perspectives. BMC Health Serv Res. 2016 Jul 26;16:300.

13. Zhou RA, Baicker K, Taubman S, Finkelstein AN. The Uninsured Do Not Use The Emergency Department More-They Use Other Care Less. Health Aff (Millwood). 2017 Dec;36(12):2115–2122.

14. Enard KR, Ganelin DM. Exploring the Value Proposition of Primary Care for Safety-Net Patients Who Utilize Emergency Departments to Address Unmet Needs. J Prim Care Community Health. 2017 Oct;8(4):285–293.

15. Dowd B, Karmarker M, Swenson T, Parashuram S, Kane R, Coulam R, et al. Emergency department utilization as a measure of physician performance. Am J Med Qual. 2014 Apr;29(2):135–143.

16. Giannouchos TV, Kum H-C, Foster MJ, Ohsfeldt RL. Characteristics and predictors of adult frequent emergency department users in the United States: A systematic literature review. J Eval Clin Pract. 2019 Jun;25(3):420–433.

17. Emergency Medical Treatment & Labor Act (EMTALA) | CMS [Internet]. [cited 2021 Aug 16]. Available from: https://www.cms.gov/Regulations-and-Guidance/Legislation/EMTALA

18. Finkelstein A, Zhou A, Taubman S, Doyle J. Health Care Hotspotting - A Randomized, Controlled Trial. N Engl J Med. 2020 Jan 9;382(2):152–162.

19. Samuels-Kalow ME, Ciccolo GE, Lin MP, Schoenfeld EM, Camargo CA. The terminology of social emergency medicine: Measuring social determinants of health, social risk, and social need. Journal of the American College of Emergency Physicians Open. 2020 Oct;1(5):852–856.

20. Centers for Disease Control (CDC) NC for CDP and HP. Chronic Diseases in America [Internet]. Chronic Diseases in America. 2021 [cited 2021 Aug 15]. Available from: https://www.cdc.gov/chronicdisease/resources/infographic/chronic-diseases.htm

21. Mainous AG, Tanner RJ, Jo A, Park K, Beau De Rochars VM. Trends in Cardiovascular Disease Risk in the U.S., 1999-2014. Am J Prev Med. 2018 Jun 27;55(3):384–388.

22. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. JAMA. 2004 Mar 10;291(10):1238–1245.

23. Department of Health and Human Services. About the Affordable Care Act [Internet]. About the ACA. 2021 [cited 2021 Aug 16]. Available from: https://www.hhs.gov/healthcare/about-the-aca/index.html

24. Sanz H, Valim C, Vegas E, Oller JM, Reverter F. SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. BMC Bioinformatics. 2018 Nov 19;19(1):432.

25. Escanilla NS, Hellerstein L, Kleiman R, Kuang Z, Shull JD, Page D. Recursive feature elimination by sensitivity testing. Proc Int Conf Mach Learn Appl. 2018 Dec;2018:40–47.

26. Peng Y, Wu Z, Jiang J. A novel feature selection approach for biomedical data classification. J Biomed Inform. 2010 Feb;43(1):15–23.

27. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS One. 2020 Jun 12;15(6):e0234722.

28. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 2005 Aug;27(8):1226–1238.

29. Hapfelmeier A, Ulm K. A new variable selection approach using Random Forests. Comput Stat Data Anal. 2013 Apr;60:50–69.

30. Chandrashekar G, Sahin F. A survey on feature selection methods. Computers & Electrical Engineering. 2014 Jan;40(1):16–28.

31. Song F, Guo Z, Mei D. Feature selection using principal component analysis. 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization. IEEE; 2010. p. 27–30.

32. Roffo G, Melzi S. Features selection via eigenvector centrality. Proceedings of new frontiers in mining complex patterns (NFMCP 2016)(Oct 2016). 2016;

33. Cistola AS. FractureProof: An open source artificial intelligence process for precision public health. OSF; 2020 [cited 2020 Dec 4]. Available from: https://osf.io/dj6ub/

34. Cistola AS. Finding Value: Community and Quality Predictors of Value Based Payment Adjustments Among Florida Hospitals. Academy Health; 2021.

35. Weaver A, Lapidos A. Mental Health Interventions with Community Health Workers in the United States: A Systematic Review. J Health Care Poor Underserved. 2018;29(1):159–180.

36. Nejtek VA, Aryal S, Talari D, Wang H, O'Neill L. A pilot mobile integrated healthcare program for frequent utilizers of emergency department services. Am J Emerg Med. 2017 Nov;35(11):1702–1705.

37. Misner D. Community paramedicine: part of an integrated healthcare system. Emerg Med Serv. 2005 Apr;34(4):89–90.

38. O'Meara P, Stirling C, Ruest M, Martin A. Community paramedicine model of care: an observational, ethnographic case study. BMC Health Serv Res. 2016 Feb 2;16:39.

39. O'Meara P. Community paramedics: a scoping review of their emergence and potential impact. Int Paramed Pract. 2014 May 21;4(1):5–12.

40. Langabeer JR, Champagne-Langabeer T, Alqusairi D, Kim J, Jackson A, Persse D, et al. Cost-benefit analysis of telehealth in pre-hospital care. J Telemed Telecare. 2017 Sep;23(8):747–751.

41. Barnes AJ, Cunningham PJ, Saxe-Walker L, Britton E, Sheng Y, Boynton M, et al. Hospital use declines after implementation of virginia medicaid's addiction and recovery treatment services. Health Aff (Millwood). 2020 Feb;39(2):238–246.

42. Cunningham A, Mautner D, Ku B, Scott K, LaNoue M. Frequent emergency department visitors are frequent primary care visitors and report unmet primary care needs. J Eval Clin Pract. 2017 Jun;23(3):567–573.

43. Langabeer JR, Gonzalez M, Alqusairi D, Champagne-Langabeer T, Jackson A, Mikhail J, et al. Telehealth-Enabled Emergency Medical Services Program Reduces Ambulance Transport to Urban Emergency Departments. West J Emerg Med. 2016 Nov;17(6):713–720.

44. Chellappa DK, DeCherrie LV, Escobar C, Gregoriou D, Munjal KG. Supporting the on-call primary care physician with community paramedicine. Intern Med J. 2018 Oct;48(10):1261–1264.

45. Bennett KJ, Yuen MW, Merrell MA. Community paramedicine applied in a rural community. J Rural Health. 2018;34 Suppl 1:s39–s47.

46. Mason S, Knowles E, Colwell B, Dixon S, Wardrope J, Gorringe R, et al. Effectiveness of paramedic practitioners in attending 999 calls from elderly people in the community: cluster randomised controlled trial. BMJ. 2007 Nov 3;335(7626):919.