Analyzing the Effectiveness of COPD Drugs Through Statistical Tests and Sentiment Analysis

Indra kiran Chowdavarapu, Oklahoma State University; Dr. Scott Shepherd, OSU CHSI; Dr. Miriam McGaugh, Oklahoma State University

ABSTRACT

Chronic Obstructive Pulmonary disease (COPD) is the third leading causing of death in the United States. An estimated 24 million people suffer from COPD, and the medical costs associated with it stands at a whopping $36 billion. Besides the emotional and physical impact, a patient with COPD has to undergo severe economic burden to pay for the medication. At this juncture, identifying the best medicine to treat COPD enhances the living conditions of patients. This paper deals with analyzing the effectiveness of three popular drugs prescribed for the COPD patients through statistical tests and Sentiment analysis. The statistical analysis determines the effectiveness of these drugs on the patients in terms of mortality rates and readmission within 30 days of discharge. The impact of comorbidities, such as cardiovascular diseases, accident history, and smoking, on COPD patients is also examined. The data consists of 1 million patient encounter records obtained through Cerner Health Facts data. Base SAS is used to perform statistical analysis, combine multiple data sets to obtain each patient’s hospital records, compute readmission within 30 days’ information using lag function, and perform descriptive analysis. This paper also includes text mining of patients’ reviews about the drugs on drug portals and social media. The results obtained through sentiment analysis are then compared with the results of statistical analysis obtained earlier to determine the effectiveness of drugs prescribed to the COPD patients.

INTRODUCTION

Chronic Obstructive Pulmonary Disease, is a chronic lung disease that makes it hard to breathe because less air flows in and out of the airways in your lungs. When you’re getting less air, less oxygen gets into body tissues and it gets harder to get rid of the waste gas carbon dioxide. This results in shortness of breath during everyday activities. People with COPD can experience fatigue, chronic cough and frequent respiratory infections as well. COPD doesn't just have a physical impact—living with chronic disease also can affect mental health. The current research is aimed at determining three objectives:

1. Analyze the effectiveness of 3 COPD drugs: Advair, Symbicort, Spiriva, the combination of Advair/Spiriva and Symbicort/Spiriva. The effectiveness of these drugs are tested in terms of patients
   i. Length of Stay in the hospital
   ii. Readmission within 30 days of discharge
   iii. Mortality from COPD

2. Understand the effect of various Comorbidities on the outcome of COPD patients. The Comorbidities tested in this research are Accidents, Heart diseases and Smoking. The influence of these Comorbidities are seen in terms of COPD patients
   i. Length of Stay in the hospital
   ii. Readmission within 30 days of discharge
   iii. Mortality from COPD and Comorbidity

3. Sentiment analysis on drug portals to determine the sentiments of patients using Advair, Symbicort and Spiriva.
DATA PREPARATION

The Cerner Health Facts data set used in this project is a comprehensive collection of Patient information, Discharge information, Hospital information, Medication information of patients diagnosed with COPD. The ICD codes used for the analysis are given below:

<table>
<thead>
<tr>
<th>DIAGNOSIS_CODE</th>
<th>DIAGNOSIS_DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>490</td>
<td>Obstructive Chronic Bronchitis with (Acute) Exacerbation</td>
</tr>
<tr>
<td>491</td>
<td>Other Emphysema</td>
</tr>
<tr>
<td>491.1</td>
<td>Other Emphysema</td>
</tr>
<tr>
<td>491.2</td>
<td>Asthma, Unspecified</td>
</tr>
<tr>
<td>491.21</td>
<td>Asthma, Unspecified</td>
</tr>
<tr>
<td>491.22</td>
<td>Chronic Airway Obstruction, Not Elsewhere Classified</td>
</tr>
<tr>
<td>491.8</td>
<td>Obstructive Chronic Bronchitis with (Acute) Exacerbation</td>
</tr>
<tr>
<td>491.9</td>
<td>Other Emphysema</td>
</tr>
<tr>
<td>492</td>
<td>Obstructive Chronic Bronchitis with (Acute) Exacerbation</td>
</tr>
<tr>
<td>492.8</td>
<td>Asthma, Unspecified</td>
</tr>
</tbody>
</table>

Table 1. Diagnosis codes used in the analysis

The date set consists of a total of 593,932 patient records. These records include patients with all the medications prescribed for the patients admitted with COPD. For the current scope of study, we have identified the patients who have been given either of Advair, Symbicort, Spiriva or combination of Advair/Symbicort with Spiriva. Some of the patients were prescribed these medications for other ailments other than COPD. To eliminate this problem, the patient records are extracted based on the above ICD codes. The final data set consists of 83,395 unique patient records have used the mentioned drugs with diagnostic codes related to analysis. Figure 2 represents the selection process of the data for the purpose of this study.

Figure 2. Overview of data used in the study
**TARGET VARIABLES**

For the current analysis, three target variables are created from the existing data set

1. **Length of stay of patient:**

   The variable length of stay is calculated by subtracting the discharge date of patient with his admitted time. The length of stay is calculated at the level of patient’s each encounter to the hospital but not on overall visits of patient to the hospital.

2. **Mortality flag for Mortality analysis:**

   The variable Mortality_Flag is extracted from the column Discharge_disposition_id in the date set. The Discharge_disposition_id is a categorical variable that contains integer numbers for each of the discharge types. For eg, the discharge to Skilled Nursing facility is denoted by integer 12. In the current analysis, the main focus is on the final status of the patient irrespective of the discharge facility patient has been prescribed. In order to achieve that, all the patients who are discharged from the hospital are grouped as Target=0. The patients who were expired during or after the treatment are categorized into Target=1. Some of the levels of target variables can be seen below.

   ![Figure 2. Description of Mortality_flag variable](image_url)

3. **Readmission flag for readmission analysis:**

   The variable readmit30 is created from the existing data set. The variable is a binary variable that defines the readmission status of the patient with COPD.

   - Readmit30='0' => The patient is not readmitted into the hospital within 30 days of discharge from the hospital
   - Readmit30='1' => The patient is readmitted into the hospital within 30 days of discharge from the hospital

   The below analysis is used to identify patients who are readmitted to hospital within 30 days of discharge from the hospital. The steps followed for the analysis is shown below:

   ```
   /*Calculating Length of Stay and converting date format to calculate readmission*/
   data &out ;
   set &in ;
     date_dis = datepart(discharged_dt_tm);
     /*format date mmddyy10.;*/
     date_adm = datepart(admitted_dt_tm);
     /*format date mmddyy10.;*/
     length_stay = date_dis - date_adm;
   run;
   ```
/* Identifying patients with readmission within 30 days of discharge*/
data copd.indra_rank;
set &out;
Rank = 1;
run;

PROC SORT DATA=copd.indra_rank;
  BY patient_sk DESCENDING date_adm date_dis;
RUN ;

data copd.indra_seq;
  set copd.indra_rank;
  by patient_sk;
  if first.patient_sk then Seq_No=0;
Seq_No+rank;
run;

PROC SORT DATA=copd.indra_seq OUT=copd.indra_seq ;
  BY patient_sk Seq_No;
RUN ;

data copd.indra_seq;
set copd.indra_seq;
by patient_sk Seq_No;
ref_dt_id=LAG(date_adm);
 gap = ref_dt_id - date_dis;
length_of_stay = date_dis - date_adm;
if First.patient_sk then do;
  ref_dt_id = .;
  Gap = .;
  readmit30 = .;
End;
If 0 <= Gap <= 30 then
readmit30 = 1; /* Identify a readmission and assign value 1 to readmit30;*/
else readmit30 = 0;
Run;
STATISTICAL TESTS TO MEASURE EFFECTIVENESS OF DRUGS

The drugs used in the current study are Advair, Symbicort, Spiriva and combination of Advair/Symbicort with Spiriva. To eliminate the biases, each combination of drug is considered individually. Patient who has taken combination of Advair and Spiriva is considered different from patient who has taken only Advair or Symbicort. The distribution of medications on various measures can be seen in the figure 2.

![Distribution of Drug Percentages](image)

Figure 2. Distribution of drug percentages among the patients

1. LENGTH OF STAY:

The length of stay is computed for each patient's encounter to the hospital. ANOVA test is performed to understand the effectiveness of drugs on patient's length of stay. The ANOVA test compares the means between the groups you are interested in and determines whether any of those means are statistically significantly different from each other. Specifically, it tests the null hypothesis:

\[ H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k \]

where \( \mu = \) group mean and \( k = \) number of groups. If, however, the one-way ANOVA returns a statistically significant result, we accept the alternative hypothesis (HA), which is that there are at least two group means that are statistically significantly different from each other.

/*ANOVA test to determine effectiveness of medicines*/
```
proc anova data=&data;
    class medication;
    model length_of_stay=medication;
    means medication/ duncan;
    run;
```
The p-value for the ANOVA is less than 0.05 which means at least two groups have means different from others.

Post-hoc Tests:

The one-way ANOVA is an omnibus test statistic and cannot tell you which specific groups were statistically significantly different from each other, only that at least two groups were. To determine which specific groups differed from each other, you need to use a post hoc test. For this analysis, Duncan’s test is used to determine the differences between the groups.

This Post-hoc test determines if there exists difference between certain groups of Medication. As seen from the output, there is no statistically significant difference for Spiriva and Symbicort. The combination of Symbicort and Spiriva, Advair and Spiriva and Advair are significantly different from each other. The medications Symbicort combined with Spiriva has the highest length of stay for a patient with an average of 6.54 days per visit.

Out of all the medications, Advair has the best length of stay with an average of 5.72 days per visit.

Assumptions of ANOVA:

To use ANOVA results, we need to test if it is satisfying various assumptions. One of the assumption tested in this analysis is to verify if there is any heteroscedasticity in the data.

/*Testing and adjusting for unequal variances (heteroscedasticity)*/

```sas
proc anova data=&data;
  class medication;
    model length_of_stay = medication;
    means medication / hovtest welch;
run;
```
2. MORTALITY ANALYSIS:

Total Number of Patients considered: 83394 (130986 encounters)
No. of patients discharged: 80044 (96.7%)
No. of patients expired: 3350 (3.3%)

Chi-square test for the relationship between mortality and medications:

The Chi-square test is used to test if any relationship exists between two categorical variables. In our current analysis, Chi-square test can be used to understand if type of drug prescribed for the patient has any outcome in improving patient's chances of recovery from COPD.

Assumptions:

Null Hypothesis: There is no association between Drug administered and mortality of the patient.
Alternate Hypothesis: There is an association between Drug administered and mortality of the patient.

/*chi-sq tests on Mortality*/
proc freq data=&data;
tables mortality_flag * medication/chisq norow nopercent;
run;

<table>
<thead>
<tr>
<th>Frequency Col Pct</th>
<th>Table of mortality_flag by medication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>medication</td>
</tr>
<tr>
<td></td>
<td>advair</td>
</tr>
<tr>
<td>mortality_flag</td>
<td>advair</td>
</tr>
<tr>
<td>0</td>
<td>27179</td>
</tr>
<tr>
<td></td>
<td>96.44</td>
</tr>
<tr>
<td>1</td>
<td>1002</td>
</tr>
<tr>
<td></td>
<td>3.56</td>
</tr>
<tr>
<td>Total</td>
<td>28181</td>
</tr>
</tbody>
</table>
The chi-square value is 0.001 which is less than the p-value of 0.05. It means with 95% confidence we can say that, there is an association between the type of drug administered to the patient and the chances of survival of patient.

**Multiple Comparison tests on proportions:**
Tukey's multiple comparison test is performed to determine if there exist any differences between different groups of medicines on mortality.

Out of the various medications, the differences in proportions for 1 vs 5 (Advair vs Symbicort Spiriva), 1 vs 2 (Advair vs Advair Spiriva) and Advair vs Spiriva are significant.

**ODDS RATIO:**
An odds ratio (OR) is a measure of association between an exposure and an outcome. The OR represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure.

```sas
/*ODDS Ratio for Mortality Analysis*/
Proc logistic data=indra.medicine_mortality_final;
class medication(ref = 'advair');
model mortality_flag(event='1') = medication;
run;
```
The Odds of Mortality (‘1’) is 22% more for combination of Advair, Spiriva than for Advair alone. Based on all the odds, Advair has the less outcome for Mortality.

**Male vs Female:**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>medication advair spiriva vs advair</td>
<td>1.227</td>
<td>1.117 1.348</td>
</tr>
<tr>
<td>medication spiriva vs advair</td>
<td>1.195</td>
<td>1.086 1.315</td>
</tr>
<tr>
<td>medication symbicort vs advair</td>
<td>1.100</td>
<td>0.979 1.236</td>
</tr>
<tr>
<td>medication symbicort spiriva vs advair</td>
<td>1.308</td>
<td>1.145 1.493</td>
</tr>
</tbody>
</table>

**Urban vs Rural:**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>URBAN_RURAL_STATUS Rural vs Urban</td>
<td>1.138</td>
<td>1.041 1.244</td>
</tr>
</tbody>
</table>
3. READMISSION ANALYSIS:

Total Number of Encounters considered: 130004

No. of patients discharged: 118112 (90.85%)

No. of patients readmitted within 30 days: 11892 (9.14%)

A total of 11892 patients were readmitted within 30 days of discharge from hospital. This comprises of 9.14% of total patients admitted into hospital with COPD.

Chi-square test for the relationship between two categorical variables:

<table>
<thead>
<tr>
<th>readmit_flag</th>
<th>advair</th>
<th>advair spiriva</th>
<th>spiriva</th>
<th>symbicort</th>
<th>symbicort spiriva</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>39623</td>
<td>30079</td>
<td>25035</td>
<td>13851</td>
<td>9524</td>
<td>118112</td>
</tr>
<tr>
<td></td>
<td>91.73</td>
<td>89.06</td>
<td>91.92</td>
<td>91.79</td>
<td>88.96</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3574</td>
<td>3696</td>
<td>2201</td>
<td>1239</td>
<td>1182</td>
<td>11892</td>
</tr>
<tr>
<td></td>
<td>8.27</td>
<td>10.94</td>
<td>8.08</td>
<td>8.21</td>
<td>11.04</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>43197</td>
<td>33775</td>
<td>27236</td>
<td>15090</td>
<td>10706</td>
<td>130004</td>
</tr>
</tbody>
</table>

Figure 11. Chi-square test on Readmit within 30 days and Medication

The chi-square value is 0.001 which is less than the p-value of 0.05. It means with 95% confidence we can say that, there is an association between the type of drug administered to the patient and the chances of readmission of patient within 30 days of discharge.

Post Hoc Test:
The following post hoc test determines the three drugs Spiriva, Spiriva Symbicort and Advair to be Statistically significant.

**ODDS RATIO:**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>medication advair spiriva vs advair</td>
<td>1.362</td>
<td>1.298 1.430</td>
</tr>
<tr>
<td>medication spiriva vs advair</td>
<td>0.975</td>
<td>0.922 1.030</td>
</tr>
<tr>
<td>medication symbicort vs advair</td>
<td>0.992</td>
<td>0.927 1.061</td>
</tr>
<tr>
<td>medication symbicort spiriva vs advair</td>
<td>1.376</td>
<td>1.284 1.475</td>
</tr>
</tbody>
</table>

**Figure 13. Odds ratio of medications in determining readmission**

The Odds ratio determines Spiriva to have the least odds for the readmission.

**Male vs Female:**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER Male vs Female</td>
<td>1.253</td>
<td>1.169 1.343</td>
</tr>
</tbody>
</table>

**Figure 14. Odds ratio of gender in determining readmission**

**Urban vs Rural:**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>URBAN_RURAL_STATUS Rural vs Urban</td>
<td>1.138</td>
<td>1.041 1.244</td>
</tr>
</tbody>
</table>

**Figure 15. Odds ratio of location in determining readmission**
II. STATISTICAL TESTS TO MEASURE IMPACT OF COMORBIDITIES

The Comorbidities tested in this research are Accident_history, Heart diseases and Smoking disorder. The influence of these Comorbidities are measured in terms of length of stay, readmission within 30 days and mortality of COPD patients.

![Comorbidity % among Patient Population](image)

**Figure 16. Percentage of Comorbidities among COPD patient population**

1. LENGTH OF STAY:

ANOVA test is performed to understand the impact of comorbidities on patient’s length of stay.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>116669.127</td>
<td>19444.854</td>
<td>431.52</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Error</td>
<td>74734</td>
<td>3367584.719</td>
<td>45.061</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>74740</td>
<td>3484253.846</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 17. ANOVA test on comorbidities in determining length of stay**

The p-value for the ANOVA is less than 0.05 which means at least two groups have means different from others.
Post-hoc Tests:

Figure 18. Output of Duncan’s multiple range test
The Comorbidities are significantly different from each other. The Comorbidity with Accident_history has the highest length of stay and Smoking disorder has the least length of the stay.

2. EFFECT OF COMORBIDITIES ON MORTALITY:

Chi-square test for the relationship between mortality and comorbidities:

Figure 19. Chi-square test on mortality_flag and comorbidities
The Chi-square value is significant. There is association between mortality and Comorbidity.
Tukey’s Post-hoc Tests:

![Tukey's Post-hoc Tests](image)

**Figure 20. Output for Tukey's Post hoc test**

Based on the Tukey’s test on multiple proportions, the Comorbidities Smoking, Heart-Smoking, Smoking-Accident and Accident have significant association.

**ODDS RATIO:**

![Odds Ratio Table](image)

**Figure 21. Odds ratio of comorbidities in determining mortality**

The complication Heart disease and accident with COPD has the highest odds for mortality.
3. EFFECT OF COMORBIDITIES ON READMISSION:

Chi-square test for the relationship between readmission of COPD patients and comorbidities

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Table of readmit_flag by complication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>complication</td>
</tr>
<tr>
<td></td>
<td>Accident</td>
</tr>
<tr>
<td>readmit_flag</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

Figure 22. Chi-square test on readmit within 30 days and comorbidities

The Chi-square value is significant. There is association between readmission and Comorbidity.

Tukey’s Post-hoc Tests:

Figure 23. Output for Tukey’s post hoc test

The complication smoking has an association with Accident, Heart-Accident, Heart-smoking and Heart complications.
ODDS RATIO:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Point Estimate</th>
<th>95% Wald Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>complication Accident vs Heart</td>
<td>1.068</td>
<td>0.983 1.160</td>
</tr>
<tr>
<td>complication Heart Accident vs Heart</td>
<td>1.107</td>
<td>0.991 1.236</td>
</tr>
<tr>
<td>complication Heart Smoking vs Heart</td>
<td>1.012</td>
<td>0.931 1.101</td>
</tr>
<tr>
<td>complication Heart Smoking Accident vs Heart</td>
<td>0.984</td>
<td>0.831 1.166</td>
</tr>
<tr>
<td>complication Smoking vs Heart</td>
<td>0.863</td>
<td>0.810 0.919</td>
</tr>
<tr>
<td>complication Smoking Accident vs Heart</td>
<td>0.886</td>
<td>0.786 0.998</td>
</tr>
</tbody>
</table>

Figure 24. Odds ratio of comorbidities in determining readmission

Based on the odds ratio, the complication Heart Accident has the highest odds for readmission and the smoking has the least odds for readmission.

III. SENTIMENT ANALYSIS ON COPD DRUG REVIEWS

The ubiquitous presence of internet has given rise to explosion of data coming from the people. To analyze the functioning of a product or a medicine, we can leverage the experiences of the functioning of the medicine straight out of the Patients words. Understanding the Patients sentiments associated with medications help Doctors to take better care of their patients.

For the purpose of this study, Concept links, Text Clustering, Rule Based models are implemented to understand the most commonly used terms by the patients with COPD

DATA PREPARATION:

Reviews of COPD patients using the drugs in the study are extracted to understand the patient's opinions on the drugs. 400 reviews were collected from patient's using Advair, 204 Patients reviews of Symbicort and 159 user reviews for Spiriva is collected. The reviews are extracted using a web crawler from the below websites:

- Iodine.com
- Drugs.com
- Rxlist.com
- Viewpoints.com
- Druglib.com

CONCEPT LINKS:

Concept linking is a way to find and display the terms that are highly associated with the selected term in the Terms table. The selected term is surrounded by the terms that correlate the strongest with it. The below are the various Concept links generated for various terms.
Figure 25. Concept links for Advair drug reviews

The most common terms associated with Advair can be understood through the above concept links. The word Advair is strongly associated with enjoy, breath and doctor. From this concept link, it is understandable that Advair eases the breathing of patients and makes their life easy. In other concept links, Advair is associated with causing symptoms, higher price, increased quality of life etc., The concept links are one of the ways to visualize the association between opinions of the patients.

TEXT CLUSTERING:

The process of dividing a data set into mutually exclusive groups so that the observations for each group are as close as possible to one another and different groups are as far as possible from one another. In SAS Text Miner, clustering involves discovering groups of documents that are more similar to each other than they are to the rest of the documents in the collection. When the clusters are determined, examining the words that occur in the cluster reveals the focus of the cluster. Forming clusters within the document collection can help you to understand and summarize the collection without reading every document. The clusters can reveal the central themes and key concepts that are emphasized by the collection.
Building Models to Classify Text as Positive/Negative:

SAS Sentiment Analysis Studio provides a way to quickly understand customers’/persons opinions and experiences across multiple channels. The objective of this analysis is to develop a Statistical/rule based model that automatically classifies a given file as Positive/Negative. The below portal is used to classify documents into Positive/Negative/Neutral and allow the Studio to learn from the given data.

After classifying the data, Statistical models can be built to predict the tone of the reviews. In the below model, one of the technique provided an accuracy of 54.55%.

Figure 26. Distribution of Text Clusters

Figure 27. Statistical model output for Advair reviews
RULE BASED MODELS:
This model extracts tokens from the given data files and present the tokens as Positive/Negative or Neutral tone. This gives a quick glimpse of the kind of words used by People when describing about the product.

**Positive keywords:**

<table>
<thead>
<tr>
<th>Type</th>
<th>Positive</th>
<th>Negative</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CLASSIFIER</td>
<td>run</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CLASSIFIER</td>
<td>wheeze</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CLASSIFIER</td>
<td>medicine</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CLASSIFIER</td>
<td>hope</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CLASSIFIER</td>
<td>notice</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>CLASSIFIER</td>
<td>inhaler</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>CLASSIFIER</td>
<td>blessing</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CLASSIFIER</td>
<td>level</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>CLASSIFIER</td>
<td>bit</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>CLASSIFIER</td>
<td>inhale</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>CLASSIFIER</td>
<td>love</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>CLASSIFIER</td>
<td>Deus</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>CLASSIFIER</td>
<td>life</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>CLASSIFIER</td>
<td>me</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>CLASSIFIER</td>
<td>always</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>CLASSIFIER</td>
<td>breath</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>CLASSIFIER</td>
<td>remember</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>CLASSIFIER</td>
<td>period</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>CLASSIFIER</td>
<td>wonderfully</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 28. Commonly used positive words on Advair*

**Negative keywords:**

<table>
<thead>
<tr>
<th>Type</th>
<th>Positive</th>
<th>Negative</th>
<th>Neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CLASSIFIER</td>
<td>drug</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>CLASSIFIER</td>
<td>pain</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>CLASSIFIER</td>
<td>problem</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>CLASSIFIER</td>
<td>severe</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CLASSIFIER</td>
<td>worse</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>CLASSIFIER</td>
<td>leg</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>CLASSIFIER</td>
<td>tooth</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CLASSIFIER</td>
<td>cramp</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>CLASSIFIER</td>
<td>vasa</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>CLASSIFIER</td>
<td>long</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>CLASSIFIER</td>
<td>night</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>CLASSIFIER</td>
<td>adver</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>CLASSIFIER</td>
<td>effect</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>CLASSIFIER</td>
<td>threat</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>CLASSIFIER</td>
<td>side effects</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>CLASSIFIER</td>
<td>use</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>CLASSIFIER</td>
<td>New</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>CLASSIFIER</td>
<td>pound</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>CLASSIFIER</td>
<td>gain</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 29. Commonly used negative words on Advair*

**Model Test:**
The model being built in the earlier phase can be used to understand the sentiments of users. This model is used to test the sentiments of reviews of the drugs extracted from the drug portals and categorize into positive/negative.
In the below example, out of 159 user reviews on Spiriva, 54 are categorized as Positive and 105 are categorized as negative articles.

Figure 30. Categorizing the user reviews based on their sentiment

Among Advair, Spiriva and Symbicort, Advair received highest positive reviews with 36.45%, 33.45% viewed Spiriva as positive and only 28% of the user reviews rated Symbicort as positive.
CONCLUSION

EFFECTIVENESS OF COPD DRUGS:

i. Length of Stay:
   - The average length of stay for patients with COPD is the least for patients who are prescribed Advair with 5.72 days for over 43197 patient encounters.
   - No significant difference between length of stay is observed between Symbicort and Spiriva.
   - The combination of Symbicort/Spiriva has the highest avg. length of stay with 6.54 days for over 10706 patient encounters

ii. Readmission within 30 days of discharge:
   - The patients who are administered Spiriva has less chances of readmission within 30 days of discharge compared to other drugs.
   - Patients using Spiriva has 25% less chances of readmitting compared to Advair, while Symbicort has 8% less chances of readmissions.
   - The combination of Symbicort/Spiriva has worst readmission outcome compared to other drugs with 37% more chance of readmission compared to Advair.

iii. Mortality from COPD
   - Patients administered with Advair has the best outcomes in terms of mortality from COPD. They are more likely to survive when administered Advair than other drugs.
   - Compared to Advair, Spiriva has 19% more chances of mortality, while it is 10% for Symbicort
   - The combination of Symbicort/Spiriva has the worst mortality outcome with 31% more chances of mortality compared to Advair.

EFFECT OF COMORBIDITIES ON COPD PATIENTS:

i. Length of Stay:
   - The avg. length of stay for patients with COPD is the least for patients who have Smoking disorder with 4.8 days for over 23000 patient encounters.
   - The Comorbidity of Accidents with COPD has the avg. length of stay with 7.9 days for over 8500 patient encounters.
   - The Patients with Comorbidities of Heart disease and Accidents with COPD has the highest avg. length of stay with 8.9 days for over 4000 patient encounters.

ii. Readmission within 30 days of discharge:
   - The Patients with Comorbidities of Heart disease and Accidents has more chances of readmission within 30 days of discharge compared to other drugs.
   - Compared to Heart disease, Patients with smoking disorder are 13.7% more likely to survive.
   - Overall, there is not a significant influence between different Comorbidities in Patients readmission chances.

iii. Mortality from COPD
   - The Patients with Comorbidities of Heart disease and Accidents has the worst outcome in terms of mortality from COPD. They are 85% more likely not to survive compared to single comorbidity of Heart disease.
   - Compared to Heart disease, Patients with smoking disorder are 55% more likely to survive.
• Of all the comorbidities, Patients with Smoking disorder are more likely to survive and Accident has more chances of mortality.

SENTIMENT ANALYSIS ON DRUGS:

• Of the three drugs, Advair has received more positive reviews compared to other drugs, while Spiriva has more negative reviews.

• Advair has more positive reviews in terms Asthma maintenance, while Symbicort has better reviews in terms of COPD maintenance.

LIMITATIONS OF THE STUDY

The current study did not include factors such as Patient’s condition at the time of admission, Patient’s age, COPD stage(severe/mild) etc., The study only examines the effectiveness of the three drugs and their combinations in determining the patient’s health outcome, which sometimes may not be sufficient in determining Patient’s health outcome.

REFERENCES


2) Lucie Blais, Amelie Forget, Sulabha Ramachandran. 2010. “Relative effectiveness of budesonide/formoterol and fluticasone propionate/salmeterol in a 1-year, population-based, matched cohort study of patients with chronic obstructive pulmonary disease (COPD): Effect on COPD-related exacerbations, emergency department visits and hospitalizations, medication utilization, and treatment adherence”.


ACKNOWLEDGMENTS

We thank SAS Global Forum 2017 conference committee for giving us an opportunity to present our work. We also thank Dr. William Paiva and Ms. Elvena Fong of Center for Health Systems Innovation at Oklahoma State University for granting us permission to publish our results. Special thanks to Dr. Goutam Chakraborty and Dr. Miriam McGaugh of Business Analytics department at Oklahoma State University for their continuous support and guidance.
CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Indra kiran Chowdavarapu
MS in Business Analytics
Oklahoma State University
Phone: 330-461-7511
Email: manikiran.indra@gmail.com
LinkedIn: www.linkedin.com/in/indrakiran

Indra kiran Chowdavarapu is a graduate student enrolled in Business Analytics at the Spears School of Business, Oklahoma State University. He has been working on healthcare analytics projects with the Center for Health Systems and Innovation, Tulsa as a data analytics research assistant since Jan 2016. He is a recipient of SAS Student Ambassador, 2017, highest student honor by SAS Institute. He is a SAS® Certified Advanced Programmer, and also a SAS® Certified Predictive Modeler. He has co-authored a paper presented at the SCSUG conference, San Antonio in 2016 and had a poster presentation at the SAS® Analytics conference, Las Vegas in 2016.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.