Prescribing of Opioids among Medical Professionals

Yonaton Heit

January 9, 2019

1 Objective

Using Medicare Part D data, the possibility of opioid abuse is explored from outliers detection of health care providers claims.

2 Introduction

2.1 Background

Drug related deaths have steadily increased in the few decades. In 2016, approximately 63,693 Americans died from drug-overdosing. Noting that there may be more than one drug involved, of the deaths, 42,249 (66.4%) involved at least one prescription and/or illegal opioid. Prescription opioid was involved in 17,087 (40.4%) of the opioid related deaths.[1] Due to their addictive nature, opioids are treated as controlled substances in the United States. Examples of opioids are heroin, oxycodone, hydrocodone, morphine, fentanyl, among others. Heroin is a schedule I drugs, deemed as having no medical use, unsafe to use, and illegal in all cases. Oxycodone, methadone, morphine are examples of schedule II opioids. Scheduling divides controls drugs into five categories used FDA and the DEA in order determine a drugs risk of abusing and level of medical acceptability. Lower number scheduling indicates higher potential of abuse and higher medical regulation. In the last few decades, there has been concern over the abuse of prescription opioids, in particular Oxycontin, an oxycodone hydrochloride salt which according to a 2005 study, it is the most heavily abused of the schedule II/III opioids.[2]

2.2 Case and Value Proposition:

In 2017, 17% of individuals in the US were prescribed one or more opioid; and in 2016, it was estimated that 11.5 million or 4.3% of US population over the age of twelve misused illegal or prescription opioids [3]. Not only does the misuse of opioids cost lives but there is also a financial cost. Based on 2013 data, it was estimated that the economic burden for the misuse of opioids was $78.5 billion. The health care costs to private insurance companies was $14.0 thousand per patient.[4] It is therefore in the interest of these companies to track the misuse to prescription opioids.

3 Data Wrangling

The opioid prescription data comes from a 2013 Medicare Part D dataset of the opioid prescription of health providers around the United States.[5] This dataset has the number of opioid and drug claim (new prescriptions and refills), the percentage of claims that are opoids, the zip code, the state, the speciality, and name of each health provider. It is important to note that since the data is of the drug claims to Medicare Part D, the patents that the drugs are prescribed for is an older population. Individuals on Medicare Part D are 65 or older and it is expected that they may require painkillers at a higher rate than the general public. This dataset contains 1,049,326 different health care providers. It was downloaded using the Socrata API module and the query limit set to 1.5 million so all data points would be included.

Several of the entries of the number of opioid claim and percent of claims that are opioids for health care provider were empty. These were assumed to be and filled with zeros. Two versions of this dataframe were created, one that only contained health care providers that prescribed opioids and one that included all health care providers. The complete dataframe contains 1,049,326 health care providers, the dataframe that only includes health care providers that prescribed opioids has 496,744 providers. There were 246 specialties in the compete dataset, this lowered to 169 when only considering those to prescribe opioids. Unless otherwise stated. data analysis used the dataframe only containing health care providers who made opioid claims.

In addition to the information from the 2013 Medicare data, we are also including attributes about the area in which the health care provider operates purposes of clustering. Clustering will allow better separate the data into categories for the purposes of anomaly detection. This is discuss much more in dept in Section 5.

The population of each zip code was reported in the 2010 census. The data from the census was complied and made available from an online resource.[6]. This dataset contained the zip code, city, state, longitude, latitude, and population of cities across the United States. The population of a city was used to estimate the population from which heath care providers draw their patients from. The population of a city was determined by summing the population of each zip code within the same city. After this, of the 1,049,326 health providers, 1836 (0.17%) were in areas of unknown population. Some of these operated outside the US and therefore not included in the census. In addition, some health care providers were in the zip codes missing from the census. These health care providers were removed from the data.

The longitude and latitude of several zip code in the 2010 census database were incorrect. Any zip code in the dataset that the given longitude and latitude were two degrees different from the median longitude and latitude for zip codes in the same city were assumed to be incorrect. The longitude and latitude of these zip codes were estimated from median longitude and latitude of the zip codes in the same city.

The median income of the city in which a health care provider was kaggle database provided by the Golden Oak Research Group.[7] This database gave multiple median income for zip codes most likely for zip codes that cross city borders. Since zip code is the value that is used to link attributes together, there cannot be more than one median income per city. Of the original 1,049,326 providers in the database of all health care providers, 42,164 (4.02%) were discarded because the city income could not be determined. Of the 496,744 providers database of only health care providers that prescribe opioids, 21,749 (4.03%) were discarded for the same reason.

The temperature of a city in which a health care provider resides was estimated from a kaggle database tracking world temperatures by Berkeley Earth[8]. Only the temperatures of 2013 were selected. The database only had temperatures for select cities in the United State. The temperature of zip code was estimated to be the temperature of the closest city in the temperature database. City distance was determined from the longitude and latitude. For each city, the 2013 average, minimum, and maximum was determined.

4 Statistics

4.1 Opioid Prescribed by Clinicians:

Opioid prescribed by clinicians Figure 1 examines the amount of opioid prescriptions of the ten top prescribers by specialty. Medicare Part D patients are prescribed opioids more frequently by family practice and internal medicine physicians than any specialist, 21.6 million and 18.9 million in total, respectively. The third most frequent prescribers of opioids, nurse practitioners, are responsible for far fewer opioid claims, 6.4 million. Interesting, only about 5.5% and 4.6% of the total number of prescriptions by family practice and internal medicine physicians respectively were opioids. Family practice and internal medicine physicians treat a broad range of illnesses and disorders and are generally the first clinicians that patients consult for any complaint. This can explain why these two groups are responsible for far more opioid claims but opioids only include a small portion of their prescriptions. Orthopedic surgery, physical medicine and rehabilitation, anesthesiology, and interventional pain management are the most frequent prescribers of opioids; with more than 40% of their prescriptions falling into that category. This would be expected because two treat patients in or recovering from pain (physical medicine and rehabilitation and interventional pain management) and two perform surgeries (orthopedic surgery and anesthesiology) for which opioids are commonly given for post-operative pain.

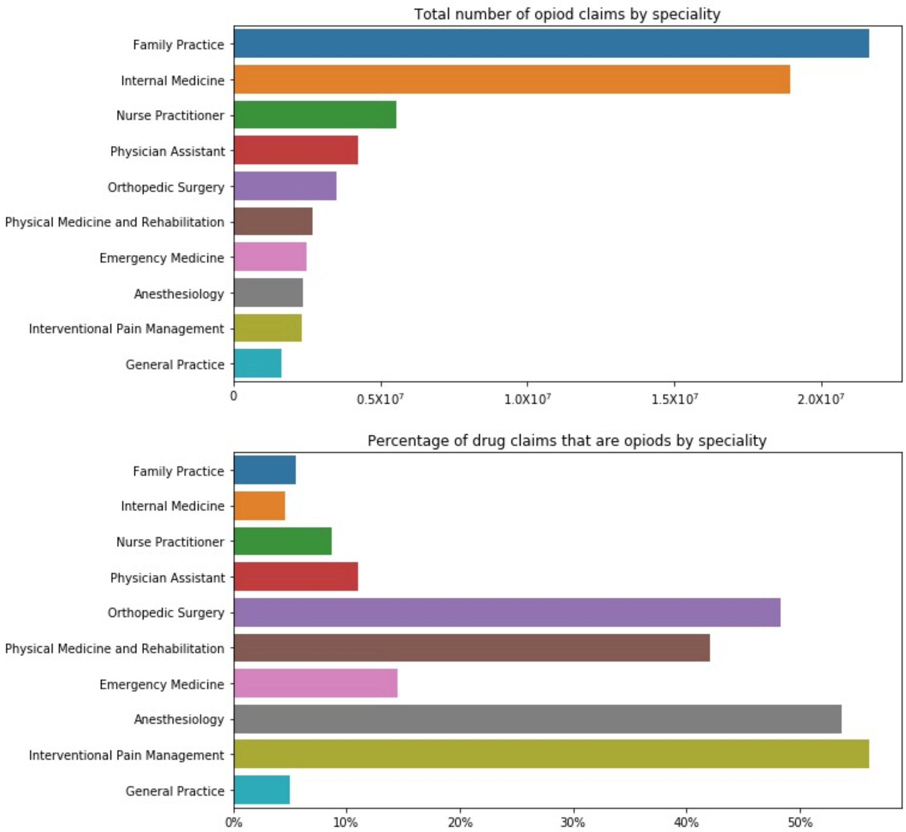

Figure 1: Prescriptions by clinician type. The top bar plot is the number of opioid claims of the 10 clinician categories who prescribe the most opioids. The bottom bar plot is percentage of the claims that are opioids by the same clinician categories.

Figure 2 examines the range of percentages of opioid prescribed by the specialists. While family practice, internal medicine, and general practice physicians have narrow spreads of their percentages (the whiskers in their boxes range from 0% to 14%), they have outliers that reach 100%. However, their outliers are only about 6% of the total number of health care providers in those categories. health care providers are considered outliers of their specialty if their percentage of their opioid claims are outside the first and third quartile by 1.5 the interquartile ranges (difference of the first to third quartile). Physical medicine and rehabilitation's boxplot has a whiskers span of 0% to 100% while orthopedic surgery, physician assistant, and anesthesiology's whiskers span most of this range. All of the top ten prescribers of opioids, as outliers or in the normal range of the plot, span 0 to 100%. Despite the wide range of values, only general practice had more than 10% of health care providers as outliers, at 11%. Nurse practitioner were slightly below at 9%. While it makes sense that specialties that deal with pain relief or surgery would have a wide range, that family practice and internal medicine health care providers have a significant number of outliers suggests that they are too lenient with opioid prescriptions

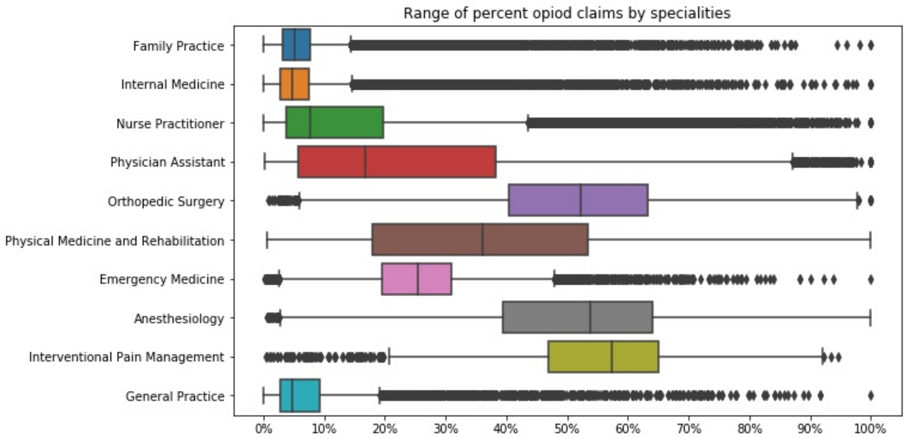

Figure 2: The range of percentage of claims that are opioids of individual health care provider separated by the specialties in Figure 3. The black line inside the colored box is the median. The edges of the colored box are the first and third quartile. The whiskers are located at a distance of 1.5 times the interquartile range (distance of the first and third quartile) for the colored box. Anything outside this range is considered an outlier.

5 Clustering

The boxplots in Figure 2 only considered specialty when separating the data. Other factors may influence the amount of opioid prescribed. For examine different geographic areas are expected to different levels of opioid prescription. In this section, clusters are create for each of the 10 specialties in Figure 1 and 2 based on multiple attributes discussed in Section 3 in order to separate data into multiple categories. Anomalies are determined based on the outliers of the cluster rather than the specialty alone. Attributes considered are location(longitude and latitude), city median income, temperature (min, max and mean), population, number of opioid claims, and percent of opioid claims. The number of total claims is left out because they can be completely dependent on number of opioid claims and percent of opioid claims. Having attributes that is completely co-dependant from other attributes is a poor practice for clustering.

5.1 Creating Cluster for each Specialty

Each attribute has different ranges and scales of values. For example, mean temperatures ranges from 1-24°C and median income ranges from \$0 to \$300000. In order to prevent attributes with larger variance from unduly weighting the destination of clusters, each attribute is standardized so that mean is zero and its standard deviation is 1. In addition, the attributes were decomposed with PCA in order to lower the number of attributes. It was found that of the 9 dimensions of the features, over 90% of the variance could be explained by the 6 dimensions. The decomposition revealed that the 4 of the mean temperature, min temperature, and latitude were highly correlated which could explain why 3 dimensions could be dropped.

Both elbow plots and silhouette scores were used to determine the optimal number of clusters to use for the K-means method. The elbow plot on Figure 3 was used to ballpark the number of clusters. Based on it, it was found that number of clusters for each specialty should be larger than 5. From this starting point ($k>5$), the silhouette score refined the value with the assumption that a larger score was a better fit.

5.2 Examination of the Clusters

Using boxplots, we can visualize the normal range of percent opioid claims of the clusters within a specialty. There were 3 distinct patterns of clustering for the specialties which is shown on Figure 4. Figure 4 also shows boxplot of opioid prescription without clustering. The first major pattern is that within a specialty, the clusters have similar medians and distribution of percent of opioids (though not identical) to the unclustered distribution. The exception is a single cluster that has a completely different median and range. The specialties that follow this pattern are family practice, internal medicine, emergency medicine and general practice. It is of interest that all these specialties prescribed a broad

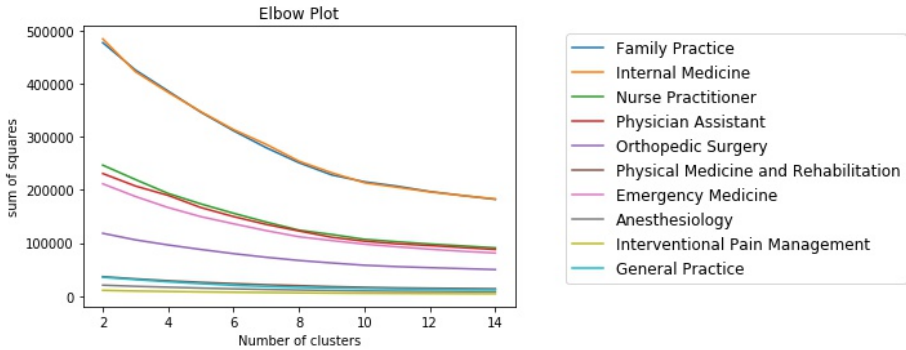

Figure 3: The elbow plot of the top 10 specialties that prescribe opioids. This embow plot indicates that larger number of clusters would be optimal. We restricted the number of clusters to be greater than 5 and refined it using the silhouette score.

set of drugs, not just opioids. A significant number of healthcare providers that were considered outliers without clustering are no longer outliers when placed into a cluster with a different range from the unclustered data. This has the effect of lowering the number of outliers with the except of emergency medicine. The single cluster with different range for emergency medicine range of values were lower than the other clusters. Since only higher opioid claim outliers are considered, clustering did not lower the number of considered outliers for emergency medicine. While clustering pattern would suggest that a lower number of clusters would suffice, this pattern was only seen for large cluster numbers (k>5).

The second cluster patterns seen for the specialties is having broad ranges in their boxplot. The range of the distribute of these unclustered specialties can covers approximately to 0% to 100% percent opioid claims which makes use of percent opioid claims useless for outliers detection. The boxplots of the clusters cover a smaller range but are still wide. Outlier detection for opioid fraud is difficult when distribution is wide since fewer data points are outside the expected range. Specialties that have this pattern are orthopedic surgery, physical medicine and rehabilitation, anesthesiology, and interventional pain management. All specialties that treat pain and would be expected include health care providers that heavily prescribe opioids.

The last cluster patterns seen having two distinct sets of similar clusters. One with a similar percent opioid boxplot to the uncluster and another set of clusters with a different range. Both nurse practitioner and physician assistant fall into this pattern. Nurse practitioners have 10 clusters, 8 with similar boxplots to the unclustered dataset and 2 similar to each other. Physician assistant have 11 clusters, 8 with similar boxplots to the unclustered dataset and 3 similar to each other.

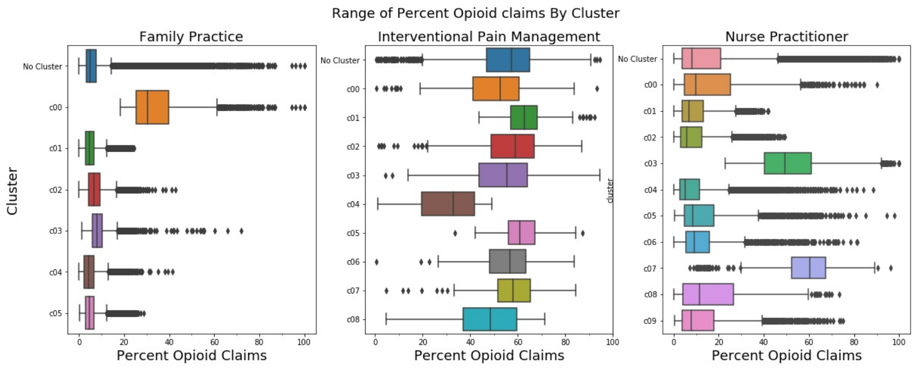

Figure 4: Each of the boxplots are represent the three clustering patterning of each specialty described in this section 5.2. The top boxplot for each set of clusters shows the unclustered data from that specialty. The boxplot below shows the data distribution for each cluster.

5.3 Check Clusters against Known Cases of Fraud

In 2018, the United States Department of Justice had a massive crackdown on the healthcare fraud [9]. 601 individuals were prosecuted, including 165 doctors, nurses and other licensed medical professionals. Of these charged, 162 were prosecuted for prescribing and distributing opioid and other narcotics. There is complete list of health care provider that have been changed. From various online articles on specific cases from this round of arrests, we were able to find 5 healthcare providers who were both persecuted for opioid fraud and were on the 2013 Medicare Part D database. One of these healthcare providers, Madhu Garg, was prosecuted before this crackdown in 2016.

Using the cluster boxplot on Figure 5, we examined whether these medical professionals were marked as outliers. These five healthcare providers are Bothra Rajendra (general surgery) [10], Eric Backo (physical medicine and rehabilitation)[10], Madhu Garg (family practice)[11], Lawrence Miller (family practice)[12], and Samson Orusa (internal medicine)[13]. General surgery, specialty practiced by Bothra Rajendra, was not one of the fields previously examined so the clustering process that has been described in Section 5.1 was performed. The field of general surgery's boxplots has a broad range of values in the distribution which makes outlier detection more difficult. Without clustering Bothra Rajendra percent opioid claims fall clearly inside the expected range and is not marked as an outlier (without clustering the boxplot distribution covers 0% to 100%). With clustering, Bothra Rajendra in a borderline datapoint (right below the top whisker). Both Madhu Garg and Samson Orusa were clear outliers within their cluster. Without clustering, Madhu Garg is right below the edge of the distribution. Lawence Miller is just below the edge of the distribution within his cluster making him a borderline case. Finally Eric Backos failed to be marked as an outlier; the median percent opioid claims of this cluster is higher the value found for Eric Backos. The final results are of the known cases of fraud, two where detected as an outlier, two were borderline, and failed to be detected.

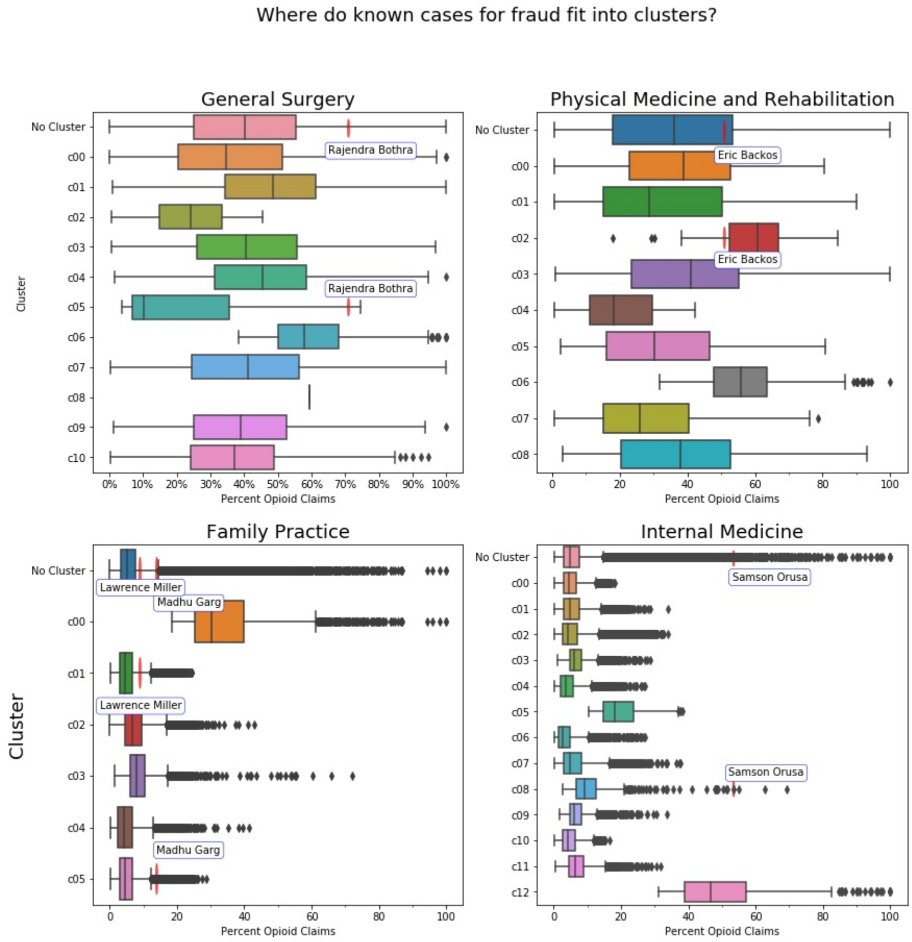

Figure 5: The five known cases of fraud where placed into the boxplot of their specialty. Similar to Figure 4, the top boxplot for each specialty is the unclustered and the rest are clustered.

6 Conclusion

The purpose of this project is to demonstrates the ability of clustering of healthcare providers in order to detect anomaly opioid prescription in hopes to determine possible cases of fraud. It is important to note that this analysis is limited by the lack of known causes of fraud. It would be more desirable to have a large cases of fraud to test the method. Since two of the five cases of known of fraud were borderline outliers, it may be warranted to examine the boundaries for determine better conditions of outliers. Considering the time scale of this project, its will have to be done in the future. In two of the cases, clustering did help identify them as outliers (or borderline outliers) where the unclustered model did not.

Based on results, these clusters can help identify outliers opioid prescribed by healthcare providers that may be fraudulent. It is recommended that outliers within their cluster (and borderline outliers) should be investigated for possible cases for fraud. This methodology was found to be more useful for specialties that have a narrow range of opioid claims such as family practice and internal medicine. For future consideration of anomaly detection, it may be beneficial to focus on the opioids that are commonly abused,such as oxycodone, rather than all opioids. An addition factor that should be consider is the dollar amount of claims which may be a strong indicator of fraud. The dollar amount of claims was considered by Gleb Esman who worked on a similar project and successfully identified fraudulent claims within the Interventional Pain Management which our model should struggle with due to the broad range of percent of opioid claims.[14]

References

[1] "Overdose Death Rates", 2018 Acessed: 31 October, 2018.

[2] Cicero, T. J.; Inciardi, J. A.; Muñoz, A. *The Journal of Pain* **2005,** *6,* 662–672.

[3] Hoots, B. H.; Seth, P. "Annual Surveillance Report of Drug-Relate", 2018 Acesssed: 31 October, 2018.

[4] Florence, C. S.; Zhou, C.; Luo, F.; Xu, L. *Medical Care* **2016,** *54,* 901–906.

[5] "Medicare Part D Opioid Prescriber Summary File 2013", 2018 Acesssed: 30 October, 2018 <https://data.cms.gov/Medicare-Claims/Medicare-Part-D-Opioid-Prescriber-Summary-File-201/yb2j-f3fp>.

[6] "US Zip Codes Database", 2018 Acesssed: 7 November, 2018 <https://simplemaps.com/data/us-zips>.

[7] "US Household Income Statistics", 2018 Acesssed: 24 December, 2018.

[8] "Climate Change: Earth Surface Temperature Data", 2016 Acesssed: 28 November 2018.

[9] "National Health Care Fraud Takedown Results in Charges Against 601 Individuals Responsible for Over $2 Billion in Fraud Losses", 2018 Acesssed: 1 January, 2019 <https://www.justice.gov/opa/pr/national-health-care-fraud-takedown-results-charges-against-601-individuals-responsible-over>.

[10] Snell, R. "Six Detroit-area doctors charged in $500M opioid scheme", 2018 Acesssed: 2 January, 2019 <https://www.detroitnews.com/story/news/local/macomb-county/2018/12/06/feds-allege-six-detroit-area-doctors-fueled-opioid-crisis-health-care-fraud-conspiracy/2225239002/>.

[11] Services, C. N. W. "Ex-Glendora Doctor Sentenced to Prison for Selling Pain Pills to Addicts in L.A.", 2016 Acesssed: 3 January, 2019 <https://patch.com/california/glendora/ex-glendora-doctor-sentenced-prison-selling-pain-pills-addicts-la>.

[12] Murray, M. "Prosecutors: 4 area doctors arrested in pill mill investigation", 2018 Acesssed: 3 January, 2019 <http://www.fox29.com/news/local-news/prosecutors-4-area-doctors-arrested-in-pill-mill-investigation>.

[13] Smith, C. "Who is Samson Orusa? Clarksville doctor facing opioid distribution, fraud charges", 2018 Acesssed: 3 January, 2019 <https://www.theleafchronicle.com/story/news/crime/2018/12/14/who-samson-orusa-clarksville-doctor-charged-opioid-fraud-scheme/2312021002/>.

[14] Esman, G. "Building a $60 Billion Data Model to Stop US Healthcare Fraud", Acesssed: 5 January, 2019 <https://www.splunk.com/blog/2017/09/28/building-a-60-billion-data-model-to-stop-us-healthcare-fraud-with-splunk-and-machine-learning.html>.