Project Report: Analysis of National Opioid Epidemic

Joseph Blankenship and Jacob Mitchell

Department of Computer Science, Bowling Green State University

CS 6010

Dr. Tian

December 12, 2021

Introduction

The opioid epidemic is a multifaceted problem in the U.S. The problem has roots in society, healthcare, and the economy. With the recent explosion in healthcare data, we can investigate more problems with data science (Lerner et al). As data grows, we may examine the problem from new angles and better develop strategies for analyzing it. If we can better understand some of the problems driving opioid deaths in the U.S., we can better direct resources and write more effective policies. This exploratory project addresses the opioid epidemic through the examination of public health data on opioid prescriptions through Medicare Part D. While there are correlations between opioid prescriptions and opioid overdose deaths, this project goes further to find relationships between opioid deaths and other public health metrics.

Problem Definition

With opioid deaths rising in part due to an increase in fentanyl use, it is important to consider where user opioid habits begin. A recent article by the New York Times, details a case where an elderly woman's opioid prescriptions were being stolen. A 2017 report from the U.S. Department of Health and Human Services, reveals that one in three Medicare Part D patients had an opioid prescription in 2016. Along with this, nearly 500,000 patients received a large amount of opioids (Nudelman, 2017). While only 3% of patients become addicted to opioids after being prescribed them, it is worth investigating if Medicare Part D opioid prescriptions have any effect on this crisis (Caution: These Are the Most, 2013). An elderly person deprived of their medication is not only a form of abuse but may unintentionally contribute to the opioid epidemic. If this is not the case, what other data points in public health data can help us better understand the problem?

Dataset and Background

The years studied included 2017, 2018, and 2019. Several datasets were used in this analysis. The opioid overdose deaths and the rate of death by opioids per state were downloaded from the Center for Disease Control (CDC) website. A dataset on Medicare prescription rates were downloaded from the Centers for Medicare and Medicaid Services. It includes rates of prescriptions by state and other related data. For general health information, ratings on different health criteria were downloaded from the United Health Foundation (UHF) website. It includes information on state smoking rates, alcohol use, education quality, mental health factors, and much more. Finally, state population estimates were downloaded from the U.S. Census Bureau.

An important distinction is made between long-acting opioid prescriptions (LA opioid prescriptions) and total opioid prescriptions. Long-acting opioids are those which release more slowly in the blood stream (Krans, 2021). They include drugs like oxycodone (Safe Use of Long-Acting). Oxycodone and other LA opioids are among the most addictive (Krans, 2021). LA opioid prescriptions, total prescriptions, and overdose deaths were normalized with state population data. General health statics through the United Health Foundation were already adjusted by state population. While 2017, 2018, 2019 was not the most recent data, they were the most readily available years. The UHF data included many public health statistics. These features were numbers between 0 and 30 that represented rates of occurrence in the given state. Table 1 shows the UHF features for our model:

UHF Feature	Description
SmokingValue	Smoking Rate
UnemploymentValue	Unemployment Rate
FreqPhysDistressValue	Rate of Frequent Physical Distress
ExcessiveDrinkingValue	Rate of Excessive Drinking
EducationLTHSValue	Rate of Education (Less than High School)
ECigValue	Rate of E-Cigarrette Use
DepressionValue	Overall Rate of Depression
Depression65UpValue	Rate of Depression of Individuals over 65
CancerValue	Cancer Rate

Table 1

Experimental Evaluation

A heatmap between variables revealed a positive correlation between Medicare LA opioid prescriptions and overdose deaths. There was also a positive correlation between Medicare Total opioid prescriptions and overdose deaths. The correlation was modeled with Linear Regression (LR) and Bayesian Ridge Regression (BRR), which is similar to regular ridge regression. Then the models were mapped to 2018 data and evaluated with mean squared error and R-squared metrics. After this, the UHF features were included in several multiple linear Regression (MLR). The models were visualized in order to find trends between public health features and opioid overdose deaths.

To better understand the many variables that contribute to the opioid epidemic, it was then necessary to use more complicated models. To investigate underlying trends between public health features, Medicare Part D prescriptions, and opioid deaths, we used K-Means clustering which includes principle component analysis (PCA). We created 9 models. There were three models for 2017, 2018, and 2019. Model UHF_Pres included UHF features with the prescription rate of opioids through Medicare Part D. Model UHF_NoPres included UHF features but not the prescription rate. Model UHF_Reduc included a smaller set of UHF features which were estimated to contribute to opioid deaths.

After creating the clusters, the quality of the clusters were examined for their relationship with normalized opioid deaths in each state. The death rate was added to each state in the clusters. Then the Standard Deviation (SD) of the death rates in the clusters were taken. A cluster with low SD indicated that the model performed well in identifying states with trends in the same direction. A cluster with high SD indicated worse performance. When a cluster was formed well, the features used for creating that cluster were a better indicators of death rate. Table 2 shows the features included in each model.

Model Features		
UHF_Pres	UHF_NoPres	UHF_Reduc
Prescription Rates	X	X
SmokingValue	SmokingValue	SmokingValue
UnemploymentValue	UnemploymentValue	UnemploymentValue
FreqPhysDistressValue	FreqPhysDistressValue	FreqPhysDistressValue
ExcessiveDrinkingValue	ExcessiveDrinkingValue	ExcessiveDrinkingValue
EducationLTHSValue	EducationLTHSValue	EducationLTHSValue
ECigValue	ECigValue	DepressionValue
DepressionValue	DepressionValue	
Depression65UpValue	Depression65UpValue	
CancerValue	CancerValue	

Table 2

The features included in the model were numerical values that represented each state by the number of people who fell into the following categories: depression, depression in people ages

65 and up, education less than high school, frequent physical distress, excessive drinkers, smokers, e-cigarette users, cancer patients, and unemployment rate.

Results

When the death rate was normalized for population, some of the top states were Florida, Ohio, Virginia, and Tennessee. A heat map showed a strong positive correlation between LA opioid prescriptions, total opioid prescriptions and opioid overdose deaths. A simple linear regression also showed a strong correlation. The BRR regression showed a similar relationship and preformed almost exactly the same in terms of R-squared. This was expected because there were just two variables. The results can be seen in Fig. 1,2 and in Table 3.

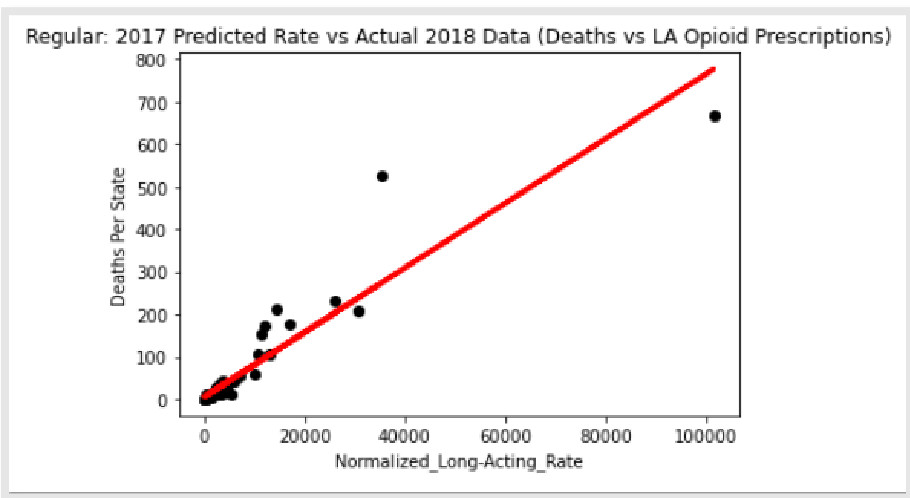

Fig 1.

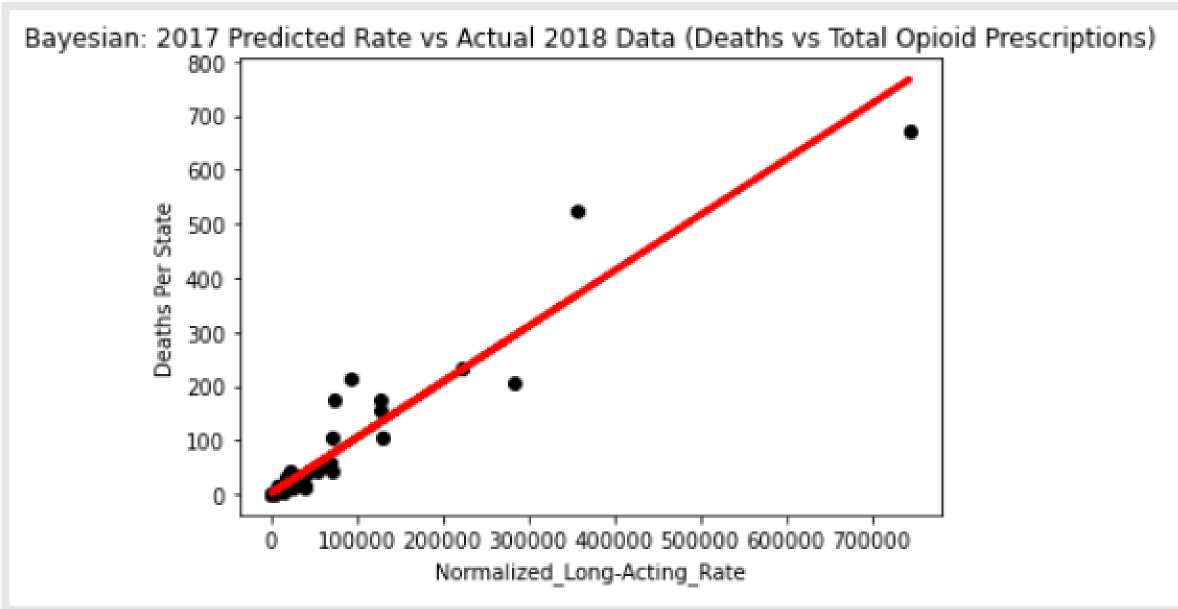

Fig 2

	Linear Regression	Bayesian Ridge Regression
MSE	20.5	20.52
R^2	0.9	0.9

Table 3

To investigate further, we included more features from UHF datasets. Fig. 3 and Fig.5 are samples of some of the MLR regressions done. When viewing the MLR results, it was still difficult to pinpoint underlying causes for opioid deaths. Cancer rates did not seem closely tied with opioid prescription rates. When education and depression rates were included, the majority of states had opioid death rates closely tied with these features. This did not necessary mean that these were important factors. It was clear a stronger multidimensional model was needed to find deeper meaning in the data.

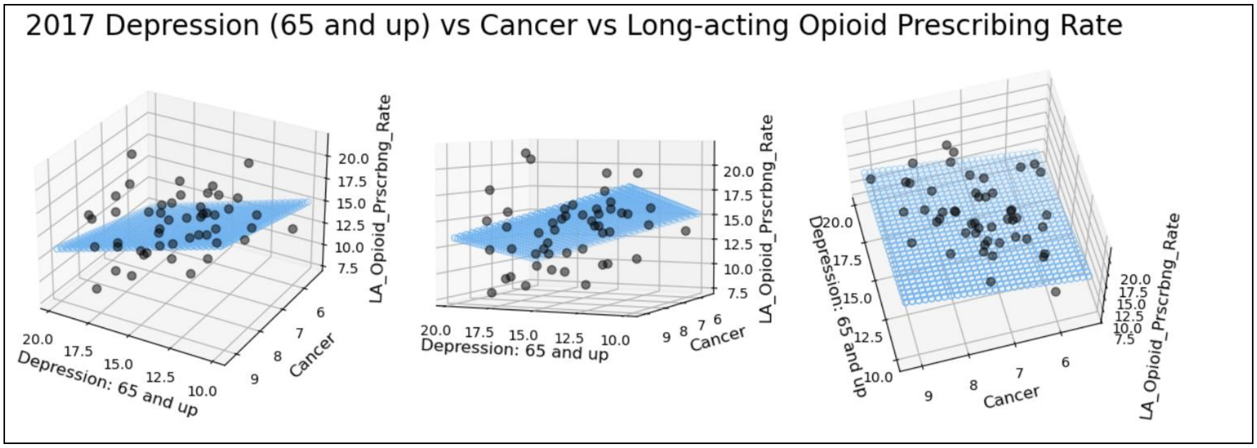

Fig 3

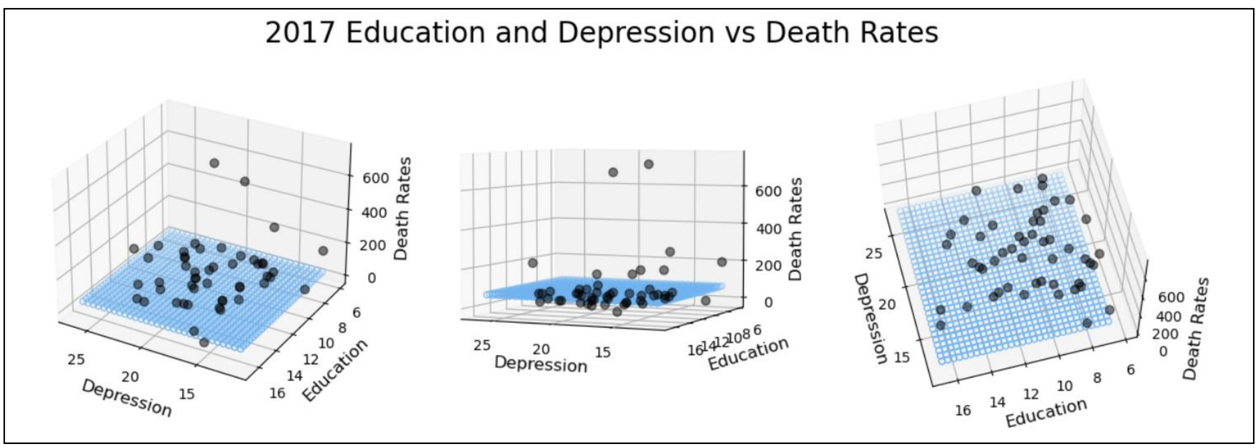

Fig. 4

In table 4, the SDs for each cluster in each model can be seen. California tended to cluster alone in the UHF_Pres model. The model clustered more evenly with 2019 data. We were not concerned with the number of clusters. We were concerned with the SD with each cluster or how well the states were grouped by opioid deaths. The models with opioid prescriptions had a lot of variability within clusters. This suggested that they did not cluster as well as the other models. The models without Medicare Part D prescription rates performed better in grouping the states to minimize the death rate SD. While there was variation between years, each model changed in a similar way when the year changed. The UHF_Pres model general generated more clusters. The SD was lower in the UHF_Reduc model.

	2017	2018	2019
UHF_Pres	cluster 0 584.053280 1 1051.690227 2 567.583401 3 453.744212 4 203.428601 5 California 6 195.868578 7 260.779418 Sum: 3317.14	cluster 0 69.784869 1 12.399756 2 329.528807 3 33.028222 4 California 5 296.586166 6 51.404086 7 101.528366 Sum: 898.23	cluster 0 9.374414 1 8.425843 2 12.066798 3 California 4 2.701855 5 8.980835 6 7.520873 7 10.486032 Sum: 59.55
UHF_NoPres	cluster 0 219.954663 1 59.578754 2 247.327593 3 87.626598 4 10.388956 Sum: 624.87	cluster 0 46.446198 1 132.624738 2 8.140505 3 64.531612 4 183.108677 Sum: 434.85	cluster 0 19.027059 1 18.967715 2 25.208095 Sum: 63.20
UHF_Reduc	cluster 0 59.578754 1 90.816391 2 177.667477 3 207.542288 4 7.570330 Sum: 543.17	cluster 0 46.446198 1 157.340532 2 34.240653 3 183.108677 Sum: 421.13	cluster 0 11.109719 1 8.425843 2 9.374414 3 7.821714 4 California 5 8.555769 Sum: 45.28

Table 4: SD of Normalized Death Rates by Year by Features

Discussion

While the LR models were useful in getting a sense of the data, the K-means model was much more effective in finding deeper trends in the data. Even though Medicare Part D opioid rates seem to correlate strongly with opioid deaths in the LR models, they were not as effective in predicting deaths in the K-means UHF_Pres model. The lower SD for the UHF_Reduc model in comparison to the UHF_NoPres model suggests that better features can be selected in the future to contribute to predicting opioid rates. Cancer rates, E-Cigarette rates, and rates of Depression for people over 65 were excluded from the UHF_Reduc model. These features did not correlate well with opioid overdose death rates.

Even though the clusters of the UHF_Reduc model formed with relatively low SD across states, it is not clear which features were driving this change. More analysis is necessary to understand which public health factors make large impacts on the opioid epidemic. If we remove some of the worse-hit states from our analysis, then Medicare prescription rates are a little more closely tied with opioid deaths. However, we did not discover other factors which may be driving this relationship.

An important factor which might have been driving the relationship was age. Some states have older populations. Almost all recipients of Medicare are over the age of 65. The relative age of a state can have a large influence on the state's economy. Another factor not considered was distribution of recovery facilities. If one state or another had more opioid recovery centers, then some relapse cases may end up staying within that state. The opioid problem is complex and more detailed models are needed.

Future Work and Conclusion

To better define some of the factors driving opioid overdose in worse-hit states, more detailed models are necessary. With more features, models may be able to better predict opioid deaths or rule out more confounding variables in the correlation between opioid deaths and Medicare Part D prescriptions. Work should continue with neural networks because of their ability to handle large feature matrixes and discover latent relationships. Age factored into these relationship including the average age of overdose, etc. There is also room for work in Medicaid prescriptions of opioids. Medicaid serves a much younger population so there may be big differences between the two. LR and clustering may be further used to get a sense of the problem from different angles, but will not provide more than just a glimpse of the whole picture.

With such a multifaceted problem facing the U.S., there is no single solution to fix it. There is also not a single best approach for analyzing it. Finding the sources of the problem to address can also be tricky. Through experimental analysis, we can narrow down which facets may play the largest role and isolate those problems to begin to alleviate opioid addiction and overdose. This project was an attempt at exploring the different levels of the United States opioid crisis. Through our analysis, we found that the sheer number of prescriptions written had a linear relationship with opioid related death. However, we could not uncover other factors driving this relationship. We also found that using clustering alongside certain features may be able to predict the number of opioid related deaths with some sense of accuracy. Clustering may be useful directing resources at a national level or at least drawing attention to states in need of resources. However, the complexity of this issue cannot be overlooked and results using more recent machine learning techniques will likely be more reliable.

Source code can be found here: https://github.com/JoeKBlank/Opioid_Analysis_Data_Science

References

"Caution: These Are the Most Addictive Pain Meds." *Harvard Health*, 1 Nov. 2013, https://www.health.harvard.edu/diseases-and-conditions/caution-these-are-the-most-addictive-pain-meds.

"Fentanyl." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 16 Feb. 2021, https://www.cdc.gov/opioids/basics/fentanyl.html.

Krans, Brian. "9 Of the Most Addictive Prescription Drugs on the Market." *Healthline*, Healthline Media, 23 July 2021, https://www.healthline.com/health/addiction/addictive-prescription-drugs.

Lerner, I., Veil, R., Nguyen, D.-P., Luu, V. P., & Jantzen, R. (2018). Revolution in health care: How will data science impact doctor–patient relationships? Frontiers in Public Health, 6, 99. https://doi.org/10.3389/fpubh.2018.00099

Nudelman, J. U.S. Department of Health and Human Services. (2017) HHS OIG Data Brief. Opioids in Medicare Part D: Concerns about Extreme Use and Questionable Prescribing.

"Safe Use of Long-Acting Opioids." *Safe Use of Long-Acting Opioids | Michigan Medicine*, https://www.uofmhealth.org/health-library/abo7833.