

# Analysis and Prediction of Opioid Crisis Based on Data Mining

Yuankun Chen

Huazhong University of Science and Technology

chenyuankun1108@163.com

**Keywords:** opioid, crisis, prescription, proliferation, socioeconomic factors

**Abstract:** The final report released by the White House Commission on combating drug addiction and the opioid crisis (2017), noted that opioid prescriptions have quadrupled since 1999 with more than 27 million people using them in non-prescription way. Opioid crisis breaks down geographic and demographic boundaries. This report calls on the government, various institutions and the masses to work together to overcome the opioid crisis. We hope to provide some effective advice through our efforts. First, we use four charts to visually analyze synthetic opioid and heroin cases in five states and their counties. Second, in order to verify the relationship between socioeconomic factors and opioid, we filter the socioeconomic data of five states and summarize them into 16 categories of indicators according to their definitions. Finally, in response to the proliferation of heroin in the Ohio, we recommend the government's two strategies for encouraging military service, increasing adult enlistment rates, and managing immigration with drug-related crime records. In the process of solving the problem, we mainly use R language and Python to draw pictures and build models.

## 1. Introduction

In March 2017, US President Trump signed an executive order to set up a presidential committee to combat the drug addiction and opioid crisis, investigating how to deal with drug abuse, addiction and the harm caused by the opioid crisis. In October 2017, Trump spoke at the White House, emphasizing the serious abuse of opioid which can lead to the death of hundreds of people every day, and announcing a national public health emergency in response to the opioid crisis [12].

## 2. Opioid case profile for each state

In order to better describe the distribution of reported opioid cases in each state and in the specific counties of the states, we have drawn the following four types of charts after datapreprocessing.

The first chart (pie chart) shows the total proportion of reported synthetic opioid and heroin cases in each state for 8 years. We would like to know the specific proportion of these two classes of drugs in each state.

The second chart (histogram) shows the annual changes in the total amount of reported synthetic opioid and heroin cases in each state, in order to understand the trend of reported synthetic opioid and heroin cases for each year.

The third chart (histogram) lists the top 10 counties in each state, based on the seven-year sum of reported synthetic opioid and heroin cases. It also shows the annual changes, hoping to know annual trend of counties with more reported opioid cases.

The fourth table lists the top five counties with the largest number of reported synthetic opioid cases in each state, and the top five counties with the largest number of reported heroin cases in each state respectively, hoping to understand the use of specific opioid.

### 2.1 Opioid case profile for Virginia

The number of reported synthetic opioid cases in Virginia accounts for 66.72% of all reported opioid cases, and reported heroin cases accounts for 33.28%. The number of reported opioid cases in

Virginia reaches its peak in 2013. Although it has declined in recent years, it is still on the rise, and the proportion of reported heroin cases is also growing. The county with the largest number of reported opioid cases for 2010-2017 is Fairfax.

## 2.2 Opioid case profile for Kentucky

There are relatively few reported opioid cases in Kentucky. The proportion of heroin cases in the previous two years is small. However, as time goes by, the number of heroin cases has gradually increased. In Kentucky, Jefferson County is far ahead of other counties, both in terms of synthetic opioid cases and heroin cases.

## 2.3 Opioid case profile for Ohio

The seven-year total heroin cases for Ohio State exceeds the number of synthetic opioid cases. In the first few years, the number of heroin cases far exceeded the number of synthetic opioid cases. However, in last two years, the number of synthetic opioid cases has increased significantly, and exceeded the number of heroin cases. The number of synthetic opioid cases and heroin cases in Hamilton is the highest in the state of Ohio.

## 2.4 Opioid case profile for Pennsylvania

The number of heroin cases in the state of Pennsylvania accounts for a larger proportion of total opioid cases, with only the number of heroin cases in 2017 being less than the number of synthetic opioid cases. Philadelphia and Allegheny have the largest number of opioid cases in the state of Pennsylvania and are far ahead of other counties.

## 2.5 Opioid case profile for West Virginia

The number of synthetic opioid cases in West Virginia is higher than the number of heroin cases per year. In recent years, both synthetic opioid cases and heroin cases have decreased significantly. Kanawha has the largest number of opioid cases in West Virginia.

## 2.6 Summary

In general, the following figure shows the annual distribution of the number of synthetic opioid cases and heroin cases in each state. We can see that the number of synthetic opioid cases in Ohio is the highest, and the number of heroin cases in Ohio is also the highest. On the contrary, the two cases are both the least in West Virginia.

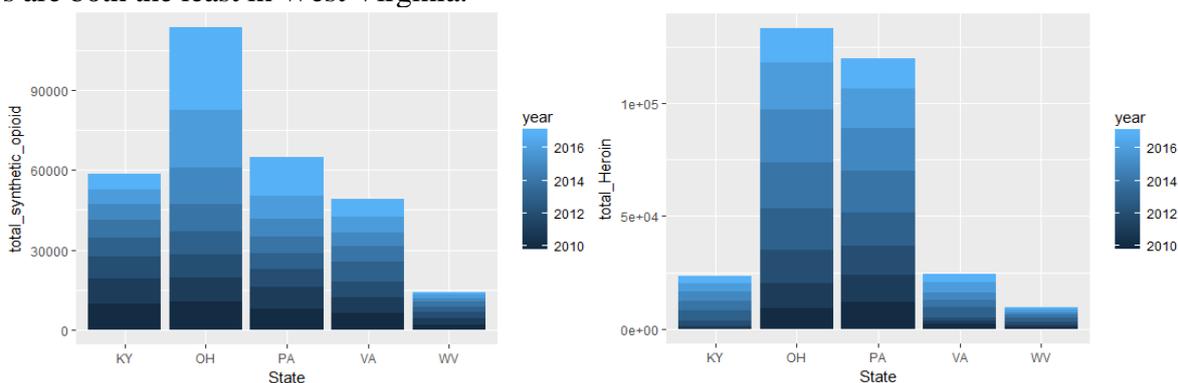


Figure 1: Opioid case profile for five states

From the perspective of map distribution, almost all counties in the state of Kentucky have reported opioid cases, which are very densely distributed. The case of heroin in Ohio is mainly concentrated in the southwestern county. Pennsylvania's synthetic opioid cases and heroin cases are loosely distributed. Virginia's synthetic opioid cases focus on the southeast region. The distribution of heroin cases is relatively loose in the northern region. There are few cases of synthetic opioid in the central part of West Virginia, and the case of heroin is mainly in the northeast.

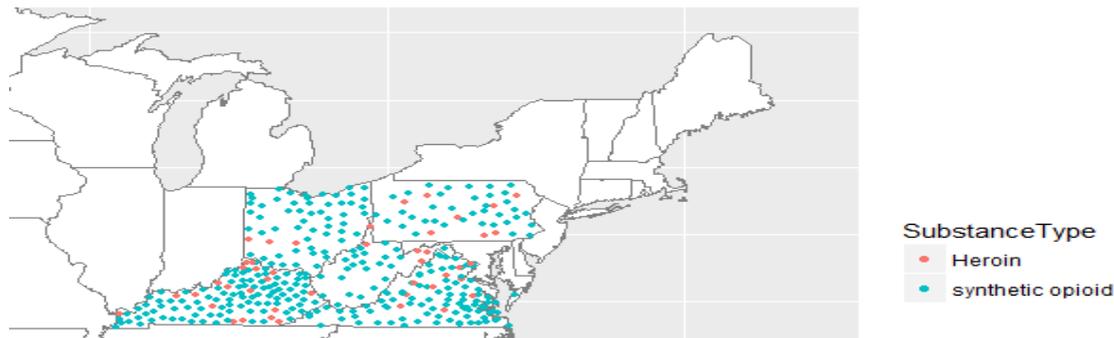


Figure 2: Opioid case map distribution

### 3. Prediction of opioid case

The least squares regression and gray prediction methods are used to fit and predict the synthetic opioid and heroin cases in each state.

Using least squares regression to predict the case of synthetic opioid and heroin in each state. Among them, the dependent variable is the number of reported cases of synthetic opioid or the number of reported cases of heroin. The independent variable is the year (1 in 2010, 2 in 2011, 3 in 2012, and so on). In order to prevent over-fitting in the prediction. We did not use high-order functions for fitting and prediction, and only used one-time and quadratic functions.

One-time fitting model:  $Y = a + bx$

Quadratic fitting model:  $Y = a + bx + cx^2$

Although the least squares regression is slightly different from the predicted value of the grey prediction, we can still find commonalities from different predictions.

(1) Synthetic opioid cases in these five states will continue to accelerate in the next three years, and from the fitted graph of synthetic opioid cases, there will be vertical growth in the next few years.

(2) From the predictions of the heroin case in the future, Ohio is facing huge challenges. The other four states have corresponding control over the heroin case in 2017, while the Ohio State heroin case is still at a high level.

(3) If the ratio of heroin to synthetic opioid is greater than 1, it is demonstrated that the number of heroin cases exceeds 50% of all opioid cases, indicating that heroin is gradually becoming the dominant opioid case, which is a dangerous sign because heroin is a Class I controlled substance, according to the Controlled Substances Act announced by Drug Enforcement Administration (DEA 2017). In the gray forecast of the state of Ohio, the ratio will exceed 1 in the next three years and is growing. Therefore, the state of Ohio faces significant pressure on heroin case control.

### 4. Strategies against the opioid crisis

The above analysis shows that the degree of heroin spread in Ohio is more serious, so the first step of this paper is to control the number of heroin reports in Ohio. The factors that have a negative impact on the heroin spread in the Ohio state include: marital status, veteran status, U.S. citizenship status, year of entry, language spoken at home and ancestry.

There are two indicators that the government can regulate and control, which are veteran status and year of entry. In Ohio, the government can encourage civilians over 18 years of age to perform military service. Military service can enhance residents' legal awareness and self-control ability and reduce opioid abuse. We increase the score of veteran status by 3%, 5%, and 7%, respectively, the rest of the socioeconomic factors remain unchanged, in order to see the change in the mean value of heroin in Ohio. The result is that the mean value of heroin in Ohio is decreasing as the score increases. Another strategy is to control immigration and control people with black history. Although the results of this measure have led to a decline in the mean value of heroin in Ohio state, the decline

is limited. After the implementation of these two strategies in Ohio, the heroin mean changes are shown in the figure below. The results of the first strategy are shown on the left and the results of the second strategy are shown on the right.

Combined with the above analysis, we can see that in the future, opioids will continue to spread. Year of entry, language spoken at home, and school enrollment negatively affect both synthetic opioid and heroin cases.

What the government can do is to raise the education level of all states and strengthen the education of the harm of opioid abuse. We increase the score of school enrollment by 3%, 5%, and 7%, respectively. The rest socioeconomic factors remain unchanged. We can see the mean value changes of synthetic opioid and heroin in the five states. The graph below shows the mean changes in total synthetic opioid and heroin in five states after implementing the strategy to improve education. Synthetic opioid is on the left and heroin is on the right. We can see that improving the education level of all states can reduce the spread of opioid.

## 5. Conclusion

In this article, we do a lot of data processing work, detailed graphical analysis of the annual synthetic opioid and heroin cases in each state. Least squares regression and grey prediction are constructed to predict the number of synthetic opioid and heroin cases in each state. At the same time, we use the principal component analysis and adaptive LASSO method to construct a quantitative model of socioeconomic indicators and opioid, in order to find the causal relationship between the two. Overall, we are satisfied with our work. Of course, this article also has advantages and disadvantages.

Advantages: 1) Accuracy: We performed reasonable cleaning and imputation of the data without using meaningless data. The final result is reliable and accurate, and we use two different models in the prediction, which is comparative. 2) The right model: We use a lot of theories and methods to support our work, each method is used reasonably. 3) Reasonable data classification: when conducting PCA, we use the existing index classification in socioeconomic data, and do not subjectively explore its classification. It is interpretable and reasonable. 4) Data interpolation method: we do not use the method of mean interpolation, but adopt the method of random forest interpolation. This method is more accurate for the interpolation of the exact data.

Disadvantages: 1) Algorithm limitations: when performing least squares and gray prediction, the amount of data is small, and our prediction results are rough. 2) Simplify the hypothesis: we must simplify the hypothesis to create a resolvable model. Therefore, some valuable data and information cannot be used.

## References

- [1] Barnett, M. L., Gray, J., Zink, A., & Jena, A. B. (2017). Coupling Policymaking with Evaluation - The Case of the Opioid Crisis. *The New England Journal Of Medicine*, 377(24), 2306–2309.
- [2] Burke, D. S. (2016). Forecasting the opioid epidemic. *Science (New York, N.Y.)*, 354(6312), 529.
- [3] Compton, W. M., Boyle, M., & Wargo, E. (2015). Prescription opioid abuse: Problems and responses. *Preventive Medicine*, 80, 5–9.
- [4] Corrigan, P. W., & Nieweglowski, K. (2018). Stigma and the public health agenda for the opioid crisis in America. *The International Journal On Drug Policy*, 59, 44–49.
- [5] DEA. Title 21 United States Code (USC) Controlled Substances Act. (2017). <https://www.deadiversion.usdoj.gov/21cfr/21usc/index.html>.
- [6] Fitzpatrick, J. J. (2019). The Nightmare of the Opioid Crisis in America. *Archives Of Psychiatric Nursing*, 33(1), 1.

- [7] Karila, L., Marillier, M., Chaumette, B., Billieux, J., Franchitto, N., & Benyamina, A. (2018). New synthetic opioids: Part of a new addiction landscape. *Neuroscience And Biobehavioral Reviews*.
- [8] Koenig, K. L. (2018). The Opioid Crisis in America: Too much, too little, too late. *The Western Journal Of Emergency Medicine*, 19(3), 557–558.
- [9] Lucas, P. (2017). Rationale for cannabis-based interventions in the opioid overdose crisis. *Harm Reduction Journal*, 14(1), 58.
- [10] Mahan, K. T. (2017). The Opioid Crisis. *The Journal Of Foot And Ankle Surgery: Official Publication Of The American College Of Foot And Ankle Surgeons*, 56(1), 1–2.
- [11] Marie, B. St., Arnstein, P., & Zimmer, P. A. (2018). Pain and Opioids: Call for Policy Action. *The Journal for Nurse Practitioners*, 14(1), 40-44.