American opioid drug crisis and response measures based on PCA

Caiwei Zhen[a], Weiqiu Wang, Jianzeng Mao

Northeast Forestry University, Harbin, China

[a]caiwei_zhen@163.com

Keywords: opioids, drug, principal component analysis (PCA), z-score, regression equation.

Abstract: The United States is experiencing a national crisis regarding the use of synthetic and non-synthetic opioids for the treatment and management of pain or for recreational purposes (illicit, over-the-counter use). After learning about the relevant background and reviewing relevant literature, our team answered the three parts of the questions proposed by the American mathematical and applied federation by using the data provided by NFLIS and the socio-economic data provided by the U.S.Census. Combined with the team's discoveries in the modeling process, we wrote a memorandum to the chief administrator, DEA/NFLIS database.

1. Introduction

Opioids are natural or synthetic compounds that can generate morphinoid pharmacological effects. Opioids are produced by a mixture of a type or a few of the opioid receptors in the peripheral and central nervous system, and it can be all excited, partly excited, excitement - antagonism and antagonism, and it helps to suppress the production and transfer of the harmful transference of the signal.[1] Opioid abuse is one of the greatest public health challenges of the 21st century. People who take low-dose opioids for more than three months have a 15-fold increased risk of addiction, according to the study. The average duration of opioid prescription medication per capita increased from 13 days in 2006 to nearly 18 days in 2015.[2]

2. Related work

2.1 Restatement of the Problem

For the opioid crisis, we focused on counties in five U.S. states: Ohio, Kentucky, West Virginia, Virginia and Pennsylvania. According to the drug identification count for narcotic analgesics and heroin reported to the DEA by crime LABS in each of the five states for 2010-2017, as well as data from the U.S. Census Bureau contained in seven other zip codes, we are required to complete the following questions:

We are required to build a mathematical model to describe the spread and characteristics of the reported synthetic opioid and heroin incidents in and between the five states and their counties over time using the NFLIS data provided. Using the model, identify any possible locations where specific opioid use might have started in each of the five states.

If the patterns and characteristics the model identified continue, are there any specific concerns the U.S. government should have? At what drug identification threshold levels do these occur? Where and when does the model predict they will occur in the future?

2.2 Assumptions

a) We do not have any error values in the processing and analysis of large quantities of data.

b) We have the right idea of the meaning and the specific nouns of each item.

c) Assuming the total number of people in the study area is constant at N, the population is divided into opioid-dependent patients and healthy people. At time t, the proportion of these two groups in the total number is s(t) and i(t).

d) The average number of effective daily exposures per opioid user is β constant, and β becomes

  DOI: 10.25236/icbcme.2019.025

the daily exposure rate. People who use opioids can be initiated when they have effective contact with people who do not.

e)The number of opioid-dependent patients who are cured each day as a proportion of the total number of opioid-dependent patients is a constant γ, which is called the daily cure rate. People who are cured can still become infected and take opioids again. Obviously 1/γ is the average infectious period.

3. The Model

3.1 Model one establishment

According to the data of NFLIS, we used Excel to draw the total drug report changing with time in five states, and drew the panorama of five states, and analyzed the total of all drug reports for the state noted in each state. The graph we drew can clearly observe the change trend of the total number of drug reports in each state over the 8 years, and form a cognition of the overall trend in the five states.

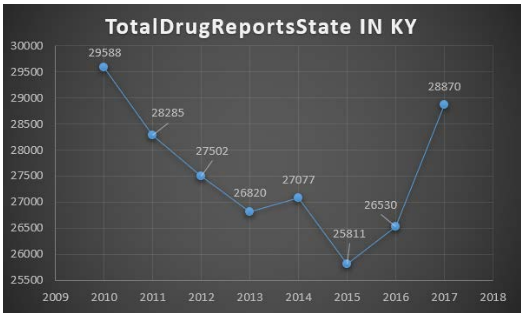

Figure1.Total Drug Reports State in Kentucky

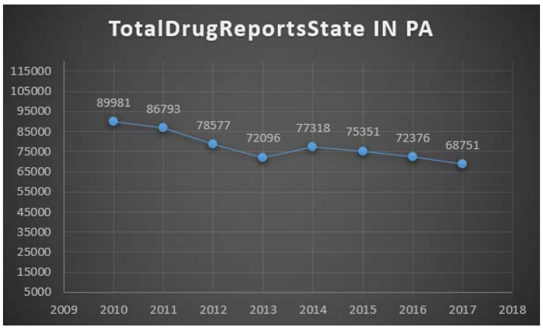

Figure2. Total Drug Reports State in Pennsylvania

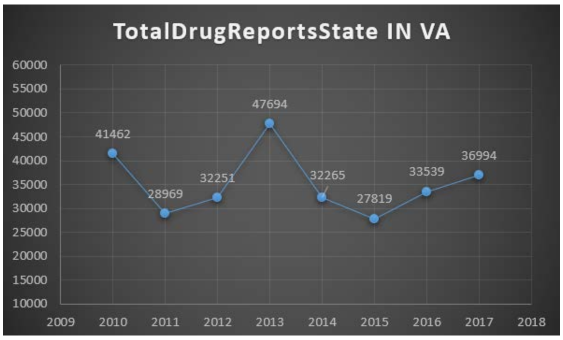

Figure3. Total Drug Reports State in Virginia

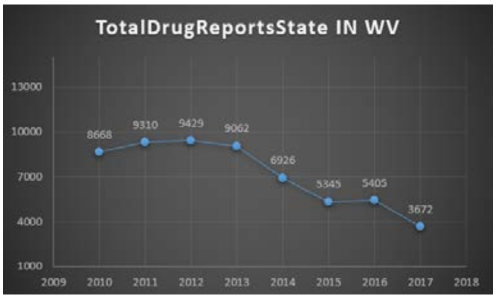

Figure4. Total Drug Reports State in West Virginia

By combining the above curves of total drug resports in each state and comparing them, it is obvious that Ohio outperforms the other four states in total drug resports and growth trend. Therefore, our team believes that Ohio is the main drug gathering and spreading point among the five states.

For the total drug report data of each state, we used Excel to make the statistical chart of drug report data of each state.

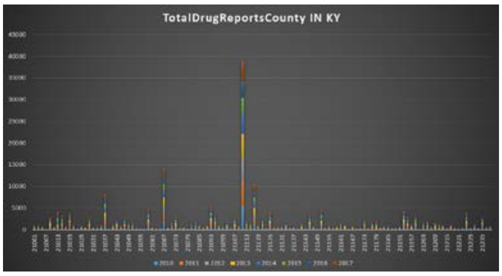

Figure5. Statistics on total drugs for 8 years in each county of Kentucky

As can be seen from the total number of all identified drug cases in a county in Kentucky, most counties in Kentucky reported a total number of drug reports below 5000 units in 8 years, showing a low overall level. Five counties have more than 5,000, and the county with the FIPS code combination of 21111 has the highest total for eight years at more than 35,000, far more than any other county, which is heavily dependent on opioids, JEFFERSON county.

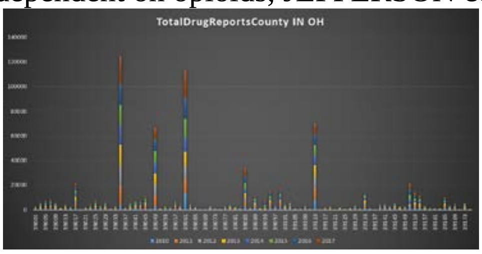

Figure6. Statistics on total drugs for 8 years in each county of Ohio

As can be seen from the statistical chart of 8-year the total number of all identified drug cases in a county of Ohio counties, 17% of the counties have more than 10,000 of the total number of all identified drug cases in a county, among which CUYAHOGA and HAMILTON counties have more than 100,000 of the total number of all identified drug cases in a county. Ohio outstrips the other four states in both total drug reports and county average drug reports.

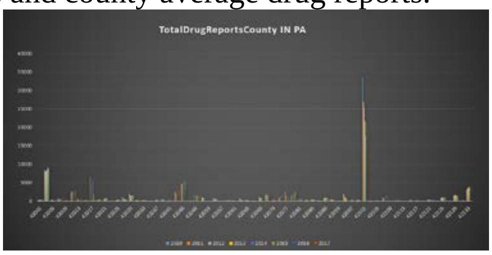

Figure7. Statistics for total drugs for 8 years in each county of Pennsylvania

Pennsylvania has a high overall level of total drug reports in 8 years, especially PHILADELPHIA, with over 180,000 of the total number of all identified drug cases in a county. It can be said that the drug epidemic is serious.

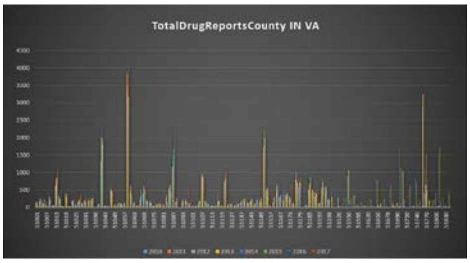

Figure8. Statistics on total drugs for 8 years in each county of Virginia

The overall level of the total drug reports in Virginia in the past 8 years is relatively low. The most prominent county is FAIRFAX, and the total number of all identified drug cases in a county is close to 30,000.

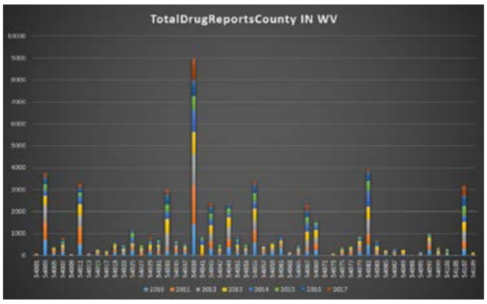

Figure9. Statistics on total drugs for 8 years in each county of West Virginia

West Virginia has the lowest total drug reports among the five states. The total drug reports has not exceeded 10,000 in 8 years, and there is little change in the total drug reports in each year. The overall level of drug control is good, and the county with the most drugs is KANAWHA.

We know from relevant literature, a fentanyl crisis is currently underway in the United States, characterized by an influx of counterfeit fentanyl-laced prescription drugs (e.g. Xanax®, Norco®, OxyContin®, and Oxycodone) now being advertised, sold and consumed by the public. The result of this counterfeit infltration into the U.S. drug supply chain has been an alarming increase in fentanyl-related overdose, deaths and seizures, due to illicitly produced products containing undeclared fentanyls and related fentanyl analogues/variants (e.g. acetyl fentanyl, butyrfentanyl, and furanylfentanyl.)[4] Therefore, we combined the data provided by NFLIS to analyze the use of fentanyl drugs in five states and counties.

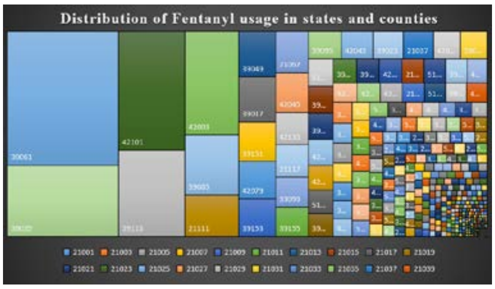

Figure10. Distribution of Fentanyl usage in states and counties

As shown in figure 10, the total number of counties with a FIPS_Combined size of 39061,39035, 39113 42101,42003 is high, and our team believes that these are likely the areas where Fentanyl started. The five districts comprise three counties in Ohio: HAMILTON, CUYAHOGA and MONTGOMERY, and two counties in Pennsylvania, PHILADELPHIA and ALLEGHENY.

In order to describe the report of the synthetic opioids and heroin incidents (cases) in the spread between the five states and counties, we adopt the SIS model. According to the assumption, every patient can turn $\beta s(t)$ healthy people into patients every day, and the number of patients at time t is $Ni(t)$, so $aNs(t)i(t)$ healthy people are infected every day.[5]

The rate of increase in patients:

$$Ndi/dt = \beta N \mathrm{si} - \gamma N \mathrm{i}$$

And

$$s(t) + i(t) = 1$$

We can derive

$$di/dt = \beta(1-i)i - \gamma i$$

Define a variable

$$\mathrm{k} = \beta / \gamma$$

It can be known that k is the average number of effective contacts per patient during the entire infectious period, which becomes the contact number.

The established SIS model is

$$\begin{cases} di/dt = -\beta i[i-(1-1/k)] \\ i(0) = i_0 \end{cases}$$

The following image can be obtained by solving the model with Matlab.

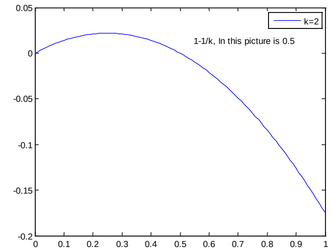

Figure11. di/dt—i (k>1)

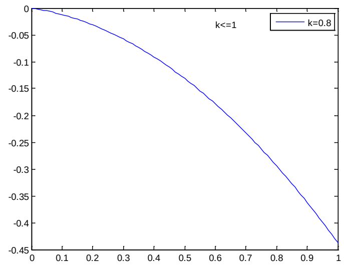

Figure12. di/dt—i (k<1)

It is not difficult to see that the contact number k=1 is a threshold. When k>1, the increase or decrease of i(t) depends on the size of i_0, but its limit value i(inf)=1-1/k increases with the increase of k. When k<=1, the proportion of patients i(t) becomes smaller and smaller, and finally tends to 0, which is due to the fact that the number of healthy patients becomes no more than the original number during the infectious period.

3.2 Model one prediction model

For the prediction of model 1 features, we use the Second exponential smoothing and the Single exponential smoothing to build the prediction model.[6] The principle formula of the Second exponential smoothing is as follows:

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$$

In the formula, $S_t^{(2)}, S_{t-1}^{(2)}$ are Second exponential smoothing values of t period and t-1 period respectively, α is the smoothing coefficient. Under the given conditions of $S_t^{(1)}$ and $S_t^{(2)}$, the prediction model of Second exponential smoothing is as follows:

$$\hat{Y}_{t+T} = a_t + b_t T$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha) S_{t-1}^{(2)}$$

$$\begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \dfrac{\alpha}{1-\alpha}\left(S_t^{(1)} - S_t^{(2)}\right) \end{cases}$$

The symbol T is the number of forecast lead periods. The formula of the Single exponential smoothing is as follows:

$$S_t^{(1)} = \alpha y_t + (1-\alpha) S_{t-1}^{(1)}$$

In the formula, $S_t^{(1)}$ and $S_{t-1}^{(1)}$ are the Single exponential smoothing values in t period and t-1 period respectively, and y_t is the actual value in t period. The curve of the total number of drug reports in five states over time was fitted with the Single exponential smoothing. The smoothing coefficient α=0.2, α= 0.5 and α=0.8 were selected for data processing respectively. According to the error between the processing result and the actual value, the smoothing coefficient with the smallest error, namely α=0.8, was selected as the smoothing coefficient of the Second exponential smoothing. (Detailed data processing steps are listed in the appendix.)

The linear quadratic exponential smoothing method of α=0.8 was used to predict the trend of the curve of the total number of drug reports over time in five states. The data of Kentucky were selected for detailed data processing steps. The processing methods of the other four states were the same.

The first step is to calculate the exponential smoothing value.

Take α=0.8, $S_0^{(2)} = S_0^{(1)} = y_1 = 29588$. Based on the formula (9), we can calculate an exponential smooth prediction as follows:

$$S_1^{(1)} = 0.8 \times y_1 + 0.2 \times S_0^{(1)} = 0.8 \times 29588 + 0.2 \times 29588 = 29588$$

$$S_2^{(1)} = 0.8 \times y_2 + 0.2 \times S_1^{(1)} = 0.8 \times 28285 + 0.2 \times 29588 = 28545.6$$

And so on for the rest $S_t^{(1)}$.

The third step is to calculate the parameter variable values a and b in each period. According to formula (8), the and b of each period can be calculated. See table ⑥ and table ⑦ respectively.

Step 4: according to equations (9) and (7), the predicted trend values of each period are calculated, as shown in the last column of the table. Such as the forecast value for 2015 is

$$\hat{Y}_{5+1} = a_5 + b_5 = 27036.99584 - 96.93184 = 26940.064$$

Extrapolate the forecast, and the forecast value for 2018 is

$$\hat{Y}_{8+1} = a_8 + b_8 = 28779.45957 + 1584.870973 = 30364.330543$$

Extrapolate the forecast, and the forecast value for 2019 is

$$\hat{Y}_{8+2} = a_8 + b_8 \times 2 = 28779.45957 + 1584.870973 \times 2 = 31949.201516$$

Based on the principles and methods of the procedure, the results of the total drug reports in the five states of 2018 to 2025 are predicted, and the results are as table II.

TABLE I. Kentucky Second exponential smoothing model

Time	t	y_t	$S_t^{(1)}$	$S_t^{(2)}$	a_t	b_t	$\hat{Y}_{t+T}$
①	②	③	④	⑤	⑥	⑦	⑧
2010	1	29588	29588	29588	29588	0	
2011	2	28285	28545.6	28754.08	28337.12	-833.92	29588
2012	3	27502	27710.72	27919.392	27502.048	-834.688	27503.2
2013	4	26820	26998.144	27182.3936	26813.8944	-736.9984	26667.36
2014	5	27077	27061.2288	27085.46176	27036.99584	-96.93184	26076.896
2015	6	25811	26061.04576	26265.92896	25856.16256	-819.5328	26940.064
2016	7	26530	26436.20915	26402.15311	26470.26519	136.2241536	25036.62976
2017	8	28870	28383.24183	27987.02409	28779.45957	1584.870973	26606.48934

TABLE II. FIVE STATES ARE REPORTING PROJECTED DRUG TOTALS FOR 2018-2025

Time	Ohio	Kentucky	West Virginia	Virginia	Pennsylvania
2018	121366	30364.33055	2456.773683	40001.46296	65642.6318
2019	129429.5618	31949.20152	1196.298865	43099.06333	62472.72015
2020	134415.2835	33534.07249	-64.17595392	46196.6637	59302.80851
2021	139401.0052	35118.94347	-1324.650772	49294.26407	56132.89686
2022	144386.7269	36703.81444	-2585.125591	52391.86444	52962.98522
2023	149372.4486	38288.68541	-3845.60041	55489.46481	49793.07357
2024	154358.1703	39873.55639	-5106.075228	58587.06518	46623.16193
2025	159343.892	41458.42736	-6366.550047	61684.66554	43453.25029

Based on the forecast data of these five states in 2018-2025 obtained by our prediction model, it can be found that the total number of reported drugs in Ohio, Kentucky and Virginia was predicted by the growth model, while that in West Virginia and Pennsylvania was predicted by the decline model.

The United States is in the midst of a devastating opioid misuse epidemic leading to over 33,000 deaths per year from both prescription and illegal opioids. Roughly half of these deaths are attributable to prescription opioids.[8] If this model feature continues, for the three growth states we believe that the US government may have some concerns in the following areas.

First, in terms of law and policy. There are three states in the five states that have continued to increase the number of opioids, indicating that the current legal policy in the United States is not strong enough in these three states. In states with high levels of drug abuse, existing laws and regulations alone will not meet U.S. targets for monitoring opioid use.

Second, in the social aspect. The epidemic of drug abuse has created a shortage of employees in many places, and sudden overdose deaths are on the rise. The latest figures show that substance abuse has reduced life expectancy in the United States, with more than 42,000 deaths due to opioid use and intoxication in 2016.[8]

Third, in terms of drug users. Opioid addicts are more likely to be accompanied by depression, which may be related to their long-term drug abuse, poverty, social discrimination, and difficulty in getting a job, so they are pessimistic and disappointed about the future. As a negative emotional response, depression can directly affect drug addicts' self-identity and coping style in the face of negative sexual events, leading to relapse or increase of suicidal ideation. Repeated suction behavior will aggravate the patient's depression, forming a vicious cycle.[9]

To define a reasonable threshold level for the United States government to worry, we've calculated the average of the drug reports in five states, 2010 to 2017, and to sum up these five averages. Each state is assigned a weighting coefficient based on the weight of the five averages in the sum. The average of the total number of drug reports in each state for eight years is multiplied by the weight coefficient of the state, and the value obtained is summed again to obtain a new summation. We use 200% of this new sum as the drug threshold. These values are all included in the table below.

TABLE III. Determination of the threshold value

State	The average	Weight coefficient	The product	The threshold value
OH	95830.125	0.393719215	37730.16159	141819.9158
KY	27560.375	0.11323213	3120.719976	
PA	77655.375	0.319048037	24775.79497	
VA	35124.125	0.144307888	5068.688289	
WV	7227.125	0.02969273	214.5930679	
Total	243397.125	1	70909.9579	

Based on our projections for the total number of drug reports in the five states for 2018-2025 and the drug identification threshold levels we identified, only Ohio out of the five states can meet and exceed this threshold in 2022. In Pennsylvania and West Virginia, where overall reported drug use is under effective control and will not reach the thresholds we set, the trend is gradually downward. For Kentucky and Virginia, which have not reached this level in the past nine years, we calculated when these two states would reach the threshold based on the prediction model.

The calculation for Kentucky is as follows:

$$\hat{Y}_{8+T} = a_8 + b_8 \times T = 28779.45957 + 1584.870973 \times T = 141819.9158$$

$$T = 71.3247060207$$

According to our definition of symbols, T should be an integer, so take 72. Therefore, Kentucky will reach and exceed the threshold in 2089 under our prediction.

The calculation for Virginia is as follows:

$$\hat{Y}_{8+T} = a_8 + b_8 \times T = 36903.86259 + 3097.600369 \times T = 141819.9158$$

$$T = 33.8701061183$$

Similarly, T should be 34 here. Therefore, we predict that Virginia will reach and exceed the threshold in 2051.

4. Evaluation of Model

4.1 Strength

In terms of data preprocessing, the given data set has been screened and checked. Standardized data processing has improved data quality and problem solving efficiency.

Our model has considered the influence of geographical location, economy, population and other factors, and has adopted the time series method and parameter setting method to obtain relatively accurate results.

4.2 Weakness

Unable to make full use of all indicators, through PCA to screen the 28 indicators to determine the regression function.

Simplified assumptions have been adopted for a solvable model, so the result may slightly digress from the truth.

This model didn’t consider the change of mean and variance, and didn’t consider the influence of policies.

References

[1] Sun li. American opioid abuse and its control measures [J]. Chinese journal of drug abuse control, 2018, 24(4):219-224.

[2] CDC. Opioid prescribing is still high and varies widely throughoutthe U.S[EB/OL]. (2017- 07- 06)[2017- 08- 22]. https://www.cdc.gov/ media/releases/2017/p0706- opioid.html.

[3] Ouyanghui, fan yinghua, gong changzhen, et al. The opioid crisis and the opportunity of acupuncture development in the United States [J]. Journal of traditional Chinese medicine, 2018.3, 24(5):1-9.

[4] Mackey T K, Kalyanam J. Detection of illicit online sales of fentanyls via Twitter [version 1; referees: 2 approved][J]. F1000Research, 2017(6):1937.

[5] Feng hao. SIS model and its analysis on dynamic complex network [D]. Shanxi University, 2017.

[6] Lu dan-dan, gu qiao-xiang, xing chao. Optimization of markov prediction model by quadratic exponential smoothing method [J].Journal of China university of metrology, 2017,28(03): 334-339.