

Analysis and Supression of Opioid Spread

Summary

We propose a model to describe the characteristics of the number, rate and direction of opioid spread in and between states and counties using transition matrices. The result shows that among the five given states, Ohio is most likely the source of opioid cases and has frequent opioid transition with Pennsylvania and Kentucky. We also find that if no effective regulatory measures are imposed, the number of drug cases in Ohio will increase up to 141,266 in the next decade. The spread of opioids between counties was analyzed in the same way, and the result is shown in Figure 3.

We determine two epidemic thresholds based on characteristics of opioid spread, with respect to the number of opioid cases and opioid spread rate respectively. We perform simulations to forecast the number of opioid cases in states and counties. The result shows that the number of narcotic analgesics cases in Ohio reaches the threshold in 2022, and the number of heroin cases in Pennsylvania reached the threshold in 2023.

We analyze the correlations between the number of opioid cases and various socio-economic factors using information entropy. We find that the number of opioids cases demonstrates high correlations with disability status, educational level, family status and adolescents. Therefore, we provide suggestions based on these factors to help suppress opioid spread. We also evaluate the effectiveness of these strategies on both the spread rate and the number of opioid cases. The result shows that when the amount of opioids spread is reduced to 74.5%, the amount of opioids will stay on a relatively low level, and become lower than that of 2010 after 5 years.

We also do sensitivity analysis of various parameters to prove the robustness of our model. The result shows that both the continuous and periodical inflows lead to increase in the number of opioids in the state, and the increased amount depends on the specific transition amount.

key words : Transition matrix; Multi-level thresholds; Information entropy

MEMO

FROM: Team 1917694 , MCM C

To: The group of Governors

Date: January 28, 2019

Dear Officials:

It is our honor to help you with analyzing the spread and characteristics of opioid. We are writing this letter to report our findings.

We analyzed the characteristics of opioid spread in the five states and counties, and found that if left uncontrolled, the opioid abuse will continue to spread, and the number of drug users will increase. We also found that there is no obvious relationship between the spread of opioid use and the geographical location. We deduced that Ohio is most likely the originated state of opioid use, and the drug cases will increase significantly over time. Of all five states, West Virginia shows a relatively healthy status. We also analyzed each county in the five states. Taking West Virginia as an example, Monongalia and Berkeley may be the originating counties for heroin transmission, and drug cases in these two counties are mainly spread to Kanawha and Berkeley.

We determined two epidemic thresholds based on characteristics of opioid spread, with respect to the number of opioid cases and opioid spread rate respectively. We found that the number of narcotic analgesics cases in Ohio reaches the threshold in 2022, and the number of heroin cases in Pennsylvania reached the threshold in 2023.

We analyzed the correlations between the number of opioid cases and various socio-economic factors. We found that the number of opioids cases shows high correlations with disability status, educational level, family status and adolescents. Therefore, we provide suggestions based on these factors to help suppress opioid spread. We also evaluate the effectiveness of these strategies. The result shows that when the amount of drug spread is less than 26.3% on before, the number of drug users will be significantly reduced in 3 years, and become lower than that of 2010 in 5 years.

We also do sensitivity analysis of our model. We found that inflows

of drugs from outside the states also plays an important role in affecting the number of opioid cases.

Based on our analysis, we provide some suggestions to effectively prevent opioid spread:

1. National and local governments play an important role in detecting and preventing opioid overdose and abuse. Monitoring systems and early warning mechanisms should be established to effectively detect and respond to the spread of opioid abuse.

2. Provide more job opportunities to alleviate people's pressure, even to set up a free subsidy program. The local government should also publicize the harm of opioids abuse.

3. Prescriptions and sales of opioids should be strictly controlled, and a feedback mechanism for mandatory postoperative hospital visits should be established. The patients should also be examined for abuse of opioids during the hospital examination.

4. Impose careful check and strict management of logistics, to reduce the impact of opioids from outside the states.

We believe that our model is useful for preventing opioid spread in the future. You are welcome to contact us at any time for further cooperation.

Sincerely yours
MCM C Team 1917694

Contents

1	Introduction	1
1.1	Restatement of the Problem	1
1.2	Our Works	1
2	Assumptions and Notations	2
2.1	Assumptions	2
2.2	Notations	2
3	Model Construction	3
3.1	Opioid Spread Model	3
3.2	Determining the Epidemic Threshold	5
3.3	Modification to the Opioid Spread Model	6
3.3.1	Determining Correlation Using Information Entropy	6
3.3.2	Modifying Our Model with Hidden States	7
3.3.3	Analysis of the Modified Model	8
4	Model Simulation and Analysis	8
4.1	Describing Characteristics of Opioid Spread	8
4.2	Suggestions	12
4.3	Evaluating the Effectiveness of Strategies	13
5	Strengths and Weaknesses	15
5.1	Strengths	15
5.2	Weaknesses	16
6	Sensitivity Analysis	16
7	Conclusion	17
	References	19
	Appendices for Code and Data	20

1 Introduction

1.1 Restatement of the Problem

Many people nowadays rely heavily on opioids and even die from opioid overdose. Reducing opioid use and suppressing opioid spread becomes an urgent task. Therefore we are facing the following problems:

- Describe the spread and characteristics of opioids in and between the given states and counties, analyze the characteristics of opioid spread, and find the source of opioid use cases.
- If opioid cases continue to spread, determine the epidemic threshold for the government to take action and analyze where and when they will occur.
- Analyze the correlation between the number of opioid cases and the socio-economic factors, then modify the model to include important factors.
- Establish effective policies to prevent the spread of opioids and evaluate the effectiveness of policies.

1.2 Our Works

- First, we build a model to analyze the characteristics of opioid spread, and determine the source of opioid cases based on the characteristics.
- Then, we determine two thresholds with respect to the spread rate and the number of opioid cases, perform simulations to evolve the opioid spread process to determine where and when the government should take actions first.
- We use information entropy to measure the correlations between the number of opioid cases and the socio-economic factors, and propose reasonable suggestions for those factors with high correlation with the number of opioid cases.

- Finally, we establish effective policies and perform simulations to evolve the spread of opioids, analyze the result to determine when opioid spread can be successfully suppressed, and quantify the goals that government needs to achieve.

2 Assumptions and Notations

2.1 Assumptions

Due to the lack of necessary data, we make the following assumptions to help us perform modeling.

1. Inflow of drugs from outside of the five states are neglected.
2. The microscopic behaviors of opioid transmission are ignored.

2.2 Notations

Here are all the notations and their meanings in this paper.

Symbol	Meaning
t	Time
N	Total reported opioid cases
N_t	Total reported drug cases
λ	Average cases induced by a single case
A_t	Status at t
E	Set containing socio-economic factors with high correlation t
T	Transition matrix
$i(t)$	Proportion of opioid cases in all drug cases at t
μ_1	Average number of drug cases induced by an existing drug case
μ_2	Number of opioid cases induced among all drug cases
γ	Drug spread slow down factor
i_0	Status at t
H	Information Entropy
p_0	Initial number of drug cases

3 Model Construction

Census data will have certain deviations. We don't consider statistical errors of the provided census data and just use the values provided.

3.1 Opioid Spread Model

As is well known, spreading of illegal drugs are elusive, therefore we do not consider the ways, actions and tools of spreading opioid uses. Instead we focus on analyzing the characteristics of drug use spreading, including the amount of transfer, rate of change and location, in order to obtain conclusions that are convenient for the administration and control of government departments.

According to the report on opioid use published by NFLIS[1], we define the drug spread characteristics as the number of cases, the rate of change, and the location of use:

- Distribution of drug cases: Indicates the number of drug cases over time in a region.
- Rate of change: Describes the growth rate of the number of drug cases over time (a positive growth rate indicates a normal trend in drug spread, while a negative growth rate indicates a downtrend).
- Location of drug use: Indicates regions where drug cases are mainly concentrated and regions where drug cases are reported.

Distribution of drug cases A state can be transferred to a next state by applying a transition matrix [2], therefore we construct a transition matrix T , an element T_{ij} in T indicates the proportion of drug cases flowing from region i to region j to the total drug cases in region i . We define a matrix A_{2010} as the drug cases reported in 2010, and similarly A_{2011} as the drug cases reported in 2011. Then under the influence of T , A_{2010} transfers to A_{2011} . We then derive the general form of drug transmission based on the reasoning above as

$$T \cdot A_{t-1} = A_t \implies T = A_t \cdot A_{t-1}^+ \quad (1)$$

After solving the transition matrix T , which are different across years, we can obtain information on the spread of drugs.

We define β as the average amount of drugs in a single drug case, T_{ij} describes the flow of drugs from region i to region j . We define η as the

proportion of drugs flowing out of region i , then an η less than 1 indicates there are drugs consumed in region i , and an η greater than 1 indicates there are drugs flowing in from other regions.

Location where specific drug use starts We first take the mean value of the transition matrices T across all given years to get the average transition matrix T_{mean} in order to eliminate the effects of anomalous propagation. Each element in the matrix T_{mean} describes the flow of drugs, and the location where specific opioid use might have started can be determined based on the largest outflow. Furthermore, without government intervention, under the effect of the transition matrix T , the evolution of drug use transmission will eventually show a stable state [3]. Therefore, we compute the drug cases vector A_i as time t approaches infinity according to Formula 1, to determine which regions have high incidences of drug cases, and then take action.

Drug transmission rate[4] We define N_t as the number of reports of all drug cases, and the initial number of drug cases is p_0 . We also define $i(t)$ as the proportion of opioid cases in all drug cases at t , and the initial proportion of opioid cases is i_0 .

We first consider the spread of drug use towards non-drug users. Assuming that the average number of drug cases induced by an existing drug case is μ_1 , we establish the following equation:

$$\frac{dN_t}{dt} = N_t\mu_1, N_t(0) = p_0 \quad (2)$$

The solution to the above equation is

$$N_t(t) = p_0 e^{\mu_1 t} \quad (3)$$

This indicates that in the absence of any government precautions, the number of drug cases will grow exponentially over time, resulting in severe harm, as shown in Figure 1(a).

We then consider the spread of opioids in all drug cases. We define μ_2 as the number of opioid use cases induced by an opioid use case among all drug cases. Therefore, $(1 - i(t))\mu_2$ represents the increase in the number of opioid use cases among all drug cases as induced by a single opioid use case. Also notice that the total number of opioid use cases is $N_t i(t)$, thus the total increase in the number of opioid use cases among all drug cases is $N_t i(t)\mu_2(1 - i(t))$. Therefore we establish the following equation:

$$N_t \frac{di}{dt} = N_t \mu_2 i(t)(1 - i(t)), i(0) = i_0 \quad (4)$$

The solution to the above equation is

$$i(t) = \frac{1}{1 + \left(\frac{1}{i_0} - 1\right)e^{-t\mu_2}} \quad (5)$$

which indicates that $i(t)$ is an S-shaped growth curve, as shown in Figure 1(b). The derivative di/dt reaches the maximum at $t = 0.5$, at which time the opioid use case grows the fastest, indicating a potential opioid use outbreak.

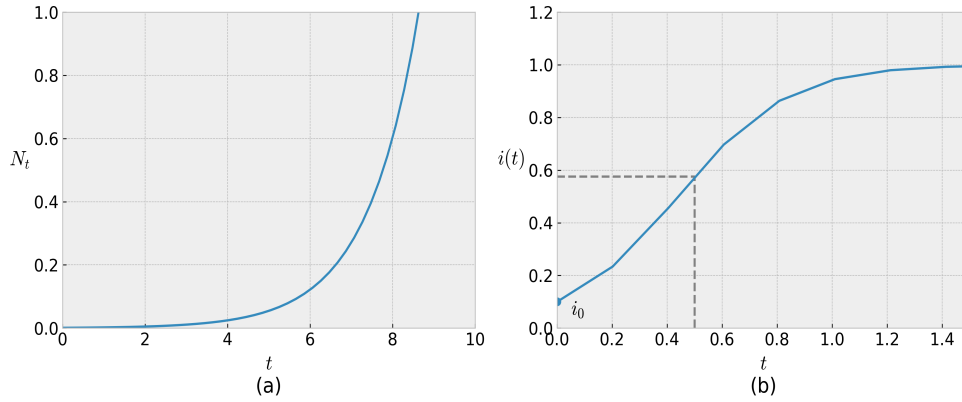


Figure 1: Drug cases and proportion of opioid cases over time

3.2 Determining the Epidemic Threshold

Since the number of drug cases, population and level of economic development are different between states and counties, the epidemic threshold is calculated from a statistical perspective, and each state and county should have different epidemic thresholds. In addition, the excess of the amount and speed of drug transmissions should both raise the government's concern, therefore two epidemic thresholds are set in consideration of both aspect, respectively.

For the epidemic threshold in the number of drug cases, $m + 2\sigma$ is generally considered as a gold standard, where m and σ denote the mean and standard deviation of the number of drug cases respectively[5]. The calculated threshold is considered constant over time, because the number of drug cases increases by year, and by setting a relatively low threshold, the government can stay vigilant and therefore suppress the number

of drug cases promptly.

For the epidemic threshold in transmission speed, we compute the derivative of $N_t(t)$, which gives us

$$N'_t(t) = \mu_1 p_0 e^{\mu_1 t} \quad (6)$$

We observe that the growth rate $N'_t(t)$ turns out to be an exponential function, and the number of drug cases will keep growing if no interventions are imposed. So we set the initial growth rate at $t = 0$ to be the baseline growth rate, which turns out to be $\mu_1 p_0$. Here we find that the growth rate threshold is related to μ_1 and p_0 , where p_0 reflects the amount of drug users in a given region, and μ_1 reflects the administrative state such as the strictness of management.

3.3 Modification to the Opioid Spread Model

3.3.1 Determining Correlation Using Information Entropy

There are uncertainties and randomness in the process of drug transmission, and the way to eliminate uncertainty is to introduce new information from the outside, and using this information can eliminate some of the uncertainties. Information entropy is a common measurement of uncertainty.

For the number of drug cases N , its uncertainty $H(N)$ is computed according to the following formula:

$$H(N) = - \sum P(N) \cdot \log_2 P(N) \quad (7)$$

Suppose the introduction of information from the socio-economic data provided by the U.S. Census Bureau can reduce uncertainty. The more uncertainties are reduced, the stronger the correlation [6]. We introduce an socio-economic factor as a piece of information Y , then we calculate the uncertainty of N under the condition of Y , i.e. the conditional entropy $H(N|Y)$:

$$H(N|Y) = - \sum P(N, Y) \log_2 P(N|Y) \quad (8)$$

The conditional entropy $H(N|Y)$ after the introduction of information Y is less than the information entropy $H(N)$, which indicates that the

introduction of the socio-economic factor reduced the uncertainty. We define Δh to be the reduction in information:

$$\Delta h = H(X) - H(X|Y) \quad (9)$$

There are multiple attributes in the provided data. although it is possible to calculate the conditional entropy $H(N|Y_1, Y_2, \dots, Y_n)$ under multiple information, it is difficult to determine the optimal order in which the information is integrated, for example, whether $H(N|Y_1, Y_2, Y_4)$ reduces more uncertainty than $H(N|Y_1, Y_2, Y_3)$. Therefore we only calculate Δh of the number of drug cases and one other socio-economic factor. After calculating Δh for all socio-economic factors, those factors with larger value of Δh are kept in a set E . The socio-economic factors in the set E demonstrate a strong correlation with the number of drug cases, and we make suggestions to the government based on the phenomena represented by these attributes.

3.3.2 Modifying Our Model with Hidden States

As depicted above, the socio-economic factors in the set E demonstrate strong correlations with the number of reported drug cases. Therefore, in order to improve our model's performance, it is necessary to integrate these factors into our opioid spread model.

The specific relationships (e.g. linear or nonlinear) between the number of drug cases and the socio-economic factors are unclear, and there does not yet exist formulae to describe the relationships between them. Although the number of drug cases are provided, there is no effective way to figure out the drug transmission processes. Therefore the numbers of drug cases A is defined as the hidden state, and the observable socio-economic factors E are defined as explicit states. Then we modify our opioid spread model to take both the explicit states and the hidden states into consideration, as illustrated in Figure 2.

In the Figure 2, the transition between hidden states can be calculated by the transition matrix T . An explicit state can be seen as a window through which the corresponding hidden state can be observed. The correlation between a hidden state and the corresponding explicit state can be calculated according to Formula (9).

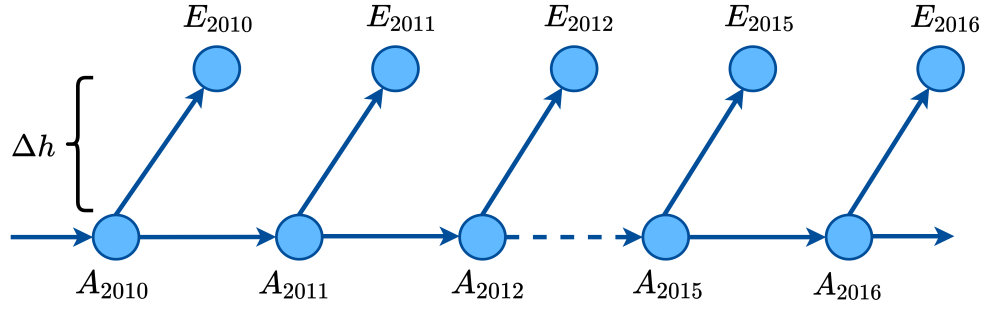


Figure 2: Illustration for model structure

3.3.3 Analysis of the Modified Model

The structure of our improved model shares lots of similarities with the hidden Markov model. Our improved model can solve the following problems [7]:

Given the information of the socio-economic factor set E in the past few years as the explicit state and the corresponding number of drug cases as the hidden state. Also given the specific explicit state value D , then the transition matrix T_{max} which is most likely to produce D can be calculated:

$$T_{max} = \text{ArgMax} \prod_{t=1}^n (P(E_t|A_t)|P(T_t|T_{t-1})) \quad (10)$$

where E_t, A_t, T_t represents the explicit state, hidden state, and transition matrix at time t . The transition matrix T with the highest probability can be obtained based on the distribution of the population, and the result becomes more stable over time. With the transition matrix, which indicates the drug case transmission, the government will be able to effectively monitor and prevent the drug spread in the region, thus reduce the number of drug cases.

4 Model Simulation and Analysis

4.1 Describing Characteristics of Opioid Spread

Using the provided data, we perform simulation on our model to simulate the drug spread in and between states and counties, then analyze the resulting spread characteristics.

Opioid spread between counties: We take West Virginia as an example to analyze the transmission of synthetic opioid and heroin inside the state. The spread status of heroin is shown in Figure 3.

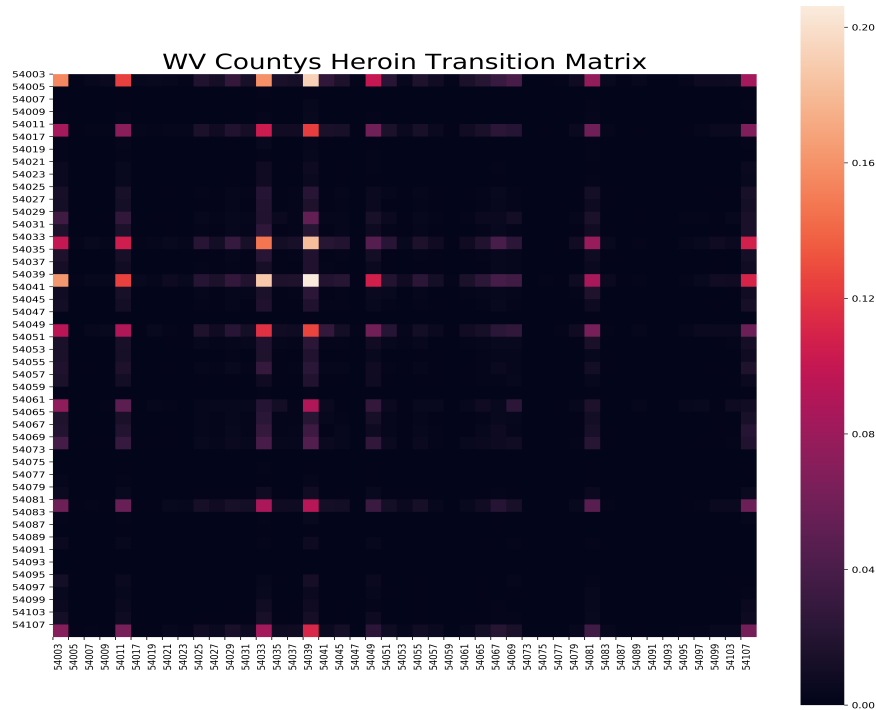


Figure 3: Transition matrix for opioid spread rate in West Virginia

We draw a the conclusion that Monongalia (54061) and Berkeley (54003) are most likely to be the originating counties for heroin transmission, and heroin in these two counties are mainly transferred to Kanawha (54039) and Berkeley. In addition, Berkeley has a large flow rate, and as both an inflow and outflow region of heroin, therefore it is necessary to strengthen supervision on heroin transmission in this region. For other counties such as Cabell (54011), Harrison (54033) and Marion (54049) which have large heroin inflows, it is also necessary to strengthen supervision on drugs and check the logistics.

We analyze the transmission of narcotic analgesics (synthetic opi-

oids) in West Virginia in the same way. We find that Mercer (54055) has a high inflow and outflow of narcotic analgesics. Therefore we can deduce that Mercer may be the manufacturing and main consumption county of analgesics. For other counties such as Raleigh (50481), Nicholas (54067), and Kanawha (54039), there are also large outflows and use cases, which indicates they are also main consumption and spread regions of analgesics in West Virginia. The transmission flows in the remaining counties are significantly lower. The resulting figures are in Appendix A.

Opioid spread between states: We analyze the transmissions of all drugs within the five states, and plot the heat map for the transition matrix of analgesics and heroin. The results are shown in Figure 4.

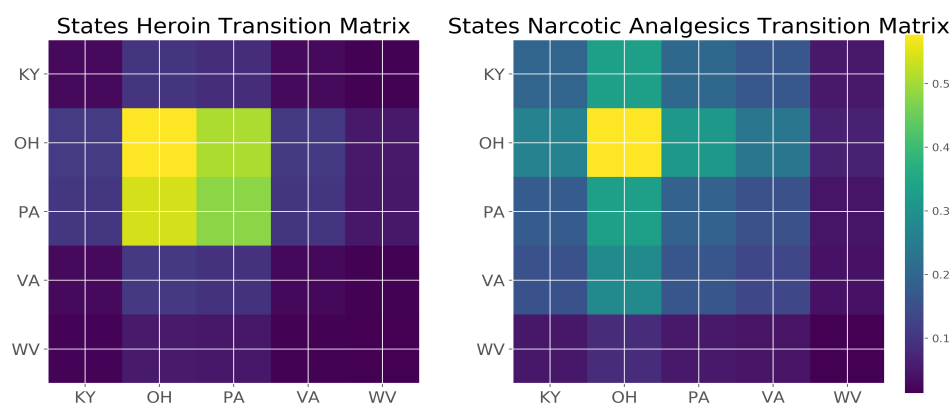


Figure 4: Heat map of transition matrix for five states

We find that analgesic and heroin transmissions between Kentucky, Ohio, and Pennsylvania are frequent. We also notice that these three states were located at edges relative to other states geographically, indicating that geographical locations generally do not prevent the spread of drugs. In addition, the consumption and outflow of drugs in Ohio are high, thus should raise the government's attention.

Source analysis of drug cases: In order to eliminate the effects of abnormal transmission, we take the average of the transmission matrices in 2010-2017 in each state as the drug transmission matrix. The resulting transition matrix is shown in Table 1.

From Table 1 we observe that Ohio has the largest drug outflow, therefore we deduce that Ohio is most likely to be the source of drug transmissions. We also find that the flow between Pennsylvania and Ohio

Table 1: Transition matrix for states

State	KY	OH	PA	VA	WV
KY	0.044	0.158	0.124	0.056	0.012
OH	0.146	0.539	0.405	0.184	0.037
PA	0.128	0.454	0.359	0.161	0.035
VA	0.057	0.199	0.158	0.07	0.015
WV	0.013	0.045	0.036	0.016	0.004

Table 2: Thresholds for narcotic analgesics and heroin for the five states

	KY	OH	PA	VA	WV
Narcotic analgesics	9507	47226	20260	7848	1877
Heroin	6192	25005	19110	6265	1463

are high, therefore logistics between these two states should be strictly monitored in order to reduce the number of drug cases. Among all five states, West Virginia has the lowest drug inflow, which indicates a relatively healthy status.

For the epidemic threshold with respect to the spread rate, which indicates the proportion of the number of drug users and policies in a region, it cannot be calculated due to the lack of necessary data. For the threshold with respect to the number of drug cases, we take the five states as an example, and calculate the mean m and the standard deviation σ for the number of reported drug cases in 2010-2017. The resulting threshold is shown in the table 2.

We evolve narcotic analgesics and heroin use in the five states in the next few years using the formula $A_{t+1} = T \cdot A_t$. The result shows that the amount of heroin use in Pennsylvania reaches the threshold in 2023, and the amount of narcotic analgesics use in Ohio reaches the threshold in 2022.

Stability analysis: We iterate the total nuamder of reported drug cases in each data over time under the influence of the transition matrices and observe its stability. The result is shown in Table 3.

We find that Ohio will gradually become a gathering place for drugs and opioids over time. Therefore, the government should strengthen

Table 3: Evolution of drug cases in states

Year	KY	OH	PA	VA	WV
2018	29403	121549	70018	37473	3740
2019	29944	123789	71309	38370	3809
2020	30450	126701	72623	39078	3879
2021	31508	128395	73962	39798	3950
2022	31631	130762	75325	40532	4023
2023	32214	133172	76714	41279	4097
2024	32808	135627	78128	42040	4173
2025	33412	138127	79568	42814	4250
2026	34028	140673	81035	43604	4328
2027	34656	141266	82528	44047	4408

management and control in Ohio and improve drug rehabilitation measures. In West Virginia, it demonstrates slower growth compared to other states. For Kentucky, Pennsylvania and Virginia, although the number of drug cases are lower than Ohio, they should also raise the government's attention, in order to be effectively controlled when the number of drug cases reaches the threshold.

4.2 Suggestions

Opioids are widely prescribed to relieve pain. However, bad habits lead to the abuse of opioids, which also causes great harm.

As indicated by Δh , the number of opioids cases demonstrates high correlations with disability status, educational level, family status (whether living alone, having children, number of family members, having elderly people) and adolescents (marital status, education, whether living alone). Therefore, we provide suggestions based primarily on these factors to help prevent the abuse of opioids.

- For the disabled: The disabled needs analgesics to relieve pain after the surgery, but overuse can lead to serious consequences. Therefore, for local hospitals, prescriptions and sales of opioids should be strictly controlled, and a feedback mechanism for mandatory postoperative hospital visits should be established. The patients should also be ex-

amined for abuse of opioids during the hospital examination.

- For educational levels: For those with higher educational levels, they are more aware of the hazards of opioids, while those with lower educational levels may not understand the potential hazards of opioids and thus lead to inappropriate use. Therefore, the local government should publicize the harm of opioids abuse.
- For the family status: It's the responsibility for the government to provide more job opportunities to alleviate people's pressure, even to set up a free subsidy program. In addition, domestic violence should be prohibited, in order to maintain a harmonious family relationship.
- Detection and response: National and local governments play an important role in detecting and preventing opioid overdose and abuse. Monitoring systems and early warning mechanisms should be established to effectively detect and respond to the spread of opioid abuse. Institutions such as the Centers for Disease Control and Prevention, the health department, and law enforcement agencies should be at the forefront to address this issue as well.

4.3 Evaluating the Effectiveness of Strategies

Drug spread in the crowd:

The spread of opioids can only be effectively suppressed under specific conditions. We evolve the spread of drugs in the future. We first analyze those socio-economic factors with high correlations to the number of drug cases, then suggest strategies based on the values of the factors.

Suppose that the government's compulsory drug rehabilitation measures and the effective prevention of drug transmission slows down the drug spread by γ , and the proportion of drug users in the population at t is $p(t)$, we modify the equation2 to get the following equation:

$$N_t \frac{dp}{dt} = N_t \mu_1 t - \gamma N_t t \quad (11)$$

We solve the equation and plot the $p(t)$ function, as shown in Figure 5. We discover in the figure that μ_1/γ serves a threshold: only when $\mu_1/\gamma \leq 1$, which indicates the control to the drugs is stronger than the spread of drugs, will the number of drug users gradually decrease to zero. We find

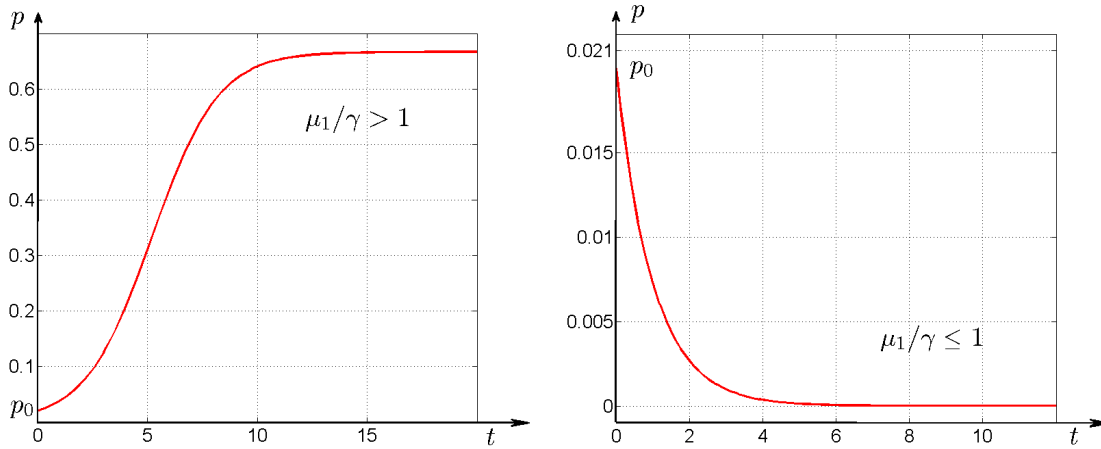


Figure 5: The proportion of drug users over time

that under the effect of policies, the number of drug cases start to decline. By analyzing the results of the drug spread depicted by the equation, we find that after four years, the drug use was effectively controlled. The parameter at this time is $\mu_1/\gamma = 0.5$, indicating that the number of drugs controlled is twice the number of opioid spread.

Under the control of policies, opioid transmissions between the five states as well as the number of opioid cases begin to decrease, and the transition matrix is reduced, as shown in Figure 6. The blue line indicates the number of opioid cases without government's interventions, while the red line indicates the number of opioid cases with certain policies imposed by the government to control the opioid spread. We interpret the results as follows:

The figure shows that the number of drug cases transmission will gradually decrease, and after 9 years, it will be basically the same as in 2010. The average transition matrix T in 2017-2027 is shown in Table 4. Comparing to Table 1, the number of opioid cases in Ohio is found to have the most significant reduction, so the supervision in Ohio needs to be strengthened. The overall outflows of the states have been reduced. We sum across the transition matrix, and get the amount of drug transmission to be 2.603 after policies are imposed, where it was 3.543 in 2010-2017. Therefore, the number of opioid cases in the five states is reduced by 26.5%.

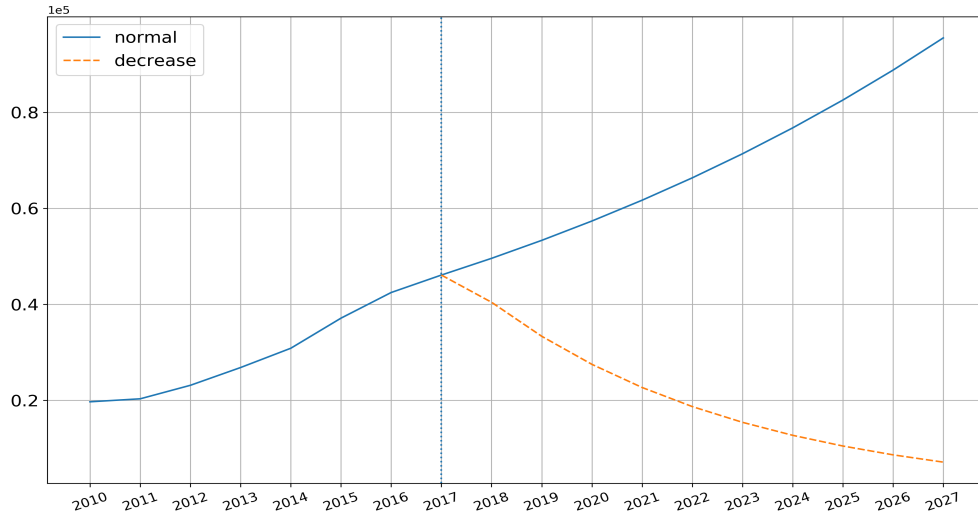


Figure 6: Comparison of number of drug cases

Table 4: Transition matrix for states

State	KY	OH	PA	VA	WV
KY	0.027	0.129	0.070	0.036	0.003
OH	0.136	0.469	0.354	0.179	0.014
PA	0.082	0.389	0.212	0.107	0.008
VA	0.035	0.167	0.091	0.047	0.004
WV	0.005	0.022	0.012	0.006	0.002

5 Strengths and Weaknesses

Based on the modeling process, we make some comments on our model as listed below.

5.1 Strengths

- We employ transition matrices in our model to analyze the spread and characteristics of opioid use as well as to provide quantitative results for further analysis, which shares a lot of similarities with the Markov chain, leading to a concise, easy-to-understand yet powerful model structure. Our model can also be adapted for other purposes

such as computing rankings and analyzing virus transmission, showing good generalizability.

- By employing conditional entropy, our model is able to effectively measure the correlations between the opioid cases and socio-economic factors without knowing the concrete relationships (e.g. linear or non-linear) between them.
- Two epidemic thresholds are determined for the number and the rate of opioid transmissions, both of which are quantified, making it convenient for government departments to make decisions.

5.2 Weaknesses

- We do not take the opioid spread outside of the five states (KY, OH, PA, VA, WV) into consideration due to the lack of data, which may affect the analysis result.
- Although reasonable suggestions are made for the socio-economic factors which shows high correlation with the number of opioid cases, we do not quantify each effect when policies are imposed, instead we represent the effects by a single parameter, the value of which we have provided.

6 Sensitivity Analysis

The analysis of the spread and flow of drugs is mainly based on the transition matrix T . In our assumptions, we assume that the transition matrix does not contain the inflow of opioids from outside the five states, but in actual cases this is not appropriate. We do sensitivity analysis of the transition matrix, in order to prove the robustness of our model. While imposing strict management inside the state, there may be inflows of opioids from outside the five states, and the inflows may be long-term continuous inflows or indirect outflows. We calculate the transition matrix T in both cases, then do sensitivity analysis of the spread results of the five states. The result is shown in Figure 7.

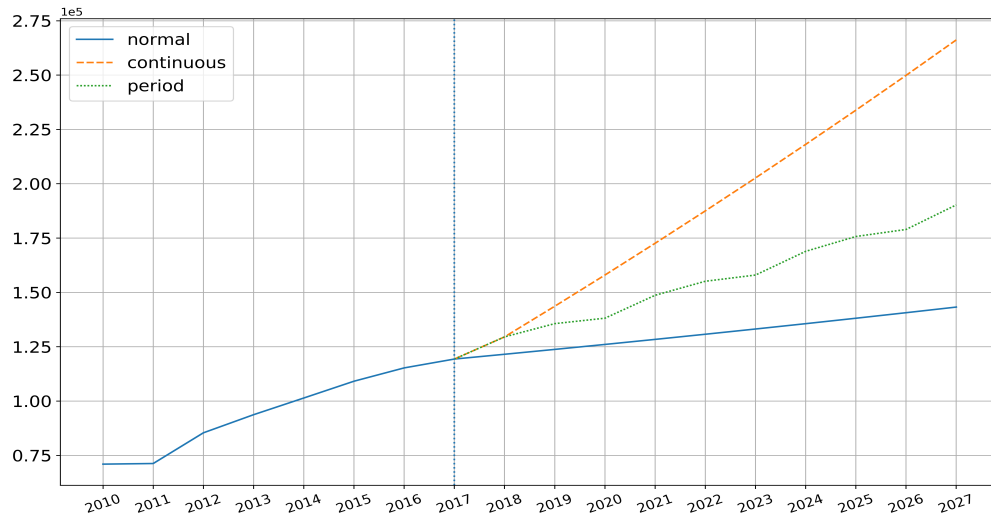


Figure 7: The number of cases

Both the continuous and periodical inflows lead to increase in the number of opioids in the state. The result shows that the continuous inflow of opioids has the greatest impact. Although the indirect inflows are incidents, they should also raise the government's concern. In conclusion, the government should not only impose strict regulations on opioids inside the state, but also pay attention to the inflow of opioids from outside the state, and monitors the logistics constantly.

7 Conclusion

We build a model to analyze the opioid spread in and between the provided states and counties, determine epidemic thresholds with respect to the number and spread rate of opioid cases, analyze the correlations between the number of opioid cases and various socio-economic factors, then we modify our model to include the most important factors, and provide suggestions to the government based on the factors with high correlations with the number of opioid cases. We also perform simulations to monitor the evolution of opioid use cases in the future. Here are some of our interesting findings:

- On a state level, we suggest that Ohio (OH) is the source of opioid use cases. Without policy interventions, there will be even more cases reported in the future.
- The epidemic threshold with respect to the spread rate is strongly related to the proportion of drug abuse population and management system of a region, and by setting a lower epidemic threshold, the number of opioid cases in the region can be effectively controlled.
- When the government's detection rate of opioid use is higher than the spread rate of opioids, the drug abuse population will be significantly reduced within 2 years. And when the transmission amount of opioid use drops by 26.3%, the number of opioid cases will decrease to the same level as in 2010.

References

- [1] “Nflis-drug 2017 annual report,” <https://www.nflis.deadiversion.usdoj.gov/DesktopModules/ReportDownloads/Reports/NFLIS-Drug-AR2017.pdf>.
- [2] L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web.” Stanford InfoLab, Technical Report 1999-66, November 1999, previous number = SIDL-WP-1999-0120.
- [3] A. Y. Ng, A. X. Zheng, and M. I. Jordan, “Stable algorithms for link analysis,” in Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ser. SIGIR ’01. New York, NY, USA: ACM, 2001, pp. 258–266.
- [4] X. Meng and L. Chen, “The dynamics of a new sir epidemic model concerning pulse vaccination strategy,” *Applied Mathematics and Computation*, vol. 197, no. 2, pp. 582 – 597, 2008.
- [5] E. L. Tay, K. Grant, M. Kirk, A. Mounts, and H. Kelly, “Exploring a proposed who method to determine thresholds for seasonal influenza surveillance,” *PloS one*, vol. 8, no. 10, p. e77244, 2013.
- [6] E. T. Jaynes, “Information theory and statistical mechanics,” *Phys. Rev.*, vol. 106, pp. 620–630, May 1957.
- [7] S.-Z. Yu, “Hidden semi-markov models,” *Artificial Intelligence*, vol. 174, no. 2, pp. 215 – 243, 2010, special Review Issue.

Appendices

Here is Code we used in our model, which python is the main development language.

Appendices A

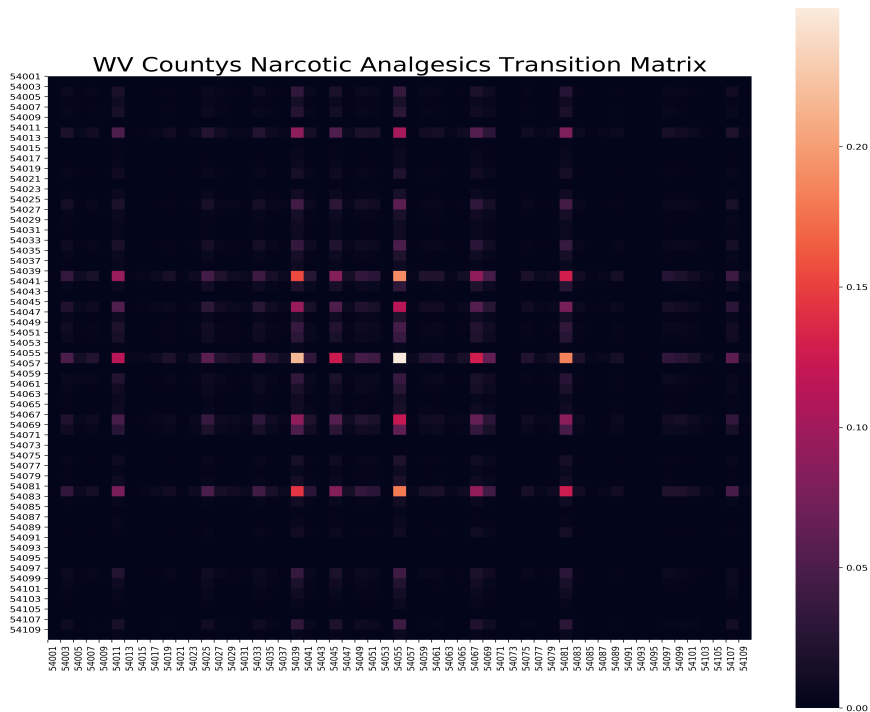


Figure 8: Transition matrix for synthetic opioid spread rate in West Virginia

Appendices B: Forecast and Draw the Drugs of 2018-2027 Cases

```
# -*- coding: utf-8 -*-
import numpy as np
import pandas as pd
import pylab as pl
import seaborn as sns

def predict(m):
    t = t_2017.copy()
    t_predict = []
    years = 10
    for _ in range(years):
        t = np.dot(t, m)
        t_predict.append(t[0])

    index = [2018+i for i in range(years)]
    t_predict = pd.DataFrame(t_predict, index=index,
                             columns=df.columns)

    return t_predict

def getMatrix(df, year_before=2010, year_after=2011):
    status_before = df[df.index == year_before].values
    status_after = df[df.index == year_after].values
    status_pinv = np.linalg.pinv(status_before)

    # dot
    transfer_matrix = np.dot(status_pinv,
                             status_after)

    return transfer_matrix

df = pd.read_csv('data/FIPS_groupby_year_state.csv',
                 engine='python', index_col=0)
m = getMatrix(df, 2016, 2017)
```

```
t_2017 = df[df.index == 2017].values

data = df.iloc[:, 1]
p = predict(m).iloc[:, 1]
data = pd.concat([data, p])

m_ = m**1.1

p = predict(m_).iloc[:, 1]
p = np.append([119349], p)
p = pd.Series(p, index=list(range(2017, 2027+1)))
data = pd.concat([data, p], axis=1)
data.columns = ['normal', 'decrease']

# plot
pl.style.use('ggplot')
pl.figure(figsize=(14, 8))
ax = pl.gca()
ax.yaxis.major.formatter.set_powerlimits((0, 0))
sns.lineplot(data=data)
pl.axvline(x=2017, linestyle=':')
pl.xticks(list(range(2010, 2027+1)), list(range(2010,
    2027+1)),
    rotation=20, fontsize=12)
pl.xlabel('Year', fontsize=15)
pl.yticks(fontsize=12)
pl.legend(fontsize=15)
pl.savefig('assets/OH_predict.png', dpi=500)
pl.show()
```

Appendices C: The Program for Calculating Information Gain

```
import numpy as np
import pandas as pd
import os
```

```

def calc_ent(x):
    x_value_list = set([x[i] for i in range(x.shape
        [0])])
    ent = 0.0
    for x_value in x_value_list:
        p = float(x[x == x_value].shape[0]) / x.shape
            [0]
        logp = np.log2(p)
        ent -= p * logp

    return ent

def calc_condition_ent(x, y):
    x_value_list = set([x[i] for i in range(x.shape
        [0])])
    ent = 0.0
    for x_value in x_value_list:
        sub_y = y[x == x_value]
        temp_ent = calc_ent(sub_y)
        ent += (float(sub_y.shape[0]) / y.shape[0]) *
            temp_ent

    return ent

def calc_info_gain(x, y):
    base_ent = calc_ent(y)
    condition_ent = calc_condition_ent(x, y)
    ent_grap = base_ent - condition_ent

    return ent_grap

years = [i for i in range(2010, 2017)]

df_opioid_all = pd.read_excel('data/MCM_NFLIS_Data.
    xlsx')
df_opioid_all = df_opioid_all[['YYYY', 'FIPS_Combined
    ', 'DrugReports']]

```

```

df_opioid_list = []
for year in years:
    df = df_opioid_all.loc[df_opioid_all['YYYY'] ==
        year]
    df.drop('YYYY', axis=1, inplace=True)
    df_opioid_list.append(df.groupby(['FIPS_Combined'
        ]).sum())

base_path = 'data/ACS_{YY}_5YR_DP02/ACS_{YY}'
            '_5YR_DP02_with_ann.csv'
filepaths = [base_path.format(YY=i) for i in range
    (10, 17)]
df_census_list = []
for filepath in filepaths:
    df = pd.read_csv(filepath, skiprows=[1],
        index_col='GEO.id2',
                        na_values=['**', '-', '***', '
                        *****', 'N', '(X)'])
    df.drop(['GEO.id', 'GEO.display-label'], axis=1,
        inplace=True)
    df.sort_index(axis=1, inplace=True)
    df_census_list.append(df)

df_attrs = pd.read_excel('data/socio-economic.xlsx')
attr_dict = dict(zip(df_attrs['code'], df_attrs['
    chinese']))

s_list = []

attr_codes = set(df_census_list[-1].columns).union(
    set(df_census_list[0].columns))
s_list.append(pd.Series([attr_dict.get(attr_code) for
    attr_code in attr_codes],
                        index=attr_codes))

for i in range(7):
    df_opioid = df_opioid_list[i]
    df_census = df_census_list[i]

```

```
info_gains = []
for attr_code in df_census:
    df_temp = pd.DataFrame()
    df_temp['attr'] = df_census[attr_code]
    df_temp['opioid'] = df_opioid['DrugReports']
    a = df_temp.values
    attr_array = a[:, 0].ravel()
    opioid_array = a[:, 1].ravel()
    attr_array = np.nan_to_num(attr_array)
    opioid_array = np.nan_to_num(opioid_array)
    info_gain = calc_info_gain(opioid_array,
                               attr_array)
    info_gains.append(info_gain)

s_list.append(pd.Series(info_gains, index=
    df_census.columns))

result_col_names = ['attr_desc'] + years
df_result = pd.DataFrame({str(result_col_names[i]):
    s_list[i] for i in range(8)})
df_result = df_result.iloc[:190]
df_result['average'] = df_result.mean(axis=1,
    numeric_only=True)

df_result.to_excel('data/result.xlsx')
```

Appendices D: Transition Matrix of Five States

```
# -*- coding: utf-8 -*-
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

df_one = pd.read_excel(
    'data/state_heroin_mean_transport_matrix.xlsx',
    index_col=0)
```

```
df_another = pd.read_excel(  
    'data/state_nonheroin_mean_transport_matrix.xlsx'  
    , index_col=0)  
  
labels = ['KY', 'OH', 'PA', 'VA', 'WV']  
  
plt.style.use('ggplot')  
plt.figure(figsize=(12, 6))  
plt.gca().set_aspect(1)  
  
plt.subplot(121)  
plt.imshow(df_one)  
plt.xticks(range(len(labels)), labels, fontsize=13)  
plt.yticks(range(len(labels)), labels, rotation=360)  
plt.title('States Heroin Transition Matrix', fontsize  
    =20)  
  
plt.subplot(122)  
plt.imshow(df_another)  
plt.xticks(range(len(labels)), labels, fontsize=13)  
plt.yticks(range(len(labels)), labels, rotation=360,  
    fontsize=13)  
plt.title('States Narcotic Analgesics Transition  
    Matrix', fontsize=20)  
plt.tight_layout()  
plt.subplots_adjust(bottom=0.1, right=0.9, top=0.9)  
cax = plt.axes([0.92, 0.1, 0.02, 0.8])  
plt.colorbar(cax=cax)  
  
plt.savefig('assets/state_drug_transition_matrix.png'  
    , dpi=500)  
# plt.show()
```
