2019

Mathematical Contest In Modeling®

Certificate of Achievement

Be It Known That The Team Of

Xingyu Liu
Wenhu Guo
Yushu Xu

With Faculty Advisor

Xiaofeng Gao

Of

Shanghai Jiao Tong University

Was Designated As

Finalist

Patrick Driscoll, Contest Director

Administered by

COMAP

With support from

AMS ASA informs MAA siam

David H. Olwell, Head Judge

For office use only	Team Control Number	For office use only
T1 _____	**1919054**	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	**C**	F4 _____

2019
MCM/ICM
Summary Sheet

Modeling Opioid Crisis Based on Cellular Automata and Chi-Square Test

Summary

The US has been in the grips on an "opioid epidemic" since the 1990s, and nowadays, the misuse and addiction to opioids have become a serious national crisis that affects public health as well as social and economic welfare. In this paper, we formulate the characteristcs of opioid cases in Kenturky (KY), Ohio (OH), Pennsylvania (PA), Virginia(VA) and West Virginia (WV).

Aimed to depict opioid spread characteristics in and between states, we establish a **Opioid Spread Model** based on **Cellular Automata**, whose update rules are determined by the Adjacency and Relative Total Drug Reports between counties. We define three types of counties: **low-risk**, **susceptible** and **high-risk**. By applying **Opioid Spread Model** iteratively, we derive the result that the transformation of large portion of susceptible counties to high-risk ones will occur in GREEN ,LEWIS, AMELIA, CRAIG ,JAMES CITY, 2021.

Based on the Opioid Spread Model we obtain, we devise **F_Score** to evaluate one county's possibility to be the origin of a specific opioid. F_Score is correlated to seven (2010-2016) years' mean and variance of the specific opioid reports in one county. Employing F_Score, we find the origin of heroin is Delaware, while oxycodone may start from Madison.

To attain the most important socio-economic factors, we construct a **Feature Selection Model** based on **Chi-Square Test**. The top 3 factors are "VETERAN STATUS - Civilian population 18 years and over", "DISABILITY STATUS - Total Civilian Noninstitutionalized Population-18 to 64 years" and "MARITAL STATUS - Never married". With the selected factors, we modify our Opioid Spread Model and propose a **Drug Strategy** which is targeted at the selected groups. Then we employ the modified Opioid Spread Model to test the strategy, and the result shows that our strategy can effectively prevent susceptible counties from transforming into high-risk ones. In addition, we identify that the strategy can success only when the coefficients of opioids' growth rate are limited to [0.3,0.8).

In the end, we make sensitivity analysis and discuss strengths and weaknesses. The sensitivity analysis shows that our model is sensitive to the parameter related to the spread rate while completely not sensitive to the parameter related to the death rate.

Keywords: Cellular Automata, Chi-Square Test

Modeling Opioid Crisis Based on Cellular Automata and Chi-Square Test

January 28, 2019

Contents

1 Introduction

1.1 Background

The problem of addiction to opioid analgesics has emerged as a major issue for the United States in the past decade and has worsened over the past few years[1], which affects public health as well as social and economic welfare.

For "opioid epidemic" problems, there are mainly three types of epidemic models: Dynamical Systems Models, Cellular Automata Models and Network Diffusion Models[2]. Dynamical Systems Models are often established via coupled dynamical equations such as in classic SIR models and their variations, which are suitable at macro scales. Wakeland et al.[3]has developed system dynamics models of nonmedical opioid use trajectories. As for Cellular Automata Models, they are typically useful to represent regional spread and evolution, while Network Diffusion Models can help capture social influences, person-to-person spread, and within-population topologies[4].

1.2 Our Work

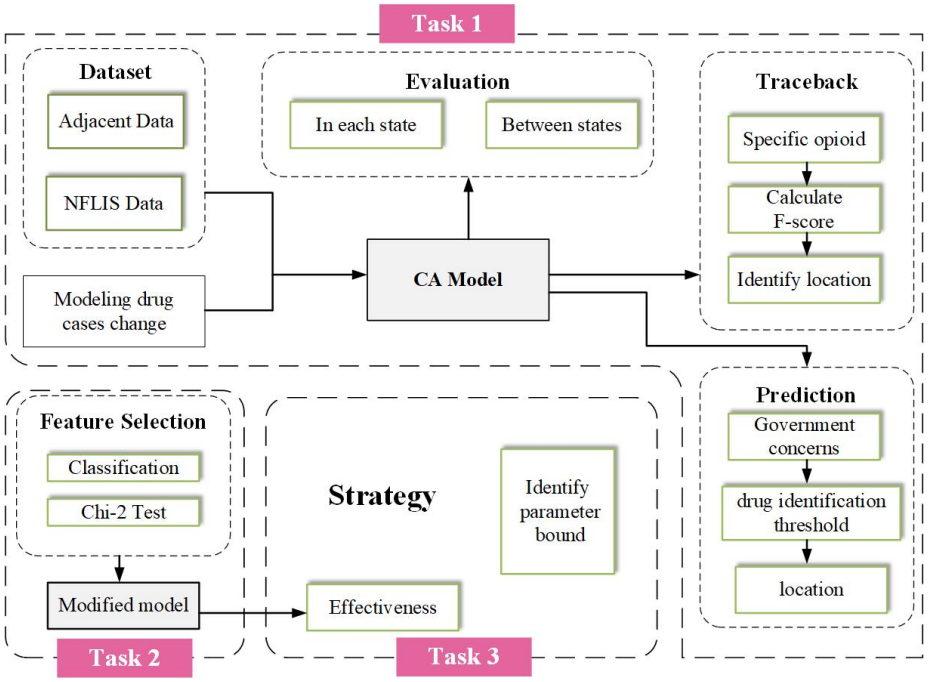


Figure 1: Overview of our study

To address this problem, we study the spread, characteristics,origin ,trend and related socio-economic attributes of drug abuse and propose feasible strategy, as is shown in Figure 1, and we summarize our work as follows:

- **Model Establishment**. Based on the DEA / NFLIS database, we established a **Opioid Spread Model** based on **Cellular Automata** capable of depicting drug spread characteristics and can be used for evaluation, traceback and prediction.
- **Evaluation**. From our original Opioid Spread Model, we visualize the spread and characteristics of drug cases on the map and analyse the results. We define three types of counties: low-risk, susceptible and high-risk.

- **Traceback**. Based on the Opioid Spread Model we obtain, we devise **F_Score** to evaluate one county's possibility to be the origin of a specific opioid. F_Score is correlated to seven (2010-2016) years' mean and variance of the specific opioid reports in one county. Employing F_Score, we find the origin of heroin is Delaware, while oxycodone may start from Madison.

- **Prediction**. After defining the concerns according to related literatures, by applying **Opioid Spread Model** iteratively, we derive that the concern (the transformation of large portion of susceptible counties to high-risk one) will occur in GREEN ,LEWIS, AMELIA, CRAIG ,JAMES CITY, 2021.

- **Feature Selection and Model Modification**. To attain the most important socio-economic factors, we construct a **Feature Selection Model** based on **Chi-Square Test**. The top 3 factors are "VETERAN STATUS - Civilian population 18 years and over", "DISABILITY STATUS - Total Civilian Noninstitutionalized Population-18 to 64 years" and "MARITAL STATUS - Never married".

- **Strategy Effectiveness Test and Bound Identification**. With the selected factors, we modify our Opioid Spread Model and propose a **Drug Strategy** which is targeting at the selected groups. Then we employ the modified Opioid Spread Model to test the strategy, and the result shows that our strategy can effectively prevent susceptible counties from transforming into high-risk ones.

2 Assumptions

First and foremost, we make some basic assumptions and explain their rationales.

Assumption1 *Drug identification counts in one county is influenced and will only influence the neighboring counties.*

Due to the spread of opioids, the number of drug cases of a county is influenced by other counties. In order to simplify the analysis, we only consider the spread between adjacent counties. This assumption is one of the rule of our model.

Assumption2 *Opioid addicts do not migrate on a large scale outside the five states.*

Large-scale migration is generally very rare. To simplify our analysis, we assume it will not happen.

Assumption3 *The characteristics of different opioids (addiction, spread) are similar.*

When analyzing all the opioids in a county, we will not consider the impact of different types of opioids.

Assumption4 *The provided data is realistic and accurate to a certain degree.*

Despite the incompleteness of the data and some tolerant error in statistics, we make this assumption to guarantee one valid solution.

3 Notation

For simplicity, we define a series of notations for relevant terms.

Table 1: Notations of Relevant Terms

Notations	Relevant Terms
$RATE_{adjacent}$	Drug cases' spread rate from adjacent counties
$RATE_{death}$	Death rate from drug cases in a county
$RATE_{infection}$	Infection rate of drug in a county
$RATE$	Total growth rate of drug cases in a county
$RATE_{socio-economic}$	Growth rate due to socio-economic factors
C_o	The number of specific opioid cases in one county
$M(C_o)$	The mean of the opioid cases number
$V(C_o)$	The variance of the opioid cases number

4 Opioid Spread Model based on Cellular Automata

In this section, we firstly do some data scrubbing based on the provided data. Then we propose Opioid Spread Model to comprehensively characterize, analyze, forecast, and evaluate the spread and characteristics of the reported synthetic opioid and heroin cases. Finally, we apply our model to predict future opioid changes.

4.1 Model Establishment: Opioid Spread Model

Considering that there are so many kinds of drugs listed, we select only the total opioid incidents to analyse the spread pattern and characteristics of the reported drug incidents in and between the five states and their counties over time. In this way, we can treat each county as a cell, and the number of opioid incidents in a county is regarded as the corresponding cell's state. Adjacent relationship between individual cells is the same as the spatial distribution of counties. Each cell has the capability of holding the information pertaining to that cell. In each iteration, the cells alter their states based on the information they capture.

Considering the spread patterns of the reported synthetic opioid and heroin incidents over time, we consider three possible reasons that may affect a county's '$TotalDrugReportsCounty$' next year :

1. County total count of all substances identified in the next year will be affected by its neighboring counties.
2. County total count of all substances identified in the next year will decrease because of the death caused by opioid abuse.
3. Counties of high susceptible risk are more likely to increase its '$TotalDrugReportsCounty$' next year.

Figure 2 summarizes our Opioid Spread Model logic, each cell represents one specific county. The cells are arranged according to its FIPS code in the simulation result and their states can vary depending on pre-determined rules. In our experiment, cells in cellular automata update their states over time as functions of adjacent state variables and their previous states. Detailed update rules are introduced below.

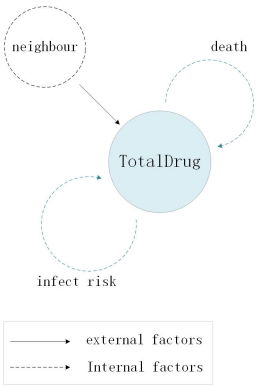


Figure 2: Overview of general logic mechanisms

Rule 1 updating state based on its neighbours' states:

$$RATE_{adjacent} = \max_{neighbour} TotalDrugReportsCounty \times \alpha$$

Rule 2 updating state based on its previous state:

$$RATE_{death} = TotalDrugReportsCounty_{own} \times \mu$$

In a nutshell, a cell's next state can be calculate through formula (1).

$$C_{i,t+1} = C_{i,t} \times (1 + 0.5 \times (RATE_{adjacent} + RATE_{death} + RATE_{infection})) \quad (1)$$

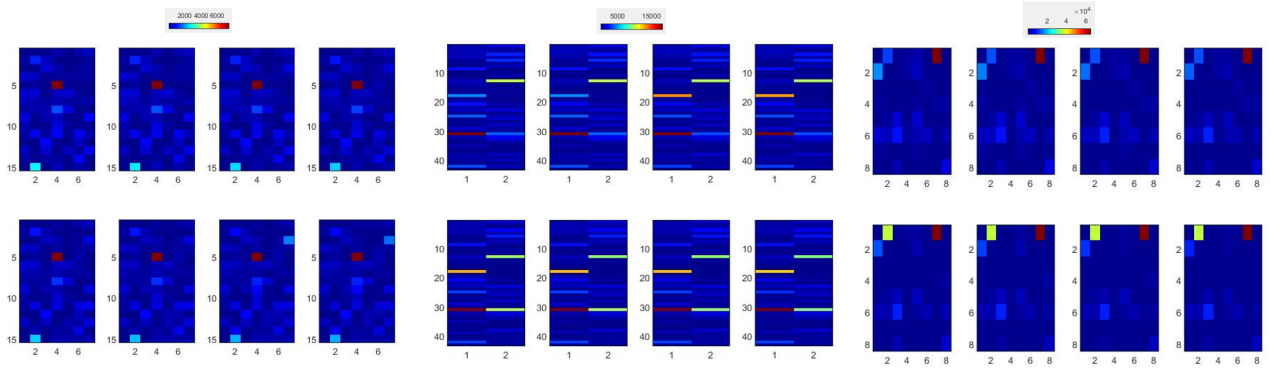
Algorithm 1: Opioid Spread Model

Input: $TotalCounty$:total number of the county
$Matrix_{adj}$: reflect the connection between two counties
$state_{ini}$: the initial number of "TotalDrugReportsCounty"
$iter$: the iteration number
$lambda$: the affection of neighbor
mu : the stochastic death rate
$Rate_{infec}$: the infection rate;
Output: $TOTALstate$
for $i = 1; i \leq iter; i++$ **do**
  $RATE = \frac{1}{2} \times (RATE_{adjacent} + RATE_{death} + RATE_{infection})$
  $NewState = OrinState \times (RATE + 1)$
  update RATE
  draw picture
end
return TOTALstate;

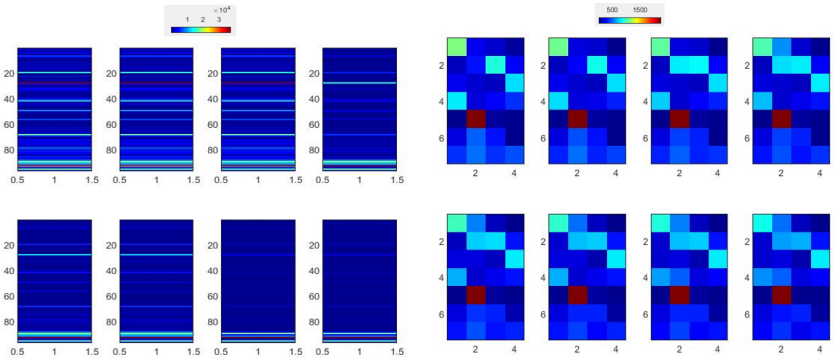
Parameter α and μ are constant. We determine α and μ by the given data from 2010 to 2017. After verification, we set $\alpha = 1$ and $\mu = 0.05$ in our experiment. Parameter α reflects the level of neighbor affection, parameter μ reflects the level of deaths caused by drug abuse. As for $RATE_{infection}$, we set four levels of $RATE_{infection}$ based on the growth rate of '$TotalDrugReportsCounty$' which is expressed by the variable λ. Finally, We construct our Opioid Spread Model based on the above rules and variables. Algorithm 1 shows the pseudo-code for our Opioid Spread Model.

$$\begin{cases} RATE_{infection} = -0.01 & \lambda_{i,t} < 0 \\ RATE_{infection} = 0 & 0 < \lambda_{i,t} < 1 \\ RATE_{infection} = 0.1 & 1 < \lambda_{i,t} < 2 \\ RATE_{infection} = 2 & \lambda_{i,t} > 2 \end{cases}$$

4.2 Evaluation:Spread & Characteristics



(a) Drug cases in Kenturky (b) Drug cases in Ohio (c) Drug cases in Pennsylvania



(d) Drug cases in Virginia (e) Drug cases in West Virginia

Figure 3: Opioid Spread in Each State Over Time (2010-2017)

To better illustrate the spread and characteristics in and between the five states, we view the five states as five blocks, and in each block, the cells change over time as they receive and give social influence to their neighbors. After each iteration, the grid is updated to reflect the modifications. Since this is a scenario-based model, the variables are set according to the provided data and adjacency data accessed from United States Census Bureau[5].

The simulation results are shown in figure 3, each grid represents one county, and the grid's changing color reflects the changes of total drug reports over the 8 years.

Based on the above five figures, we divide all counties into three categories: high-drug-risk county(the red grid), low-drug-risk county(the blue grid) and susceptible counties(green and yellow grids). The layout of the grid is arranged from small to large according to the county FIP. For example, the grid with the coordinates of (5, 3) represents the county with the FIP value of the state ranked 15 , and the corresponding points on the map can be launched from the grid.

As can be seen from figure 3.(a), in Kenturky, the high-risk county (the red grid) is stable, and the half bottom of the figure blur gradually, showing that the drug cases are spreading

to the neighbors. Besides, we can also observe a grid in the upper right corner that has changed from dark blue to green. After searching the grid's location on the map, we find that it is actually abutting the red grid. After analyzing the other four states , we can sum the spread and characteristics of drug cases in the states :

- The high-drug-risk county (the red grid) is stable and will always remain high-risk over time, and the drug cases in the neighboring low-drug-risk counties will increase as if they receive encouraging drug use signal.
- The low-drug-risk county(the blue grid) is stable and insusceptible to changes especially when it's surrounded by the low-drug-risk counties. If a low-drug-risk county (the blue grid) is surrounded by lower-drug-risk counties(the dark blue grid), the drug cases in the county will decrease as if it receives discouraging drug use signal.
- The middle-drug-risk counties (green and yellow grids) are vulnerable. The changes in this type of counties are largely dependent on the neighboring counties. If the neighbors are all middle-drug-risk counties, as time goes by, specific regions will transform to high-drug-risk county; If the neighbors are mostly low-drug-risk counties, the drug cases in the county will decrease slowly; If any high-risk-county is in its neighborhood, the drug cases in the county will increase quickly.

When considering the spread of drug cases between the states, we modified the model in one state and establish an overall adjacency matrix $\mathbf{M}$. The establish rule is also modified as follows:

$$\begin{cases} M_{ij} = M_{ji} = 0 & C_i \text{ and } C_j \text{ are not adjacent} \\ M_{ij} = M_{ji} = 1 & C_i \text{ and } C_j \text{ are adjacent and in the same state} \\ M_{ij} = M_{ji} = 0.5 & C_i \text{ and } C_j \text{ are adjacent but not in the same state} \end{cases}$$

where C denotes the county.

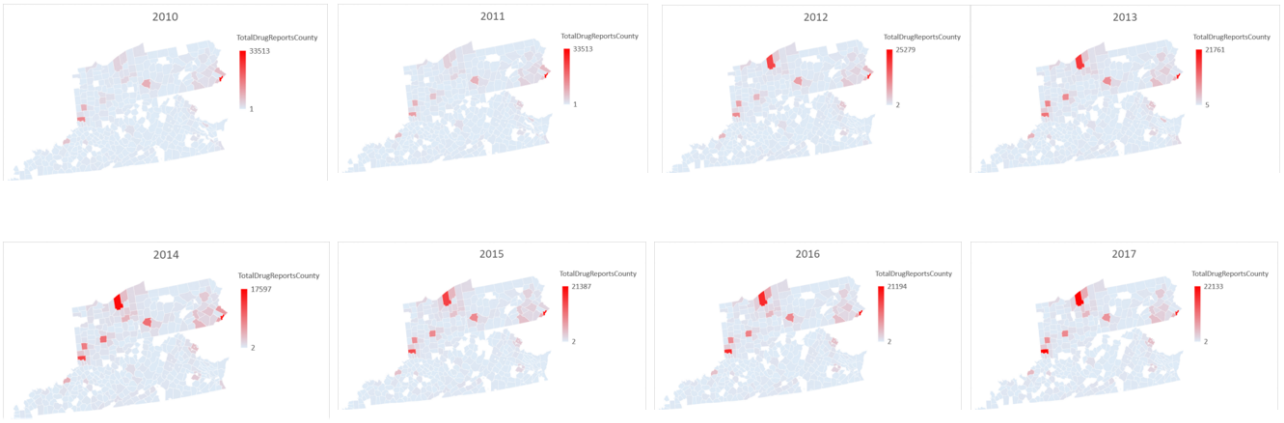


Figure 4: 2010-2018 drug spread overview

The simulation results are vividly shown in figure 4. We can observe:

- In the northwest corner (Ohio), the number of drug cases increases sharply.
- In the central region (Ohio and West Virginia), the number of drug cases increases steadily.

- In the south (Kenturky and Virginia), the number of drug cases stays at a stable and lower level.
- At the state - to - state border, the number of drug cases can easily reach a high level.

In summary, we can characterize the drug spread between the states : Drugs are mostly spread in the north (Ohio, West Virginia, Pennsylvania), while the south is hard to change; The boundary regions are high-risk area, which means the spread of drugs between states is mainly center of the border.

4.3 Traceback: the Origin of Specific Drug Abuse

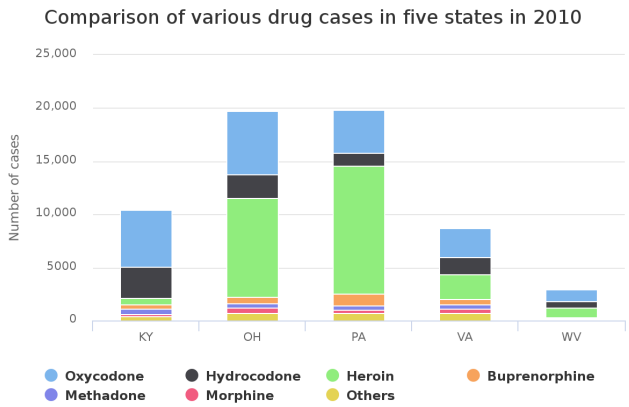


Figure 5: Comparison of various drug cases in five states in 2010

To find the locations where specific opioid starts to be used, we firstly count the data for the first year (2010), as shown in the figure 5, and select the drugs with the largest proportion of each state as possible opioid that might originate in the state. Then, for each county in the state, we use F_Score to illustrate the prevalence and stability of the possible opioid.

$$F\_Score(C_o) = M(C_o) - V(C_o) \tag{2}$$

where C_o denotes the specific opioid cases of one county, M denotes the mean of the opioid cases number during 2010-2017, V denotes the variance of the opioid cases number during 2010-2017 by calculating its mean and variance during 2010-2017. M and V are normalized to (0,1).

We assume that if a county is the origin of one opioid, then the opioid cases of this county will stay large and steady, which means F_Score is high. After processing the related data, we choose the county of the highest F_Score as the origin of this kind of opioid in each state. The result is shown in Table 2.

Table 2: Locations and specific drugs in five states

State	KY	OH	PA	VA	WV
Counties	Madison	Hamilton	Delaware	Fairfax	Raleigh
Specific Drugs	Oxycodone	Heroin	Heroin	Oxycodone	Oxycodone

4.4 Prediction:Specific Concerns

Drug addiction is a major public health problem that has a large impact on society, including on family and community , health , education , crime , work and employment. Ac-

cording to NIH, the overall cost of illicit drugs and opioids is 271.5 dollars in 2017 , adding a great burden to the society.

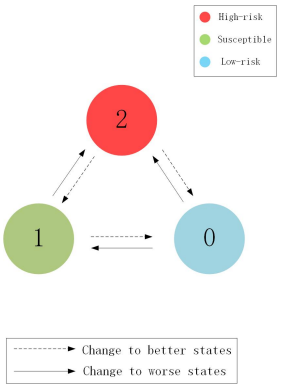


Figure 6: The state transition diagram

In order to find the specific place and time that those concerns occur, we divide the counties into three types which are low-risk county, susceptible county and high-risk county. The division standard is shown in table 3. We modified the cell states and update rules in our Opioid Spread Model. The state transition diagram is shown in figure 6. The cells alter their states based on the predicted number of total drugs of the cell. For instance, if the predicted number of total drugs is at level 1, the cell's next state will be type 1. Then we use the Opioid Spread Model we built to conduct the experiment.

Table 3: The division standard

level 1	TotalDrugReportsCounty$\leq$1000	low-risk county(type1)
level 2	1000$\leq$TotalDrugReportsCounty$\leq$10000	susceptible county(type2)
level 3	10000$\leq$TotalDrugReportsCounty	high-risk county(type3)

Using our modified model, we are able to calculate opioid incidents in each county in the next 100 years. In order to observe the change of the proportion of different types, we counted the proportion of different types of counties and draw the cell diagram to reflect the change of states. The simulation result (figure 7) shows that the number of high-risk counties is increasing slowly, so does susceptible counties. We also draw the proportion curves of three types which is shown in figure 8. From the picture, we can see that the proportion of low-risk counties decreases slightly and the proportion of high-risk counties and susceptible counties both show an increasing trend. Therefore, if without control, the opioid cases will becoming more and more common and spread from high-risk areas to susceptible ones.

So we can conclude that **if the patterns and characteristics our team identified continue, there will be more counties switch from low-risk state to susceptible state, even high-risk state under mutual influence.** Therefore, the U.S. government should have the following specific concerns:

- where are the opioid incidents significantly increase?
- how many high-risk counties are there?
- how many susceptible counties will be converted into high-risk ones?

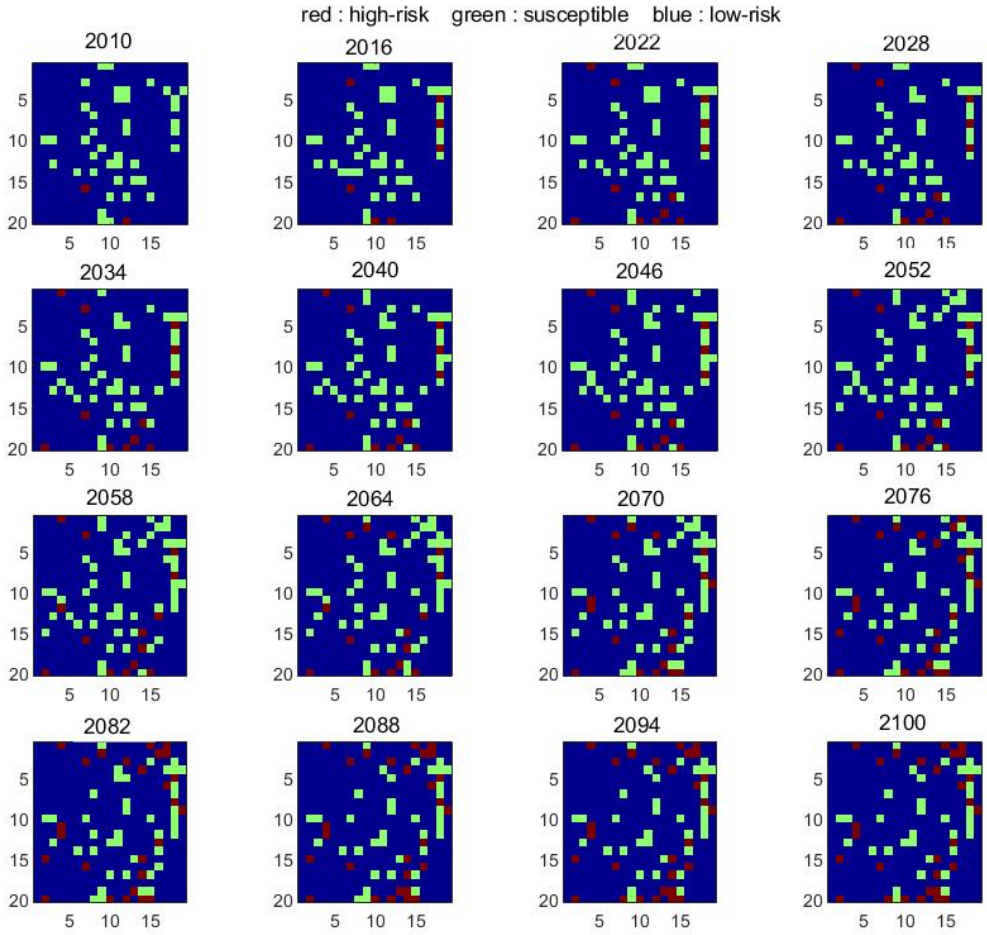


Figure 7: The distribution of counties of different types over time : three colors represent three different types

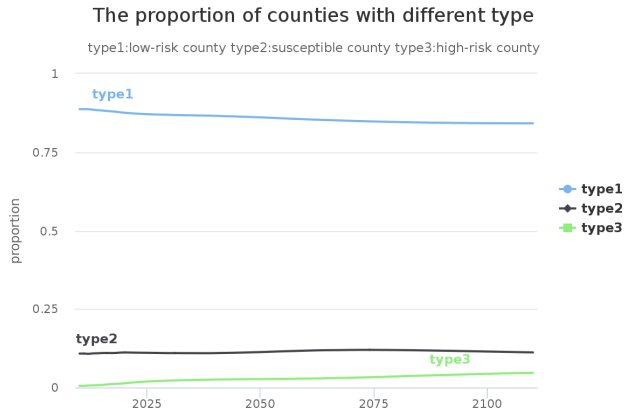
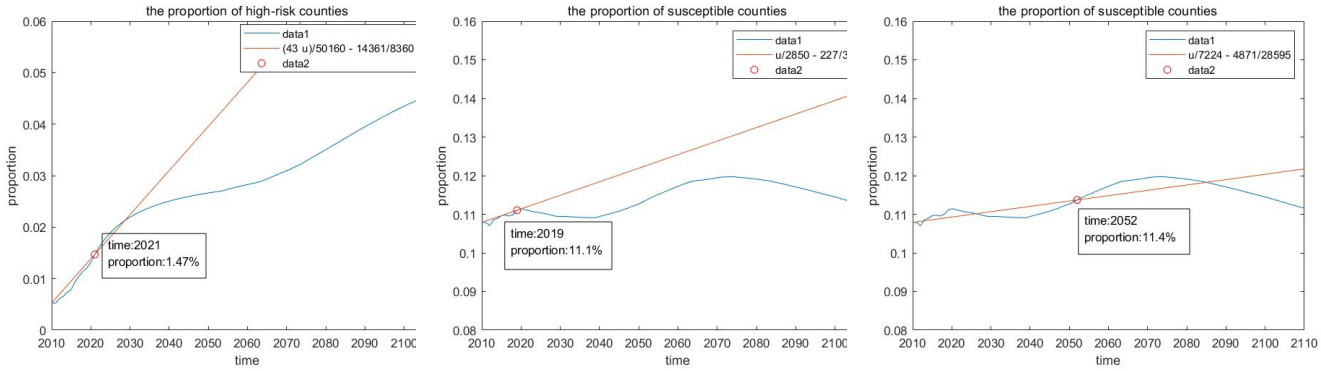


Figure 8: The proportion curves of three types from 2010 to 2110

- how many low-risk counties will be converted into susceptible counties and high-risk ones?

In order to locate where they will occur and find out when they will occur, we find the point with the highest rate of change in the proportion curve (figure 8) of susceptible and high-risk areas as shown in the figure $9a$, $9b$ and $9c$. According to picture $9a$, we can find the high-risk counties increase the fastest in the year 2021 with the proportion of 1.47%. From

picture $9b$, we can find the susceptible counties increase the fastest in the year 2019 with the proportion of 11.1%. But, after 2019, the proportion of susceptible counties decreases slowly. The third picture $9c$ indicates that the number of susceptible counties increases fast in the year 2052 with the proportion of 11.4%. And, after the year of 2052, the proportion of susceptible counties reaches a peak. **We can choose the proportion of high-risk counties in 2021 which is 14.7$\%$ as the drug identification threshold levels.** This level indicates there is a significant increase in opioid cases and the U.S. government's concerns may occur.



(a) high-risk (b) susceptible (c) susceptible

Figure 9: Find the point where the change rate is the biggest : (a)in the picture of the proportion of high-risk counties. (b)in the picture of the proportion of susceptible counties. (c)in the picture of the proportion of susceptible counties.

We can find the counties where the concerns will take place by observing the predict result. Considering concern 1 and 4, we choose the counties which change their state from low-risk into susceptible in the 2052 predict result as the places they occur. Considering concern 1 and 3, we choose the counties which change their state from susceptible into high-risk in the 2021 predict result as the places they occur. The final result is listed in the table 4. The prediction results of 2021 and 2052 are shown in figure 10.

Table 4: Places concerns occur

CONCERN	counties
concern 1 and 3 (occur in 2021)	GREEN ,KY LEWIS ,KY AMELIA ,VA CRAIG ,VA JAMES CITY ,VA
concern 1 and 4 (occur in 2052)	GALLATIN ,KY MARSHALL ,KY MARTIN ,KY MORGAN ,OH JUNIATA ,PA BATH ,VA DICKENSON ,VA DINWIDDIE ,VA KING WILLIAM ,VA NELSON ,VA POWHATAN ,VA PRINCE EDWARD ,VA RAPPAHANNOCK ,VA MINERAL ,WV

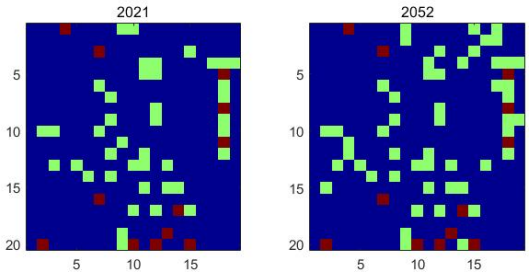


Figure 10: The prediction result of 2021 and 2052

5 Feature Selection Model and Modified Opioid Spread Model

In this section, we use Feature Selection Model based on Chi-squared test to select the socio-economic factors that contribute most to drug abuse. Then with the selected factors, we modify our model.

5.1 Feature Selection Model: Socio-economic Factors' Impact

Opioid prevalence is affected in a number of ways by individual socio-economic status and environmental socio-economic factors at both the family and community levels.[6] To find the most influential factors, we use Chi-squared test to evaluate the influence of each factor.

In Machine Learning, chi-square test is often used for categorical features of a dataset. Based on provided data in each year, we selected top100 counties as high-risk counties and bottom100 as low-risk counties, and design a classification task. We establish categorical feature array X($200 \times num\_features$) and target y(200×1), then calculate Chi-square between each feature and the target. After that , we rank the features by Chi-square scores. Chi-square scores determine whether the association between two categorical variables of the sample would reflect their real association with the population.

Chi- square score is given by :

$$\chi = \frac{(Obervedfrequency - Expectedfrequency)^2}{Expectedfrequency}$$

where:

- Observed frequency = No. of observations of class
- Expected frequency = No. of expected observations of class if there was no relationship between the feature and the target.

Table 5: Main Category and its Most Influential Subcategory

Main Category	Most Influential Subcategory
Family	
households by type	Nonfamily households
relationship	Spouse
marital status	Never married
fertility	Women who had a birth in the past 12 months
grandparents	Grandparents living with own grandchildren under 18
ancestry	German
Social	
disability status	Total Civilian Noninstitutionalized Population-18 to 64
school enrollment	Elementary school (grades 1-8)
educational attainment	High school graduate
veteran status	Civilian population 18 years and over
Race	
residence 1 year ago	Different house in the U.S.
place of birth	Native - Born in United States - State of residence
U.S. citizenship status	Foreign-born population
year of entry	Population born outside the United States
world region of birth of foreign born	Foreign-born population, excluding population born at sea
language spoken at home	English only

Using pandas and scikit-learn, we get the results and identify the most influential subcategory in each main category, as is listed in Table 5. From the table, we can infer:

1. **How opioid use got to its current level?** We think that for a county, the current opioid use level is caused by its own socio-economic factors and drug prevalence level ,drug prevalence level in neighboring counties and time . The more poor, lonely, and low-educated people in a county, the higher the level of drug use in the county. Moreover, being surrounded by high-risk counties also contributes to the increase in opioid use.

2. **Who is using/abusing it?** Poor people(disabled,migrants) , lonely people(never married,nonfamily households,veteran) and poorly-educated people(high school graduates) are abusing it.

3. **What contributes to the growth in opioid use and addiction?** The number and percentage of poor people(disabled, migrants) , lonely people(never married, nonfamily households,veteran) , poorly-educated people(high school graduates) , the number of drug abuse cases in this county and in neighboring counties.

4. **Why opioid use persists despite its known dangers?** Most of the people who use it suffers poverty , loneliness or even mental illness , which can't be easily tackled by drug preventive education. Despite its known dangers, those people still prone to drugs for relief.

Also, as Table 6 shows, we select the top three most influential attributes which will be used to modify our Opioid Spread Model.

Table 6: The top three most influential factors

Factors	chi2 statistics
VETERAN STATUS - Civilian population 18 years and over	15836053.62
DISABILITY STATUS - Total Civilian Noninstitutionalized Population-18 to 64 years	4238226.403
MARITAL STATUS - Never married	2906052.983

5.2 Modified Opioid Spread Model

To modify our original model, we take the given U.S. Census socio-economic data into consideration. After Screening the characteristics of each county, we picked out some effective features which can be used to distinguish high-risk counties from low-risk counties. To simplify, we chose the top three factors. The three factors we selected are N_{ve}(the number of Veterans in a county), N_{di}(the number of disabled individuals in a county)and N_{nm}(the number of people who never married). Figure 11 illustrates the whole logic of our modified Opioid Spread Model.

Algorithm 2: Modified CA model

Input: $TotalCounty$:total number of the county
$Matrix_{adj}$:reflect the connection between two counties
$state_{ini}$:the initial number of "TotalDrugReportsCounty"
$iter$:the iteration number
$lambda$:the affection of neighbor
mu:the stochastic death rate
$Rate_{infec}$:the infection rate;

Output: $TOTALstate$

for $i = 1; i \leq iter; i++$ **do**
  $RATE = \frac{1}{3} \times (RATE_{adjacent} + RATE_{death} + RATE_{infection} + RATE_{socio-economic})$
  $NewState = OrinState \times (RATE + 1)$
  update RATE
  draw picture
end
return TOTALstate;

Socio-economic factors:

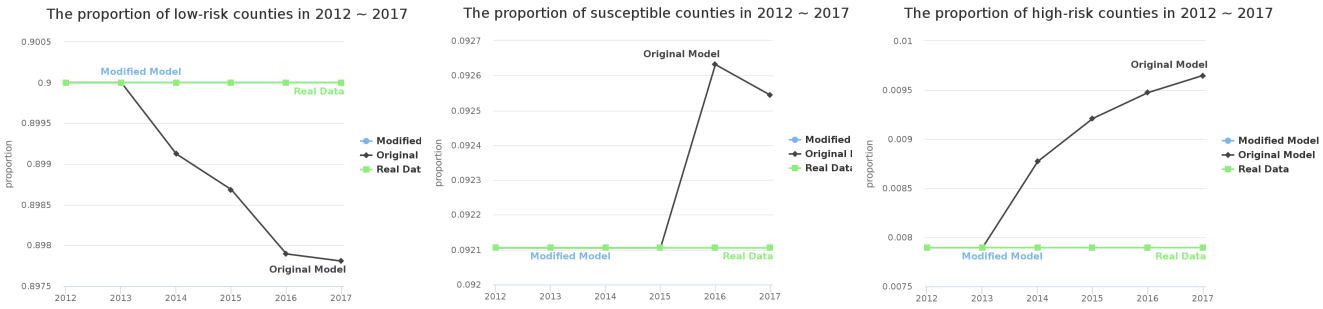
(1)VETERAN STATUS - Civilian population 18 years and over ;
(2)DISABILITY STATUS OF THE CIVILIAN NONINSTITUTIONALIZED POPULATION - 18 to 64 years ;
(3)MARITAL STATUS - Females 15 years and over - Never married

Figure 11: The pseudo-code for modified Opioid Spread Model

Each of the selected factors has a influence factor. We use the normalized influence factors as the weight. Weight coefficients are expressed as W_{ve}(the weight coefficient of Veterans), W_{di}(the weight coefficient of disabled individuals)and W_{nm}(the weight coefficient of people who never married).

In our modified Opioid Spread Model, we change the update rules based on the additional three factors we selected. The latest update rules are as follows.

$$C_{i,t+1} = C_{i,t} \times (1 + \frac{1}{3} \times (RATE_{adjacent} + RATE_{death} + RATE_{infection} + RATE_{socio-economic}))$$



(a) low-risk (b) susceptible (c) high-risk

Figure 12: The comparison of real data, original model and modified model in the proportion of three types of counties

The revise adds to the influence of three kinds of people who are possible to contribute most to the increase in total opioid cases. To testify the effectiveness of the revise in our model, we use the original model and the modified model to predict the proportion of counties with three types from 2012 to 2017. Then, we compare the result with the given real data.

Figure 12a, 12b and 12c show that the modified model can precisely reflect the real situation. Both the real data and the prediction of modified model show that the proportion of counties with three types from 2012 to 2017 remains the same. It indicates that the change rate of the proportion is small.

We can use our modified model to recalculate the distribution of three types of counties. The prediction results can be shown in figure 13. Compared with the prediction of the original model, we can notice that the proportion of high-risk areas decreases while the proportion of susceptible areas increases.

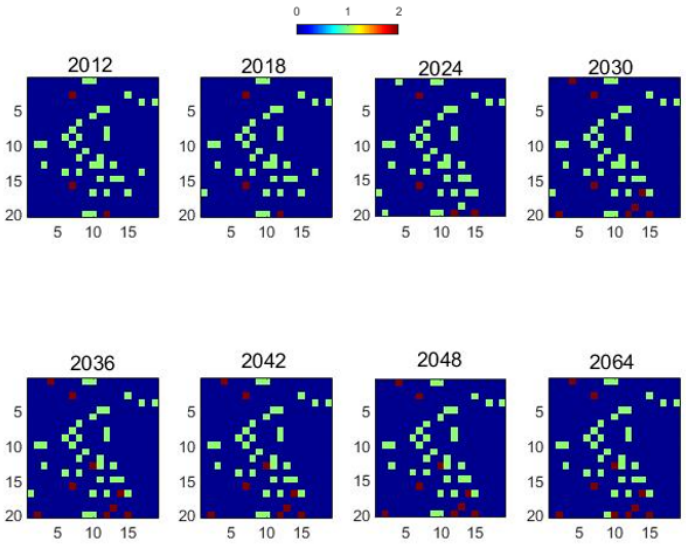


Figure 13: the prediction result of the modified model

5.3 Drug Strategy

In this subsection, we divide our strategy into two categories: prevention & treatment and transmission.

5.3.1 Prevention & Treatment

As Table 5 shows, poor people(disabled,migrants) , lonely people(never married,non family households,veteran) and poorly-educated people are susceptible to drugs. Therefore, we propose targeted approach for high risk groups.

- **Early Interference**. While substance use generally begins during the adolescent years, there are known biological, psychological, social, and environmental factors that contribute to the risk that begin accumulating as early as the prenatal period. And for children born in high-risk communities, the government can introduce policies to promote the development of projects such as Caring School Community Program [7], a program focuses on strengthening students' "sense of community", which research has shown to be pivotal in reducing drug use, violence, and mental health problems and promoting academic motivation and achievement.
- **Financial and Psychological Aid** For poor people (DISABILITY STATUS - Total Civilian Noninstitutionalized Population-18 to 64 years) and lonely people (veteran and

never married), they sometimes use drugs to cope with the physical and psychological effects. These risks can be reduced if their mental health is improved and have the ability to find and hold fulfilling employment and secure accommodations. Therefore, the government can work with local community to provide psychological aid like free psychological counseling, and establish funding to provide job opportunities and financial assistance.

- **Treatment** The government can establish and support institutions to research and give more effective and personalized treatment solutions . Moreover, the government should also develop inspectorate to supervise the commissioning of drug treatment and recovery services and the impact this can have on recovery outcomes for individuals and communities.

5.3.2 Transmission

There are many ways to transmit opioids, including the sale of pharmacies, the prescribing of doctors, sharing with others, and even the illegal inflow of foreign countries. These have led to the spread of opioids. In order to curb the spread of opioids, we should improve the strategy also for opioid transmission.

- **Strengthen opioid control in the United States.** Nowadays, drug control in the United States is very loose, and opioids can be purchased at any pharmacy. Also, many doctors do not pay attention to the safety of opioids, and provide opioids to patients at will. This has largely contributed to the transmission of the opioids.

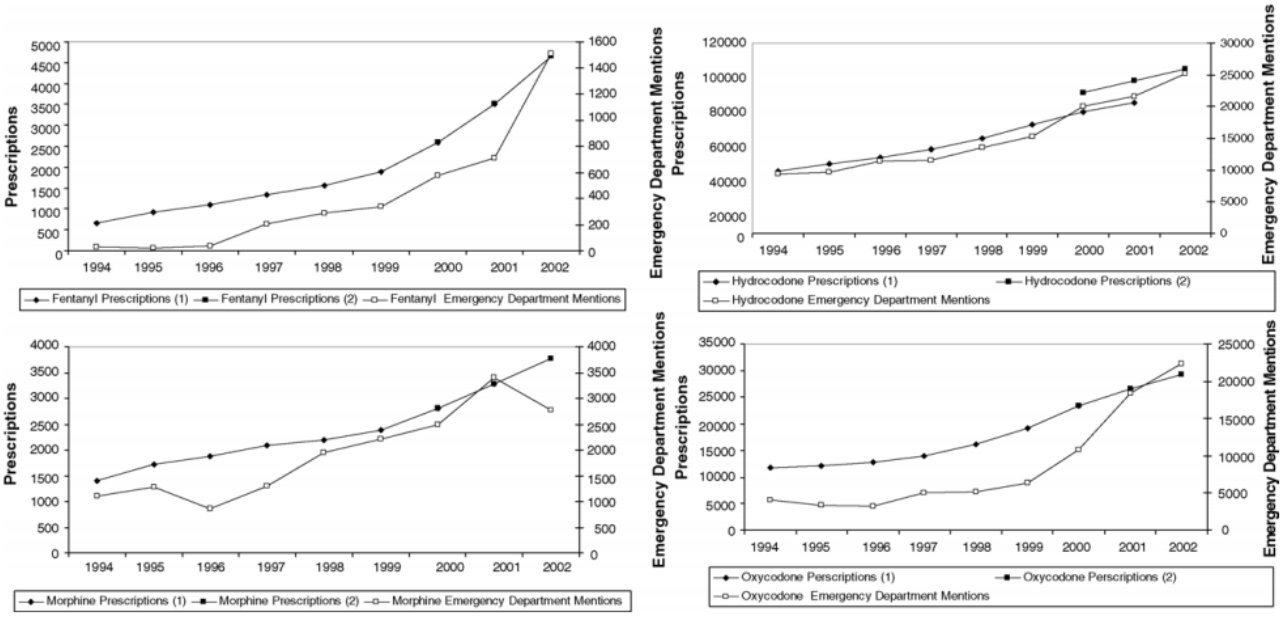


Figure 14: Prescriptions of four opioids in recent years [8]

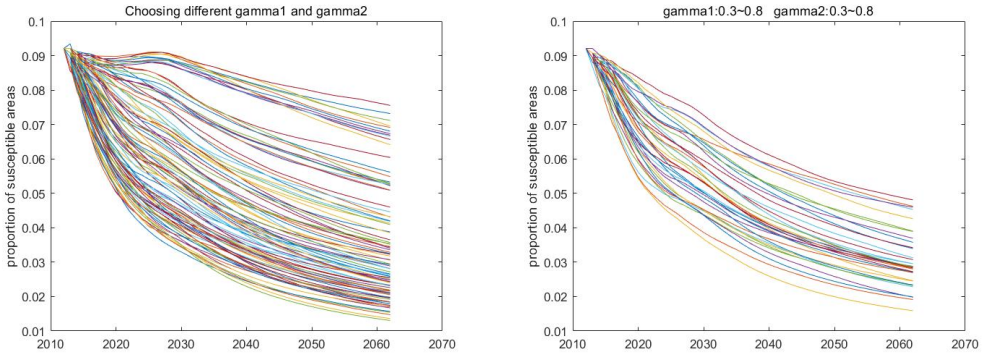
Figure 14 shows the growing opioid prescriptions. President Trump proposes training for opioid safety medications for medical personnel employed by the federal government. This is a significant measure.

- **Prevent the flow of opioids into the United States.** On April 21, 2016, American superstar Prince died, and his death was due to the large amount of Fentanyl (opioid

analgesics). Over the years, Fentanyl (opioid analgesics) has flowed into the United States and has become a serious drug in the opioid crisis. In the United States, many anonymous sales and purchases of such drugs have been discovered on the Internet. In June last year, U.S. Customs and the border stated that nearly 90 kilograms of Fentanyl had been investigated. The U.S. government should step up efforts to control entry drugs.

5.4 Effectiveness Test for Our Strategy

By using the strategy we proposed, we assume that the rate of drug spread between two adjacent counties will decrease and the number of people we find in task 2 who is using or abusing drugs will also on a declining trend. Based on these hypotheses, we add two parameters to our modified model, the first one is γ_1 which acts as the max value of a random variable (p) that can reduce the $RATE_{socio-economic}$, the second one is γ_2 which aims to reduce the $RATE_{adjacent}$. p is a random variable which ranges from 0 to γ_1 and varies in each iteration. Parameter γ_1 is a constant ranging from 0.1 to 1. γ_2 is a constant set artificially which ranges from 0.1 to 1. Through choosing different values of these two parameters, we can observe the the effectiveness of the strategy and decide the bound of the parameters. We judge if the strategy is effective according to the value of the reduced percentage of susceptible counties. The experiment result is show in figure 15a and 15b. From the two pictures we can identify the bound of parameter γ_1 and γ_2. The strategy will work successfully if γ_1 is in the range of 0.3 to 0.8 and γ_2 is in the range of 0.3 to 0.8.



(a) $\gamma_1 : 0.1 \sim 1, \gamma_2 : 0.1 \sim 1$ (b) $\gamma_1 : 0.3 \sim 0.8, \gamma_2 : 0.3 \sim 0.8$

Figure 15: The proportion of susceptible counties of different parameter values

By choosing $\gamma_1 = 0.4$, $\gamma_2 = 0.6$, the prediction result of the distribution of different types of counties is shown in figure 16. We can clearly see from the prediction result that the proportion of susceptible counties decreases significantly from 2018 to 2036. Although the decrease of the proportion of high-risk counties is not as obvious as the susceptible counties, the transition from susceptible areas to high-risk areas declines. Those conclusions illustrate that our strategy effectively prevents susceptible ones from changing into high-risk ones. At the same time, there are more and more susceptible ones change into low-risk ones which means the strategy hinders the spread of drugs and reduce the number of people who used to using or abusing drugs.

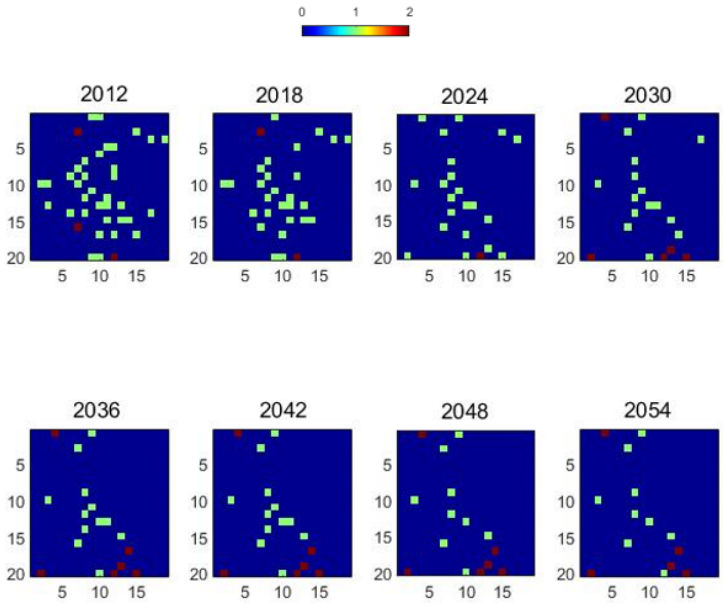


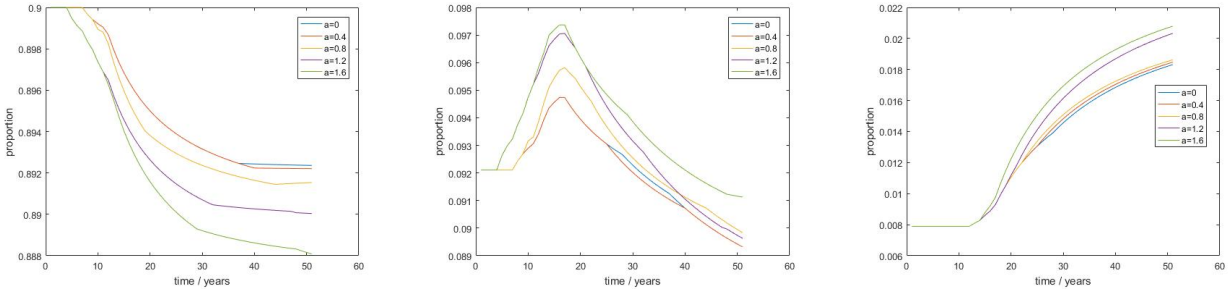
Figure 16: The prediction result : after using the strategy

6 Sensitivity Analysis

For our Opioid Spread Model, the structure of the model is decided by two parameters: α and μ. Parameter α is related to the spread rate while parameter μ is related to the death rate. Both parameters are determined by our test. Therefore, these two parameters are uncertain. We need to analyze the sensitivity of the model to them.

6.1 Impact of the Parameter α in Opioid Spread Model

The original value of parameter α is 1. We will obtain different results by changing the value of the parameter α range from 0 to 1.8. The range of variation is 100%. Figure 17 shows the change in the proportion of the three types of counties.



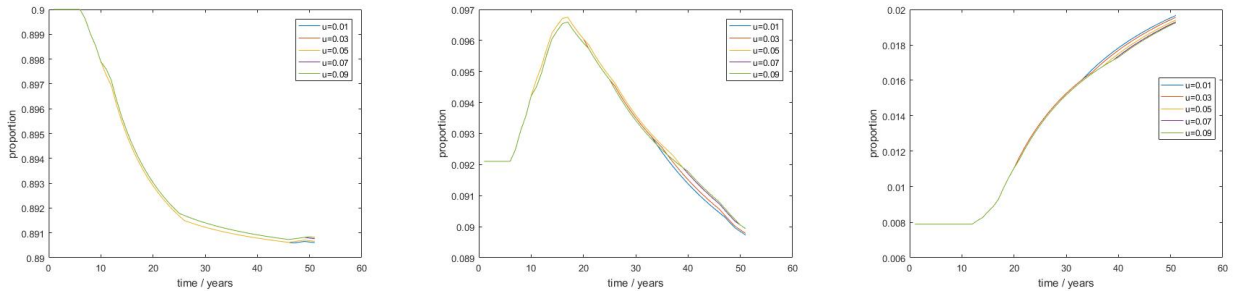
(a) Proportion of low-drug-risk counties (b) Proportion of susceptible counties (c) Proportion of high-drug-risk counties

Figure 17: Sensitivity Analysis of Parameter α in Opioid Spread model

It can be concluded from Figure 17 that the proportion is sensitive to the parameter α while α is taken as a relatively large value. However, when α is taken as a small value(close to 0), the proportion is much less sensitive to α.

6.2 Impact of the Parameter μ in Opioid Spread Model

The original value of parameter μ is 0.05. We will obtain different results by changing the value of the parameter α range from 0.01 to 0.09. The range of variation is nearly 100%. Figure 18 shows the change in the proportion of the three types of counties.



(a) Proportion of low-drug-risk counties (b) Proportion of susceptible counties (c) Proportion of high-drug-risk counties

Figure 18: Sensitivity Analysis of Parameter μ in Opioid Spread Model

Obviously, it can be concluded from Figure 18 that with the change of μ, the proportion of the three types of counties hardly change. So we can know that our Opioid Spread Model is not sensitive to parameter μ.

7 Strengths and Weakness

7.1 Strengths

- **Reliable Calculations:** Rich and detailed data guarantees the reliability of our calculations for each county in five states.
- **Accurate to the County's Research Objects:** Our Opioid Spread Model is studied in every county with high accuracy.
- **Comprehensive Consideration of Influencing Factors:** When analyzing socio-economic factors' impact, we select all factors of the known data without missing.
- **Practical Strategy:** After the model verification, our proposed strategy has a good effect in countering the opioid crisis.
- **Expandable Model:** As long as data is available, our model can calculate the distribution of drug cases in other counties and even other states.

7.2 Weaknesses

- **No Involvement of Other States:** We do not consider states other than KY, OH, PA, VA, and WV, due to the lack of relevant data.
- **No Consideration of Migration:** To simplify our model, we assume that there are no large-scale migrations, but in fact small-scale migrations exist. This brings a slight error.

8 Conclusion

In this paper, firstly, we establish the original Opioid Spread Model by analyzing the NFLIS data and adjacent data. Secondly, we use our model to characterize the spread and characteristics of the reported synthetic opioid and heroin incidents, trace back the drug origins in each state and predict the future trend of drugs cases. By observing the predict result, we propose possible concerns the U.S. government should have and give the results about when and where the concerns may occur. Then we use the selected socio-economic features to modify our model. Next, we propose a strategy targeting on mitigating the opioid crisis. Using the modified Opioid Spread Model, we test the effectiveness of our strategy. What's more, in order to make the strategy work successfully, we set a bound to the two parameters γ_1 and γ_2. Finally, we conduct sensitivity analysis of parameter α and μ in our Modified Opioid Spread Model and discuss the strengths and weakness of our work.

9 Memo

To: Chief Administrator
From: Team 1919054
Date: January 28, 2019

Recently, we do some data mining based on the DEA / NFLIS database. During modeling, we got some significant results which can be used to alleviate drug abuse problem, so I am writing this memo to you to introduce our work and results.

We study the spread, characteristics,origin, trend and related socio-economic attributes of drug abuse and derive the following results:

- **Model Establishment**. Based on the DEA / NFLIS database, we established a Opioid Spread Model based on Cellular Automata capable of depicting drug spread characteristics and can be used for evaluation, traceback and prediction.
- **Evaluation**. From our original Opioid Spread Model , we obtained the spread and characteristics of drug cases, as is shown in Figure19.
  - **Spread**: In one state, drugs are mainly spread between neighboring counties. The more the number of drugs, the greater the impact on the surrounding counties. Between states, drugs are mainly spread between adjacent states, and on the state's borders, the proliferation of drugs is prone to occur.
  - **Characteristics** : We found that around the high-drug-risk region, as well as the middle-drug-risk region, the number of drugs is prone to a surge in the number of drugs.

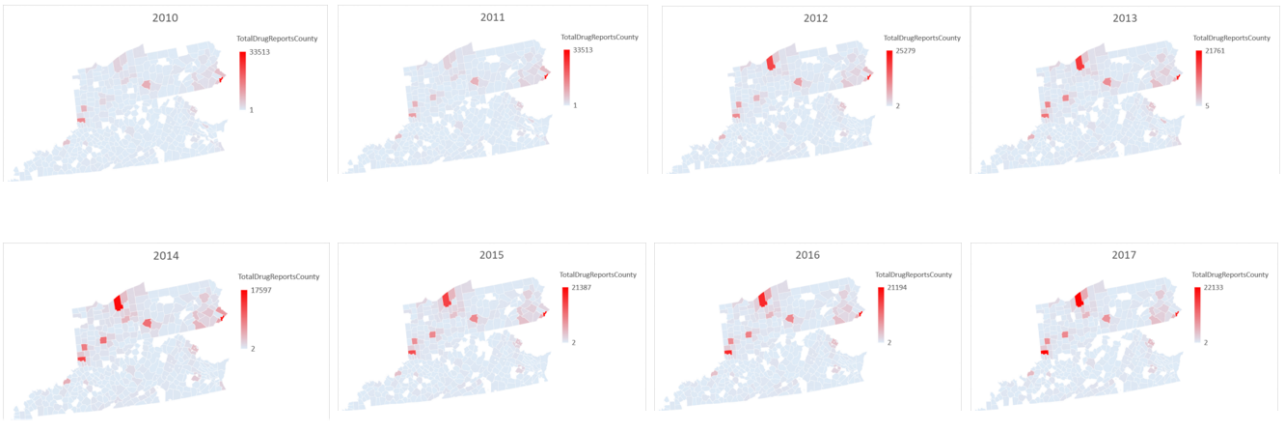


Figure 19: 2010-2018 opioid spread overview

- **Traceback**. Based on the Opioid Spread Model we obtain, we devise F_Score to evaluate one county's possibility to be the origin of a specific opioid. F_Score is correlated to seven (2010-2016) years' mean and variance of the specific opioid reports in one county. The results can help you to identify the origin. We selected the counties with highest F_Score and is shown in Table 7.
- **Prediction**. After defining the concerns according to related literature, by applying Opioid Spread Model iteratively, we get the occurring time and location for specific concerns, as is shown in Table 9

Table 7: Locations and specific drugs in five states

State	KY	OH	PA	VA	WV
Counties	Madison	Hamilton	Delaware	Fairfax	Raleigh
Specific Drugs	Oxycodone	Heroin	Heroin	Oxycodone	Oxycodone

- **Feature Selection and Model Modification**. To attain the most important socio-economic factors, we construct a Feature Selection Model based on Chi-Square Test, as is shown in Table 8. According to the result, we modified our Opioid Spread Model.

Table 8: The top three most influential factors

Factors	chi2 statistics
VETERAN STATUS - Civilian population 18 years and over	15836053.62
DISABILITY STATUS - Total Civilian Noninstitutionalized Population-18 to 64 years	4238226.403
MARITAL STATUS - Never married	2906052.983

- **Strategy Effectiveness Test and Bound Identification**. We proposed a strategy and applied the Modified Opioid Spread Model to evaluating the strategy's effectiveness, and the evaluation results is fairly good. Our strategy is :
  - **Prevention & Treatment**. Make early intervention, offer financial and psychological aid, give effective treatment, and arrage long-term tracking for the target people in Table 8.
  - **Transmission-blocking**. Strengthen opioid control and prevent the flow of foreign opioids.

Table 9 :Specific concern, occurring time and location

CONCERN	YEAR	COUNTIES
The surge of opioid cases & Large portion of susceptible counties transform into high-risk ones	2021	GREEN ,KY
		LEWIS ,KY
		AMELIA ,VA
		CRAIG ,VA
		JAMES CITY ,VA
The surge of opioid cases & Large portion of low-risk counties transform into susceptible counties or high-risk ones	2052	GALLATIN ,KY
		MARSHALL ,KY
		MARTIN ,KY
		MORGAN ,OH
		JUNIATA ,PA
		BATH ,VA
		DICKENSON ,VA
		DINWIDDIE ,VA
		KING WILLIAM ,VA
		NELSON ,VA
		POWHATAN ,VA
		PRINCE EDWARD ,VA
		RAPPAHANNOCK ,VA
		MINERAL ,WV

Hopefully our results and methods can help you.

Sincerely,

Team 1919054

References

[1] W. M. Compton and N. D. Volkow, "Major increases in opioid analgesic abuse in the united states: concerns and strategies," *Drug and alcohol dependence*, vol. 81, no. 2, pp. 103–107, 2006.

[2] J. Benneyan, J. Garrahan, I. Ilieş, and X. Duan, "Modeling approaches, challenges, and preliminary results for the opioid and heroin co-epidemic crisis," in *Simulation Conference (WSC), 2017 Winter*. IEEE, 2017, pp. 2821–2832.

[3] W. Wakeland, A. Nielsen, and P. Geissert, "Dynamic model of nonmedical opioid use trajectories and potential policy interventions," *The American journal of drug and alcohol abuse*, vol. 41, no. 6, pp. 508–518, 2015.

[4] B. Pfeifer, K. Kugler, M. M. Tejada, C. Baumgartner, M. Seger, M. Osl, M. Netzer, M. Handler, A. Dander, M. Wurz *et al.*, "A cellular automaton framework for infectious disease spread simulation," *The open medical informatics journal*, vol. 2, p. 70, 2008.

[5] "County adjacency file," United States Census Bureau. [Online]. Available: https://www.census.gov/geo/reference/county-adjacency.html

[6] V. Dabbaghian, V. Spicer, S. K. Singh, P. Borwein, and P. Brantingham, "The social impact in a high-risk community: A cellular automata model," *Journal of Computational Science*, vol. 2, no. 3, pp. 238–246, 2011.

[7] "rinciples of substance abuse prevention for early childhood," NIDA. [Online]. Available: https://www.drugabuse.gov/publications/principles-substance-abuse-prevention-early-childhood

[8] W. M. Compton and N. D. Volkow, "Major increases in opioid analgesic abuse in the united states: concerns and strategies," *Drug and alcohol dependence*, vol. 81, no. 2, pp. 103–107, 2006.

Appendices

Appendix A MAIN_SIMU.m

```
%% This is the main code
load('adj.mat');
load('ini.mat');
load('inc.mat');
load('weight.mat')
load('realp.mat')
TotalCounty=380;
weight=weight';
ADJ_matrix=adj;
ini_state=new_ini;
iter=50;     %2012-2062
error=0;
min=10000;
key1=0;
key2=0;
mu=0.05;
inc_rate=new_inc;
infec_rate=LEVEL_OF_INFECT(TotalCounty,inc_rate,iter);
p2=zeros(100,iter+1);
% rate_sus=zeros(100,iter);
count=0;
```

```
for lambda=0.3:0.1:0.8
    for gamma=0.3:0.1:0.8
        count=count+1;
        add_rate=AddFactor(weight,gamma);    % 380 * year
        [~,~,~,TOTALstate,TotalLevel,total_rate]=CA(TotalCounty,...
        ADJ_matrix,ini_state,iter,lambda,mu,infec_rate,add_rate);
        p=propOfThreeLevels(TotalLevel,iter);
        p2(count,:)=p(2,:);  %susceptible counties
%         temp=zeros(1,iter);
%         for k=1:iter
%             temp(k)=p2(k+1)-p(k);
%         end
%         rate_sus(count,:)=temp;
    end
end
```

Appendix B CA.m

```
function [~,~,~,TOTALstate,TotalLevel,total_rate]=CA(TotalCounty,...
ADJ_matrix,ini_state,iter,lambda,mu,infec_rate,add_rate)
% The function applies CA model
% input:TotalCounty:total number of the county
%       ADJ_matrix:reflect the connection between two counties
%       ini_state:the initial number of "TotalDrugReportsCounty"
%       iter:the iteration number
%       lambda:the affection of neighbor
%       mu:the stochastic death rate
%       infec_rate:the infection rate
% output:TOTALstate:the simulation result
% state transition:
% state 1:low-risk county=0
% state 2:susceptible county=1
% state 3:high-risk county=2
% The redder the color, the more the number
% The bluer the color, the less the number
%% I. initialize the state
% find the initial level of each county
TotalLevel=zeros(TotalCounty,iter+1);   %all levels over time
LEVEL=FindLevel(ini_state);
TOTALstate=zeros(TotalCounty,iter+1);
total_rate=zeros(TotalCounty,iter);
% z=zeros(TotalCounty,iter+1);
o=ones(TotalCounty,1);
state=ini_state;
subplot(2,4,1)
imagesc(reshape(LEVEL,20,19));
colormap('jet')
axis equal
axis tight
TOTALstate(:,1)=state;
TotalLevel(:,1)=LEVEL;
[rate_matrix,q,newq]=RATE(ini_state,ADJ_matrix,lambda,mu);
k=2;
%% II. Update states and determine the county's risk level
for i=1:iter
    [rate_matrix,q,newq]=RATE(state,ADJ_matrix,lambda,mu);
    rate_matrix=0.3*rate_matrix+0.3*infec_rate(:,i)+0.3*add_rate(:,i);
    total_rate(:,i)=rate_matrix;
    state=state.*(rate_matrix+o);
    TOTALstate(:,i+1)=state;    %All states contain
    LEVEL=FindLevel(state);
    TotalLevel(:,i+1)=LEVEL;    %All levels contain
    if rem(i,6)==0 && i<48
        subplot(2,4,k)
        imagesc(reshape(LEVEL,20,19))
        colormap('jet')
        axis equal
        axis tight
        pause(1)
        drawnow
        k=k+1;
    end
end
end
```

Appendix C RATE.m

```
function [RATE,Q,newQ]=RATE(state,ADJ_matrix,lambda,mu)
% The function applies Opitiod Spread Model
% input:ADJ_matrix:reflect the connection between two counties
%       state:the initial number of "TotalDrugReportsCounty"
%       lambda:the affection of neighbor
%       mu:the stochastic death rate
%       infec_rate:the infection rate
% output:Q:a matrix stand for the change rate of "TotalDrugReportsCounty"
%% Calculate the RATE Matrix
n=size(state,1);
Q=zeros(n,1);
max=0;
for i=1:n
    for j=1:n
        if ADJ_matrix(i,j)==1
            if state(j)>max
                max=state(j);
            end
        end
    end
    Q(i)=max;
    max=0;
end
newQ=NORMALIZE(Q);
normstate=NORMALIZE(state);
RATE=lambda.*newQ-mu.*normstate;
end
```

Appendix D FindLevel.m

```
function [level_matrix]=FindLevel(state_matrix)
%this function aims to find the corresponding level
%three levels in total
%total drugs: 0-1000 low-risk(0)
%             1000-10000 susceptible(1)
%             10000- high-risk(2)
thr1=1000;
thr2=10000;
n=size(state_matrix,1);
level_matrix=zeros(n,1);
for i=1:n
    if state_matrix(i)<=thr1
        level_matrix(i)=0;
    elseif state_matrix(i)<=thr2
        level_matrix(i)=1;
    else
        level_matrix(i)=2;
    end
end
end
```

Appendix E LEVEL_OF_INFECT.m

```
function infec_rate=LEVEL_OF_INFECT(TotalCounty,inc_rate,iter)
infec_rate=zeros(TotalCounty,iter);
% This function aims to find out a county's infect risk
% input : TotalCounty : total number of counties
%         inc_rate : increase rate
% output : infec_rate : the infect risk
for i=1:iter
    if i<6
        for j=1:TotalCounty
            temp=inc_rate(j,i);
            if temp<=0
                infec_rate(j,i)=-0.01;
            elseif temp<=1
                infec_rate(j,i)=0;
            elseif temp<=2
                infec_rate(j,i)=0.1;
```

```
            else
                infec_rate(j,i)=2;
            end
        end
    else
        for j=1:TotalCounty
            temp=inc_rate(j,5);
            if temp<=0
                infec_rate(j,i)=-0.01;
            elseif temp<=1
                infec_rate(j,i)=0;
            elseif temp<=2
                infec_rate(j,i)=0.1;
            else
                infec_rate(j,i)=2;
            end
        end
    end
end
```

Appendix F propOfThreeLevels.m

```
function [prop]=propOfThreeLevels(TotalLevel,iter)
% This function aims to find the proportion of different kinds of counties.
prop=zeros(3,iter+1);
total=zeros(1,3)
n=size(TotalLevel,1);
for i=1:iter+1
    for j=1:n
        if TotalLevel(j,i)==0
            total(1)=total(1)+1;
        elseif TotalLevel(j,i)==1
            total(2)=total(2)+1;
        else
            total(3)=total(3)+1;
        end
    end
    for k=1:3
        prop(k,i)=total(k)/sum(total);
    end
end
%% Judge if the proportion reach the threshold
threshold=
for i=1:iter+1
    if prop(3,i)>=threshold
        time=mat2str(2009+iter);
        place=i;
    end
end
```

Appendix G FindLoc.m

```
%% find locations
function [loc]=FindLoc(TotalLevel)
load('TotalLevel.mat')
loc=zeros(2,380);
ini=TotalLevel(:,8);
y21=TotalLevel(:,12);
y52=TotalLevel(:,43);
k=1;
k1=1;
for i=1:380
    if ini(i)==0 && y21(i)==2
        loc(1,k)=i;
        k=k+1;
    end
    if ini(i)==0 && y52(i)==1
        loc(2,k1)=i;
        k1=k1+1;
    end
end
end
```

Appendix H NORMALIZE.m

```
function res=NORMALIZE(matrix)
% This function is to normalize the input matrix
n=size(matrix,1);
sum_all=sum(matrix);
res=matrix./sum_all;
end
```

Appendix I pic_prop.m

```
load('prop.mat')
x=2010:1:2110;
plot(x,p(2,:))
hold on
max_ratio=0;
point=1;
for i=10:100
    temp=p(2,i+1)-p(2,i);
    if temp>max_ratio
        max_ratio=temp;
        point=i+1;
    end
end
k=(p(2,point)-p(2,1))/(point-1);
syms u;
v=k*(u-2010)+p(2,1);
fplot(v)
xlim([2010,2110])
ylim([0.08,0.16])
hold on
scatter(point+2009,p(2,point),'r')
```

Appendix J AddFactor.m

```
function [AddRate]=AddFactor(weight,gamma)
% this function is to calculate the additional
% increase caused by socio-economic factors
% input : the weight matrix of selected factors
% output : the additional increase
load('sub.mat');          % load the socio-economic data(380*15)
k2=0;
AddRate=zeros(380,50);  % 2012-2062
for i=1:50
    p_decre=rand(380,1)*gamma;
    if i<6
        k1=k2+1;
        k2=k1+2;
        newWeight=NORMALIZE(weight(:,i)); % 3*1
        subs=sub(:,k1:k2);   %380 * 3
        AddRate(:,i)=NORMALIZE(subs*newWeight)-p_decre;
    else
        AddRate(:,i)=AddRate(:,5);
    end
end
end
```