



Research article

Research and prediction of opioid crisis based on BP neural network and Markov chain

Wanchun Fan^{1,†}, Yan Jiang^{2,†}, Songyang Huang^{1,†} and Weiguo Liu^{2,*}

¹ Department of Finance, GuangDong University of Finance & Economics, China

² Department of statistics and mathematics, GuangDong University of Finance & Economics, China

† These authors contributed equally to this work and should be considered co-first authors.

* **Correspondence:** Email: david.liu@gdufe.edu.cn; Tel: +86-13322959344.

Abstract: Nowadays, in the United States, opioid abuse is so serious that it has become a crisis, causing serious health impacts and huge losses to the US economy. To study the social and economic data and the relationship between opioid drug abuse situation, this paper uses Grey Relation Analysis to analyze the identification and counting of synthetic opioids and heroin provided by NFLIS and the data related to socioeconomic factors provided by the United States Census Bureau. After that, the orderly clustering is used to introduce the corresponding level. Then, BP neural network and Markov model are built to forecast the degree of the flood of opioid. The Proportion of Low Education Level People, Number of New Pregnant Women, The Proportion of The Elderly Living Alone and the other 7 factors were selected as the input nodes of the neural network. Based on the prediction results of the BP neural network, Markov Chain is used to correct the residual sequence. It is found that the deviation of prediction is reduced from [-10.99%, 22.33%] to [-8.29%, 2.81%], making the modified value closer to the measured value. This method combines the advantages of the BP neural network and the Markov network, which improves the accuracy of prediction and provides certain reference values for opioid prediction. Finally, we propose strategies to address the opioid crisis and test their effectiveness.

Keywords: opioid crisis; sequential cluster method; grey correlation analysis; BP neural network; Markov chain

Mathematics Subject Classification: 62P25, 97K80, 97M10

1. Introduction

The United States is experiencing a national crisis regarding the use of synthetic and non-synthetic opioids, either for the treatment and management of pain (legal, prescription use) or for recreational

purposes (illegal, non-prescription use). It is said that opioid abuse is one of the greatest public health challenges of the 21st century. In the United States, this problem can date back to the 1980s. Some scholars think of opioid addiction is not high, and the major producers of opioid drugs also aggressively marketing its products in pledge, opioid drugs are not addictive, which makes the dosage of opioid present surge trend year by year. The National Forensic Laboratory Information System (NFLIS), as part of the Drug Enforcement Administration's (DEA) Office of Diversion Control, publishes a data-heavy annual report addressing "drug identification results and associated information from drug cases analyzed by federal, state, and local forensic laboratories". The database within NFLIS includes data from crime laboratories that handle over 88% of the nation's estimated 1.2 million annual state and local drug cases. For this problem, we focus on the individual counties located in five U.S. states: Ohio, Kentucky, West Virginia, Virginia, and Tennessee. According to statistics, the number of people taking oxycodone for non-medical use rose rapidly from 400,000 in 1999 to more than 2.8 million in 2003. It may be a huge opioid market that makes the motivation for many pharmaceutical companies to take risks. In recent years, opioids have accounted for more than half of drug overdose deaths in the United States. In terms of health insurance costs, post-addiction treatment, and criminal justice investigations, the United States is losing about \$80 billion a year to opioid abuse.

As for the abuse of opioids, the literature at home and abroad mainly focuses on the mechanism of addiction and the qualitative study of socio-economic factors, but there is still no corresponding data research and quantitative analysis, and the prediction analysis and research of the abuse of opioids. But for data prediction, there are many kinds of forecasting methods, including the grey system forecasting method, BP neural network forecasting method, and mathematical statistics forecasting method. At the same time, when studying the situation of drug abuse in the region, it is often affected by local social and economic factors, so BP neural network can adapt to this situation well. It is found that the prediction accuracy can be improved by using the Markov chain to correct the residual value.

Therefore, based on the work done by predecessors and considering the opioid crisis, we take solutions to analyze the drug incidents that happened in Ohio, Pennsylvania, Kentucky, Virginia, and West Virginia. Then propose strategies to deal with the crisis. At first, we use the Sequential Cluster Method to divide the trendless sequence into 6 intervals, determine the drug identification threshold level and introduce State Response Level (SRL) to reflect hazard levels. After that, we develop a Markov Model to forecast the state of future drug incidents. Then, we take the Grey Correlation Analysis to test whether there are correlations between opioid and heroin cases and socio-economic factors. Taking the factors that we select according to the correlation coefficients into account, we build the BP Neural Network and write programs to simulate it. Though BP Neural Network has a certain degree of fitting in the prediction, there is still some error in the results. Hence, Markov Chain Method is adopted to modify them. Finally, we identify strategies for countering the opioid crisis and verify the effectiveness of the strategies.

2. Methodology

2.1. Sequential cluster method

The Sequential clustering method, a kind of clustering analysis, was first proposed by Fisher in 1958 [1]. It has the characteristics of objective and optimal, which can help minimize the difference between samples within the sample segment and maximize the difference between segment and

segment. In this paper, this method is used to segment the occurrence times of synthetic opioid and heroin incidents in each county of the six states. The segmented interval is called state $E_j (j = 1, 2, 3, \dots)$. Furthermore, we also used this method to introduce State Response Level (SRL), which can reflect hazard levels of opioid events.

The clustering process is as follows:

Step1. Calculate class diameter

In Sequential cluster method, “class diameter” is used to represent the length of each interval, that is, the difference degree of each state. The smaller the diameter, the smaller the difference within the interval will be. The diameters of the classes $x_i, x_{i+1}, \dots, x_j, (1 \leq i \leq j \leq n)$ are denoted by $D(i, j)$, and the diameter is determined by the sum of squared deviations, which is: $D(i, j) = \sum_{l=i}^j (X_l - \bar{X}_{ij})^2$.

Step2. Calculate the error function

To determine how to segment the interval optimally, the error function is introduced to indicate whether the segmentation is good or bad:

$$E[P(n, k)] = \sum_{l=1}^k D(i_l, i_{l+1} - 1) \quad (2.1)$$

where:

- $E[P(n, k)]$ is defined as the sum of the diameters of the intervals. When n and k fixed, the segmentation that minimize the value of $E[P(n, k)]$ is defined as the optimal segmentation.

Step3. Determine the optimal number of sections We evaluated the segmentation effect by using β the ratio of the sum of squared deviations within the segment No further dividing is necessary when the value approaches 1. The specific formula is as follows:

$$\beta = \frac{E[P(n, k)]}{E[P(n, k + 1)]} \quad (2.2)$$

2.2. Grey correlation analysis

Model Preparation Based on the closeness of the trend of each factor, Grey correlation analysis is to quantitatively describe the close relationship between each reference sequence and the comparison sequence. As a measure of the degree of association between factors, it is very suitable for dynamic process analysis. The correlation coefficient formula:

$$\xi(k) = \frac{\min_i(\Delta_i(\min)) + \rho \max_i(\Delta_i(\max))}{|x_0(k) - x_i(k)| + \rho \max_i(\Delta_i(\max))} \quad (2.3)$$

where

- $\xi(k)$ means correlation coefficient between data from indicators k and 38 and reported opioid and heroin incidents;

- ρ is the resolution coefficient, $0 \leq \rho \leq 1$.

Among which:

$$\min_i(\Delta_i(\min)) = \min_k(\min |x_0(k) - x_i(k)|) \quad (2.4)$$

$$\max_i(\Delta_i(\max)) = \max_k(\max |x_0(k) - x_i(k)|) \quad (2.5)$$

The general expression of the degree of association is:

$$r_i = 1/N \sum_{k=1}^N \xi_i(k) \quad (2.6)$$

2.3. BP neural network prediction

The composition and working mechanism of BP network can be described by the three-layer neural network structure in Figure 1.

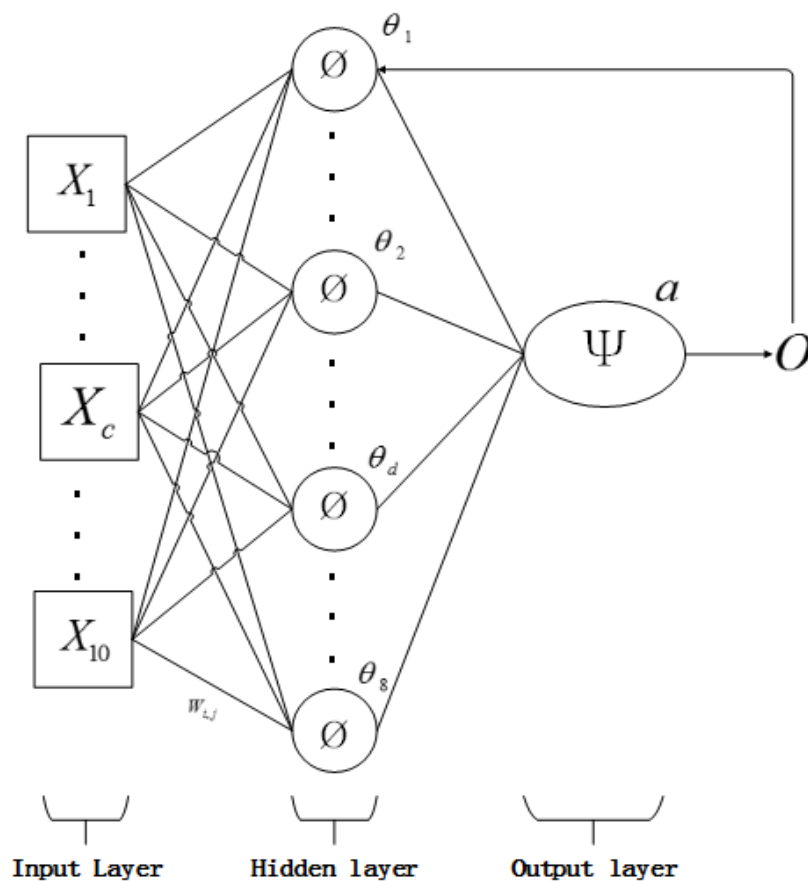


Figure 1. Structure diagram of the three-layer BP neural network.

where:

- X represents the input from the outside world, the subscript represents the serial number of the node.
- W represents the weight of each layer in the network, and the subscript represents the order of nodes.
- I is hidden layer node.
- J is the input layer node.
- K is the output layer node.
- Θ is the thresholds of the hidden layer.
- α is the thresholds of the output layer, respectively.
- ϕ is the incentive function of the hidden layer.
- ψ is the excitation function of the output layer.
- O is the output portion, the subscripts are used to indicate the order of nodes.

2.4. Markov chain

2.4.1. Determine the state space of the Markov chain

The data is serialized into different states through sequential cluster method, denoted by $E_m (m = 1, 2, 3, \dots)$.

2.4.2. State transition probability and state transition matrix

State transition probability P_{ij} :

$$P_{ij}^{(k)} = \frac{m_{ij}^{(k)}}{M_i} \quad (2.7)$$

where

- M_i represents E_i the number of occurrences each.
- $m_{ij}^{(k)}$ means from E_i to E_j requires k steps to complete.
- m is the number of states divided.

The state transition matrix is:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1m} \\ P_{21} & P_{22} & \cdots & P_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nm} \end{bmatrix} \quad (2.8)$$

2.4.3. The results of Markov revised BP network prediction

$$F = \frac{F_g}{1 - q} \quad (2.9)$$

where

- F_g represents the predicted value of the BP neural network.
- q represents the boundary value of the Markov chain state interval.

3. Empirical research

Table 1. Total amount of opioids in each state.

	Philadelphia,PA	Jefferson,KY	Hamilton,OH
2010	33513	5426	13780
2011	26969	5953	10837
2012	25279	5494	10453
2013	21761	5184	11063
2014	17597	4512	12208
2015	21387	3836	14406
2016	21194	3825	17376
2017	19614	4635	22074

Based on the NFLIS drug identification count for synthetic opioids and heroin from each county in five states between 2010 and 2017 (Table 1), and data on socio-economic factors provided by the U.S. census bureau, our research ideas are shown in the following flow chart (Figure 2):

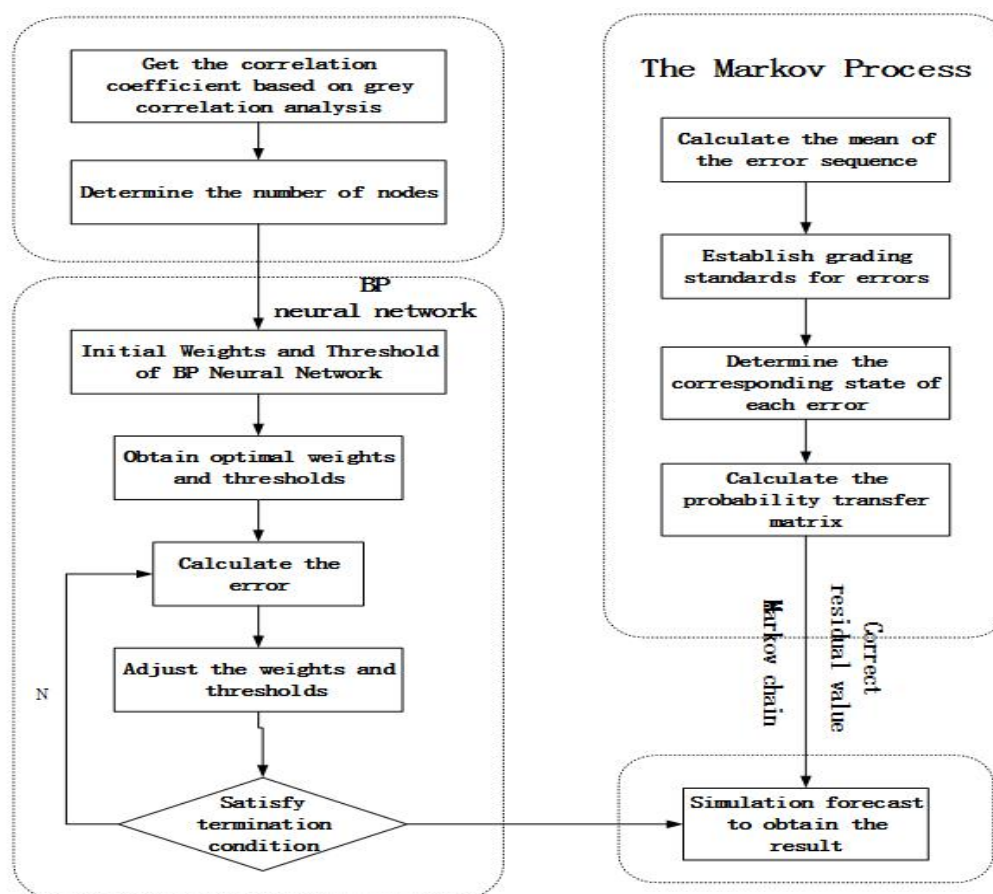


Figure 2. Model algorithm flow chart.

3.1. Grey correlation analysis

According to the sociological theory of drug abuse [2, 4], we screened 38 indicators from the 605 indicators given and used MATLAB to calculate correlation coefficients between the 38 indicators and the number of reported opioids and heroin events in Philadelphia County, Jefferson County, and Hamilton County. The top 7 indicators for each county's relevance to opioid and heroin incidents are as follows (Table 2):

Table 2. Indicators with high correlation with synthetic opioid and heroin incidents.

	Philadelphia,PA	Jefferson,KY	Hamilton,OH
I_1	0.859333	0.830618	0.836953
I_2	0.839719	0.785214	0.822447
I_3	0.835311	0.798635	0.785641
I_4	0.834103	0.77384	0.802659
I_5	0.795508	0.763527	0.816289
I_6	0.700659	0.774867	0.812202
I_7	0.781134	0.76568	0.798186

where:

- I_1 represents the proportion of low education level people.
- I_2 represents female bad marriage rate.
- I_3 represents the proportion of the elderly living alone .
- I_4 is male bad marriage rate.
- I_5 is the percentage of persons with disabilities.
- I_6 is the proportion of veteran.
- I_7 is the number of new pregnant women.

The above chart shows the results of the grey relational analysis clearly. The correlation coefficient return by MATLAB indicates that the use of opioids and heroin is somewhat related to the socio-economic data provided by the US Census. We can find that the top 7 indicators of the three counties are the same, but the order is slightly different. In Philadelphia County and Jefferson County, the top three indicators of association with opioid and heroin incidents are: The proportion of low education level people, Female Bad Marriage Rate and the Proportion of the elderly living alone. And in Hamilton County, the top three indicators of association with opioid and heroin incidents are:the proportion of low education level people, Female Bad Marriage Rate and the Percentage of persons with disability.

3.2. BP neural network prediction

According to the ranking result of the correlation coefficient, we finally choose the top 7 indicators with high correlation degrees as the input node of the BP network. The reported number of opioids and heroin events is a node of the output layer. After repeated tests, it is found that there are 8 hidden layer nodes and the network structure is 10-8-1. Then we need to determine the optimal parameter setting. The processed data samples are divided into a training set, a verification set, and a test set.

By means of cross-validation, corresponding parameters are constantly modified to minimize the error of the verification set. So we set the learning rate as 0.1, the training momentum coefficient as 0.8, the maximum convergence time as 10000, and the minimum error allowed by the training target as 0.00001. The number of training samples is 4 (from 2010 to 2014), and the number of predicted samples is 2. At last, the predicted results are shown in Tables 3, 4 and 5. The Absolute Error and Relative Error formula:

$$\text{AbsoluteError} = \text{PredictedValue} - \text{ActualValue} \quad (3.1)$$

$$\text{RelativeError} = (\text{PredictedValue} - \text{ActualValue})/\text{ActualValue} \quad (3.2)$$

Table 3. Philadelphia's predicted results.

Year	Actual Value	Predicted Value	Absolute Error	Relative Error
2015	21387	18623	-2764	-0.129237
2016	21194	25927	4733	0.223318

Table 4. Hamilton's predicted results.

Year	Actual Value	Predicted Value	Absolute Error	Relative Error
2015	14406	12490	-1916	-0.133000
2016	17376	15363.9	-2012.1	-0.115798

Table 5. Jefferson's predicted results.

Year	Actual Value	Predicted Value	Absolute Error	Relative Error
2015	4512	3894	-1916	-0.115798
2016	5184	4656.26	-527.74	-0.101802

As can be seen from the tables, the BP network has a certain degree of fitting in the prediction. However, there is still some error in the results. Then, the Markov chain method is adopted to modify the results, so as to obtain a better prediction.

3.3. Modifying of error residual by Markov chain

In order to get the result closer to the real value, we use Markov method to correct the residual error of the neural network. Since the interval of relative error in 2014 is located in the interval of E_3 , it can be predicted that the relative error of the prediction sequence is located in the interval of E_1 and E_2 respectively. Then we take Markov method to modify, and the modifying results are shown in Table 6.

Table 6. Philadelphia's relative error distribution.

Year	BP Neural Network			Markov Modification	
	Actual Value	Predicted Value	Relative Error	Corrected Value	Relative Error
2015	21387	18623	-0.129237	21061.34	-0.01523
2016	21194	25927	0.223318	21788.54	0.028052

Obviously, after using Markov correction, the prediction error is reduced from 0.03437 and 0.049446 to 0.01523 and 0.028052, and the prediction accuracy is further improved, which verifies the feasibility of this method. Similarly, the results of other states and counties can be obtained. Some of the results show as Table 7 and Table 8.

Table 7. Hamilton's relative error distribution.

Year	BP Neural Network			Markov Modification	
	Actual Value	Predicted Value	Relative Error	Corrected Value	Relative Error
2015	14406	12490	-0.133000	13974.66	-0.02994
2016	17376	15363.9	-0.115798	16181.26	-0.06876

Table 8. Jefferson's relative error distribution.

Year	BP Neural Network			Markov Modification	
	Actual Value	Predicted Value	Relative Error	Corrected Value	Relative Error
2015	4512	3894	-0.115798	4281.214	-0.05115
2016	5184	4656.261	-0.101802	4754.166	-0.08292

In conclusion, the method of using Markov to modify the neural network has great accuracy in the prediction of opioid events and the analysis of fluctuation characteristics, which can be applied to the prediction of a large number of addictive drug events.

Using the methods above, the future state of opioid abuse in each county of the five states was obtained, and the six states correspond to six colors, which were presented by the heat map as follows (Figure 3):

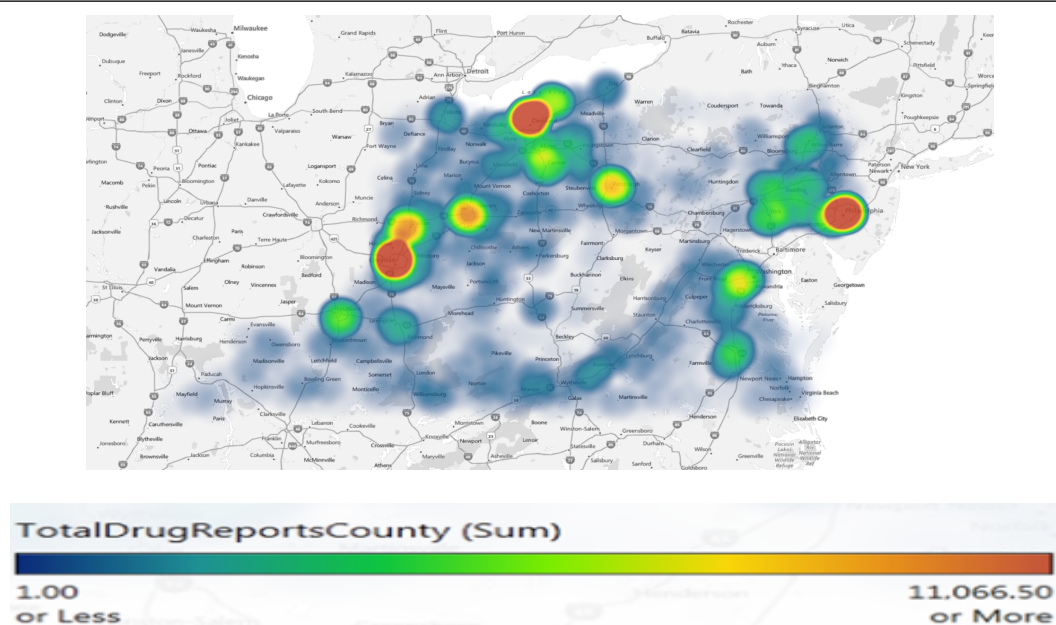


Figure 3. The heat map of the future state of opioid abuse.

Obviously, opioid abuse in Pennsylvania, Ohio and Kentucky is more serious. Therefore, in order to effectively control the situation of opioid abuse in the region, we proposed corresponding strategies and tested their effectiveness below.

3.4. Strategy and effectiveness

Based on the results given in part 2, we can find that the drug use status of each county is related to the socio-economic data provided by the US Census. For the sake of identifying strategies to address the opioids crisis, we should take measures to control indicators with a high correlation to achieve the effect of controlling drug abuse [3]. In most counties, the educational level and the proportion of elderly people living alone over 65 years of age are highly correlated with drug abuse. On the basis of the (sociology literature), policies should take up with improving the educational level of the region, strengthening the popularization of drug science, improving the level of social welfare to reduce the proportion of elderly living alone, and improving the well-being of the elderly living alone, which can help reduce the number of opioids abuse and addictions to achieving the goal of controlling the regional opioids crisis.

To test the effectiveness of this strategy, we will take Philadelphia as an example and use our model to simulate the prediction of drug identification counts after the policy takes effect from 2013 to 2017 and compare with the actual values. In the test, to effectively test the effect of the policy and eliminate the influence of interfering factors, we will adopt the control variable method, assuming that all the other indicators remain unchanged except the factors tested from 2013 to 2017.

3.4.1. Improve the educational level of the region

Raising the level of education in a region means that the proportion of low education in a region decreases. It is assumed that after the policy measures were adopted in 2012, the level of low education

in the region steadily declined year by year. Add the adjusted index data from 2013 to 2017 to the trained neural network for prediction. Figure 4 shows the effect of the annual percentage decline on Philadelphia's drug identification count: As can be seen from the figure, improving the education level in the region can help reduce the drug use and addiction in the region, thus alleviating the crisis of opioids use. However, it should be noted that when the educational level rises by 5%, the marginal utility of the policy reaches the maximum value, which can be considered as the limit of the success of the policy.

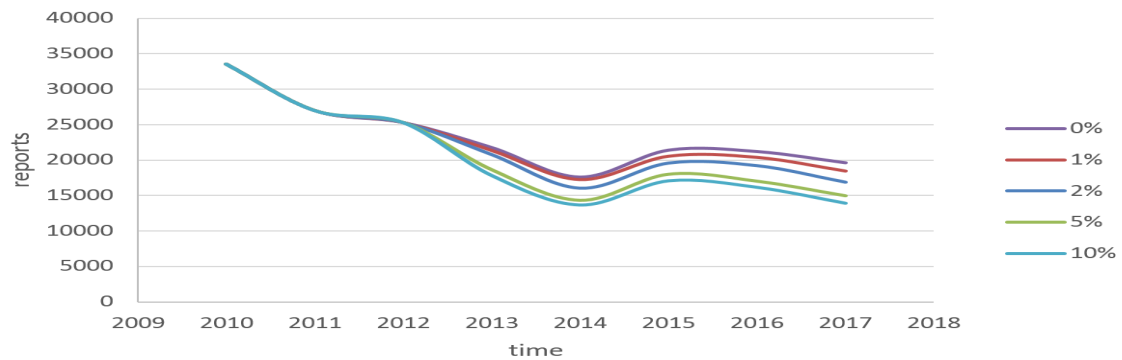


Figure 4. Sensitivity analysis of educational level.

3.4.2. Reduce the proportion of elderly people living alone

By raising the level of welfare in the region and increasing the attention to the elderly living alone, it will help reduce the excessive use of opioids by elderly living alone due to emptiness, pain and other reasons. assuming that policy measures are taken in 2012, the proportion of elderly people living alone declines, add the adjusted 2013–2017 indicator data to the trained neural network for prediction. The adjusted index data from 2013 to 2017 are added into the trained neural network for prediction. Figure 5 shows the effect of the annual percentage decline (referring to the effectiveness of the policy) on Philadelphia's drug identification count:

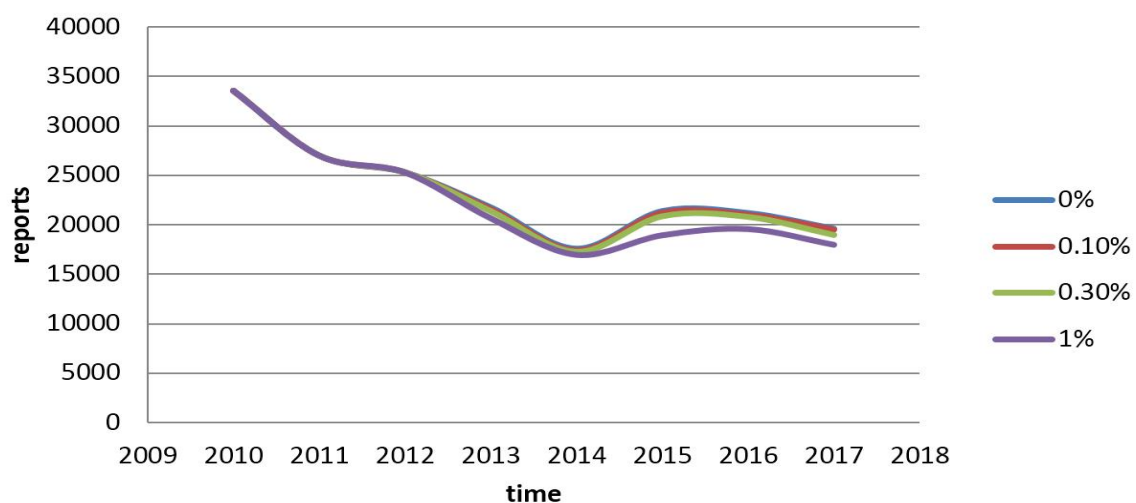


Figure 5. Sensitivity analysis of the proportion of elderly people living alone.

As can be seen from the figure, although reducing the proportion of elderly people living alone can reduce the Philadelphia's drug identification count, the effect is not deep and has a certain time lag. To have a significant impact, the annual percentage reduction must be more than 1%, which is difficult to achieve in practice. At the same time, since policies are indirectly applied to the main body through society, there is no direct impact on reducing the proportion of elderly people living alone, and the effectiveness of policies is further weakened. As a result, the policy has tested poorly to contain the opioids crisis.

4. Conclusion

This paper manages to analyze the use of synthetic and non-synthetic opioids. We use many theories and methods such as Sequential Clustering, Heat Map, Grey Correlation Analysis, Markov Chain. Our model exhibits great potential in concluding below:

- We visualize the data and visually demonstrate the temporal and spatial characteristics of the reported dose of drug events.
- We derive the relation between the socio-economic factors and the drug incidents reported from Grey Relational Analysis.
- We make predictions about future drug incident reports and point out the concerns of the US government.
- Finally, we offer measures for the factors that are highly correlated with the reported amount of drug accidents, and verified the effectiveness of the measures.

Conflict of interest

The authors declare that there is no conflicts of interest in this paper.

References

1. X. Xianhai, *Fisher ordered clustering method and its application in extraction of temperature anomalies in furnace tubes*, Guangxi Dianli, (2005), 15–17.
2. H. Dan, *A review of sociological theories on drug abuse abroad—a case study of relevant research results in the United States*, Realism, (2010), 86–89.
3. C. S. Davis, A. J. Lieberman, H. Hernandez-Delgado, et al. *Laws limiting the prescribing or dispensing of opioids for acute pain in the United States: A national systematic legal review.*, Drug Alcohol Depen., **194** (2019), 166–172.
4. W. Ke, P. Hemin, G. Jiatai, *Research on the spread of opioid abuse and its trend prediction*, Journal of Technology Wind, (2019), 244.



AIMS Press

©2019 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)