For office use only	Team Control Number	For office use only
T1 ________	**1901213**	F1 ________
T2 ________		F2 ________
T3 ________	Problem Chosen	F3 ________
T4 ________	**C**	F4 ________

2019
MCM/ICM
Summary Sheet

Take Me Home:
Preventing Journeys Down the Opioid Addiction Road

In recent years, overdose has been the leading cause of accidental deaths in the United States, and prescription opioids and heroin are among the heaviest offenders in that category. While many people need opioids to manage their chronic and severe pain, a common consequence of these treatments are abuse, addiction, and escalation to worse substances. A variety of strategies exist to combat the spread of drugs, like education, rehabilitation, and law enforcement. However, a more targeted strategy is necessary for opioids, given their new ubiquity in American society.

We have developed a model that robustly and accurately predicts the spread of opioids within and between the states of Ohio, Pennsylvania, Virginia, West Virginia, and Kentucky. In doing so, we:

- Visualized the movement and spread of opioids within the region between 2010 and 2016 by treating opioid addiction as a disease that spreads deterministically between neighbors and assuming that the spread of it can be modeled in a Markovian fashion. This will allow us to find a transition matrix that tells the influence of each county on one another. We also factored into this matrix the distances between counties.
- Next, we modeled the effect of socioeconomic factors throughout the data set, and correlated these changes with how opioid use in that county grows or shrinks over time.
- We then combined the models in two ways - in a linear and parallel fashion. We used this to estimate *epicenters* from where the drug problem emanates.
- Finally, we ran multiple simulations and predicted the drug problem well into the future to develop a number of strategies to tackle the epidemic from different perspectives, selecting from the variables that contribute most heavily to the spread of opioids.

Our model treats illicit opioid use as a disease that is spread more frequently when more people in a given area have it. This allowed us to design it such that it can be generalized to a larger region in the future. By visualizing this spread, we were able to witness predicted opioid use spread through and along major roads over longer distances than simple adjacency would predict. The counties connected in this way include both the epicenters and the vulnerable ones.

To evaluate our external model, we gave it the first two years of drug report data from all the counties, and then allowed it to propagate through the year 2016. Our predictions have an error on the order of 10^{-5} drug reports per capita. When evaluating our internal model, we realized that while socioeconomic factors are highly correlated, they could not accurately predict opioid abuse.

After this, we modified various initial conditions in our model such as socioeconomic factors and influence of the epicenters. By doing so, we were able to find effective and highly targeted strategies which will drastically reduce and reverse the prevalence of opioids in the region provided.

Contents

1 Introduction

In 2015, 33,091 people died of an overdose on heroin and other opioids, representing more than 60 percent of the total overdose deaths in the United States that year [1]. Given that overdose was also the leading cause of accidental deaths that year [1], it is clear that the dispersion of opioids is a serious problem in this country. Opioids are a class of narcotic painkillers, including the illicit drug heroin, that are derived from the poppy flower, or are synthesized artificially and have a similar structure to other opioids. Some examples of opioid prescription drugs are morphine, hyrocodone, oxycodone, dihydromorphone, and fentanyl. While almost every drug in this class can be prescribed to manage chronic pain, it is clear that the spread of opioid use has spread far beyond the reach of prescribed medicine and presented itself as an epidemic. This plague, while deadly, is highly treatable with proper targeting for law enforcement and strategies for potential victims. For this reason, it is of the utmost importance that we understand and analyze the spread of opioids and figure out how to reverse the trend. In this report, we do the same by building mathematically rigorous models. In summary, we will take the data pertaining to counties of five different states - Kentucky, West Virginia, Virginia, Ohio and Pennsylvania - and develop a predictive model to analyze how drug abuse spreads and infects the counties. Then, we will attempt to find solutions to the same problem using the model.

1.1 Problem Summary

- Visualize drug report data for each county in the states of Ohio, Kentucky, Pennsylvania, Virginia, and West Virginia for the years 2010-2016.

- Develop a predictive model that anticipates the spread of opioid use over the given time frame, and use it to find potential epicenters for the spread of opioids within those states.
- Analyze key socioeconomic factors to determine the relationship between these factors and opioid use. Employ these relationships to find counties that are demographically prone to abusing opioids.
- Determine a course of action to prevent the spread of new addictions and make current ones harder to maintain.

1.2 Our Model

- Initially we treated opioid addiction as a disease that spreads deterministically between neighbors, by assuming that the spread of it can be modeled in a Markovian fashion. This will allow us to find a transition matrix that tells the influence of each county on one another. We also factored into this matrix the distances between counties.
- Next, we modeled the effect of socioeconomic factors throughout the data set, and correlated these changes with how opioid use in that county grows or shrinks over time.
- We then combined the models in two ways – in a linear and parallel fashion. We used this to estimate *epicenters* from where the drug problem emanates.
- Finally, we ran multiple simulations and predicted the drug problem well into the future to develop a number of strategies to tackle the epidemic from different perspectives, selecting from the variables that contribute most heavily to the spread of opioids.

In our model, we analyze the general effect of socioeconomic factors as well as the drug problem of neighboring counties. Due to the general approach taken in model, the model will very likely scale well to larger regions.

The rest of this report is organized as follows: Section 2 will provide background information necessary for understanding the condition we are modeling. Section 3 will briefly discuss the provided data for this analysis. In Section 4 we will make the necessary assumptions to allow the situation to be modeled efficiently. Section 5 will encapsulate the details of our models, both external and internal. In Section 6 we solve for the parameters of our model, and discuss the results we get. In Section 7, we will modify key factors that appeared to have great influence over the model, to determine its sensitivity to the initial conditions, and give insight into which the most effective strategies for combating the drug problem. In Section 8, we will evaluate the strengths and weaknesses of our model. In Section 9 will lay out the strategies we have found will work best to reverse the drug crisis. Finally, we will conclude the report in Section 10 by summarizing our key findings. Additionally, in Section 11, we have written a brief memo to the Drug Enforcement Administration suggesting policy changes to combat the drug crisis.

2 Background

Before modeling the spread of opioids, it is important to understand the science behind addiction. When an opioid is taken, it mimics natural neurotransmitters by locking onto and preventing the re-uptake of these chemicals. This results in the brain being flooded with dopamine, the chemical the brain normally releases as part of its reward system. This can sometimes result in an unhealthy cycle of needing the drug that creates this feeling, and taking the drug at higher and higher doses due to tolerance. When a person has taken this cycle to the point where the drug becomes one of their basic needs, like food or water, and they suffer extreme adverse effects in their life as a result, they have an addiction. Due to the addiction potential of opioids and the frequency with which they are prescribed, the spread of opioid use and abuse can be rapid.

In 2012, enough prescriptions for opioid medication were written for every American adult to have their own bottle of pills [2]. It is important to understand how those around one will factor into one's likelihood to use opioids, due to how common opioids are in everyday life now. Along with one's surroundings, data supports that certain socioeconomic factors can be correlated to an increase use in opioids [3]. Given these possibilities, it is paramount to find if, and how much, these factors can predict future drug use, in order to stop the vicious cycle of addiction for so many victims.

3 Data

In this section, we will give a brief overview of the data we used to develop the models described in subsequent sections.

3.1 Given data

We were given 8 different data files. The first contained drug report counts from counties of five states in USA (Kentucky, West Virginia, Virginia, Ohio, Pennsylvania). A drug is reported in this case as part of an investigation by local police departments. The data provided gives total drug reports for each county, and any counts of drug reports for prescription opioids or heroin, courtesy of the National Forensics Laboratory Information System (NFLIS). For our purposes, we used just the opioid and heroin counts. The other 7 data files contain US Census data for these counties from years 2010-2016. These represent a common set of about 134 socioeconomic factors.

While cleaning up the data, we realized that the number of counties differed in some years. If a county was missing drug report data in the first file for any year, we set the drug reports to zero. Further, we realized that Bedford City, VA was incorporated into Bedford County, VA in 2013. So, we fixed this discrepancy by adding the data from Bedford City from years 2010-2012 into those of Bedford County. Also, since 2013, the US Census data divided a few of the larger categories. So, during the cleanup process, we had to account for those discrepancies manually.

3.2 Additional data

As we were not given any data about the distances between counties, but were allowed to get such data from outside sources, we obtained the distance between counties, as measured in

2010, from [4]. Once again, we had to remove Bedford City from this list as it is no longer independent from the county of the same name.

4 Assumptions

When designing a model, it is necessary to make assumptions to simplify the problem to an effective representation of reality i.e., the data provided, to be able to efficiently and closely predict the future. Our model is no different and requires assumptions be made about the factors that contribute to drug abuse and the social reaction to witnessing drug use. The following subsections discuss each of our assumptions in the order of their importance.

4.1 As a disease

Our model treats the opioid crisis as an epidemic and each drug report as an infected individual. This gives us the ability to interpret the given data in a more familiar context.

To see why the opioid addiction can be thought of as a disease, consider what happens during addiction. The drug overstimulates the brain's reward system and as the user builds tolerance - the overstimulated system becomes the 'new normal - a positive feedback loop is created where the user needs larger doses to get the same high. This implies that opioid addiction is not just dependent on intrinsic factors such as tolerance and drug use habits. It also depends on external factors such as peers, drug availability, legality, etc. This dependence on external and internal factors is analogous to diseases. And treating our model in such a manner gives us simple interpretations to our results.

4.2 The Markovian assumption

In order to reduce computation power, our model is based on the *Markovian assumption*. The Markovian assumption, otherwise known as the Markov property, states that the next state depends only on the current state. This is also referred to as memorylessness because the model need not remember the past to predict the future. Although this is often used to describe probabilistic models, we will use the same assumption to help build our model, which is not entirely probabilistic. Apart from reduced need for excessive computation, we have made this assumption because it has been successfully used to model population dynamics [5,6] and epidemiology [7,8] in the past.

To be precise, we will use the second order Markov property that says that the future depends on the current state and the previous state i.e., state at time $t + 1$ depends on the states at times $t, t - 1$.

4.3 Proximity to neighboring counties

The third assumption is that distance to other counties plays an important role in each county's opioid problem. In other words, counties that are farther will have a different, not necessarily small, effect than counties that are closer. Although we have no way of estimating this relation with distance *a priori*, we can intuitively hypothesize that it will be an inverse relation.

This assumption also follows from the analysis of opioid crisis as an epidemic - a person who is 100 miles away from me and suffering from common cold has a different effect from a person who is in my town with the same condition.

4.4 All opioids are the same

Another assumption made in our model is that all opioids are approximately the same, having the same health risks and effects. This allows us to analyze the total reports of opioids in each county, rather than analyze every opioid individually. While this assumption is not entirely realistic, we justify this assumption because the counts of drug reports, sorted by individual opioids, are very sparse. The sparseness of the data, when analyzed by individual drugs, could severely skew our model. This can also be justified as, the effect of all opioids are very similar, just different in magnitude.

A consequence of this is that all opioid addictions are also cured in same manner. This allows us to explore intervention strategies using a single model.

4.5 Data homogeneity across counties

Finally, we assumed that data from each county is collected with the same accuracy across each county. For instance, we assume that the number of opioid reports in a county reflect the true use of opioids in that county, rather than the diligence, or lack thereof, to report opioid use.

While this might not be realistic, we assume homogeneity because we have no means of verifying and correcting our data to reflect any inhomogeneity. Further, this assumption also simplifies our analysis and ensures that the data provided can be represented by our model.

5 Model

Nomenclature

$\mathbf{n_t}$ = *Health* of all counties in year t sorted by FIPS code
$\mathbf{x_t}$ = *Opioid problem* of all counties in year t sorted by FIPS code
A = Adjacency matrix
Q = Transition matrix
δ = Threshold for opioid reports per capita to declare a county as *infected*
S_t = 30 highly correlated socioeconomic factors of all counties in year t sorted by FIPS code
$\beta_\mathbf{1}$ = Parameters of the linear regression model using S_t
$\beta_\mathbf{2}$ = Parameters of logisitc regression model using S_t

As hinted in Section 4, we will model the opioid crisis as a disease. Treating this as a disease helps us decouple the larger problem into two smaller problems – modeling the effect of external factors, and modeling the effect of internal factors. This can be thought of as follows: the future health of an individual depends on the current health of their neighbors and the current health of the individual. Equation 1 shows this relationship.

$$\underbrace{\quad}_{\substack{\text{Future}\\\text{opioid problem}}} = \underbrace{\quad}_{\substack{\text{Surrounding}\\\text{opioid problem}}} + \underbrace{\quad}_{\substack{\text{Internal}\\\text{well being}}} \tag{1}$$

First we will develop a model that predicts health based on the health of all counties. Then, we will create a model that factors in the internal factors i.e., socioeconomic factors for each county. Finally, we will combine these together to form a holistic model.

5.1 Defining *health*

So far, we have loosely used the term *health* to refer to the amount of drug crisis in each county. Here, we will formally define and quantify health. Using the raw number of drug reports in each county can be misleading because the population of counties are different. So, instead, we normalize the number of drug reports every year by the population of that county. This gives us a better representation of the *health* of the counties.

We will use $n_t^{(k)}$ to denote this health measure of country k in year t, and $\mathbf{n_t}$ to denote the health of all counties sorted by the FIPS code in year t. If we have N counties, $\mathbf{n_t} \in \mathbb{R}^{N\times 1}$.

5.2 Modeling external factors

Our model for predicting the drug crisis using external factors takes into account the opioid problem in all counties and the distances between counties. Our model for external factors is summarized in Equation 2.

$$\mathbf{x_{t+1}} = QA\mathbf{x_t} \tag{2}$$

Now, we will look at what each of these terms represent.

5.2.1 Understanding $\mathbf{x_t}$

$\mathbf{x_t}$ to represent the *opioid problem* of each county in year t. We use $\mathbf{x_t}$ to encapsulate the current health and the trend in health from last year. This makes sense because the current opioid use *and* the trend of that county's opioid use will have an effect on the neighboring counties. We calculate $\mathbf{x_t}$ as:

$$\begin{aligned}\mathbf{x_t} &= \mathbf{n_t} + (\mathbf{n_t} - \mathbf{n_{t-1}})\\ &= 2\mathbf{n_t} - \mathbf{n_{t-1}}\end{aligned}$$

Thus, $\mathbf{x_t} \in \mathbb{R}^{N\times 1}$. This is also where our second order Markovian assumption comes into play. Going back to Equation 2, notice how $\mathbf{x_t}$ predicts $\mathbf{x_{t+1}}$ i.e., the *opioid problem* of counties in year $t+1$ depends on the health of counties in years t and $t-1$. Thus, our model is temporally aware of itself.

5.2.2 Understanding A

A refers to a modified adjacency matrix. This accounts for the distances between counties. Each element in $A \in \mathbb{R}^{N \times N}$ is given by:

$$a_{i,j} = \begin{cases} \frac{1}{\text{distance between county } i \text{ and } j} & , i \neq j \\ 1 & , i = j \end{cases}$$

Thus, all counties have values less than 1 in A, and farther counties will have a smaller value. To understand how this works, let us look at each element of the vector $A\mathbf{x_t}$:

$$(A\mathbf{x_t})_i = a_{i,1}x_t^{(1)} + a_{i,2}x_t^{(2)} + \cdots + a_{i,N}x_t^{(N)}$$

As we can see, the opioid problem from a county that is at larger distance from country i contributes a smaller value in the vector $A\mathbf{x_t}$. Thus, the i-th element of $A\mathbf{x_t}$ encodes the information about both, proximity between counties *and*, the opioid problem across counties. Thus, our model is spatially aware of itself.

5.2.3 Understanding Q

Q is the transition matrix that propagates the opioid problem at year t into the next year. $Q \in \mathbb{R}^{N \times N}$ propagates the model by understanding how the opioid problem in each county affects every other county including itself. Thus, we can use the elements of Q to determine which counties have the largest influences on other counties.

$$Q_{i,j} = \text{how much does county } j \text{ affect county } i \text{ independent of distance}$$

After normalizing the rows, we can use the column sums to estimate the most influential counties in the opioid crisis.

5.2.4 Estimating Q

At this point, we have $\mathbf{x_t}$ and A from data. We need to estimate Q, the transition matrix. Notice that we need not find a Q that works exactly for all t. In fact, depending on our data, such a Q might not exist at all. Even if we do find such a Q, it might not generalize well for examples outside the dataset. Instead, we want to find a Q that *approximately* works best for all t. We can do this by solving a modified least squares estimation problem:

$$Q = \min_{Q'} \sum_t \|\mathbf{x_{t+1}} - Q' A\mathbf{x_t}\| \tag{3}$$

Here, the summation of the norms is over all t in our dataset.

Thus, we have a model accounting for external factors that is spatiotemporally aware of the opioid problem. Now, we can focus on constructing a model that is based on the internal factors.

5.3 Modeling internal factors

Going back to the analogy of disease, we can argue that the opioid problem in a county is also dependent on the internal properties of the county i.e., the socioeconomic factors of the county. So, we can develop a model that takes these into account and see the effect of socioeconomic factors on the opioid crisis.

The data set given to us has approximately 150 socioeconomic factors. Instead of building a model that uses all the factors, we can instead select the 30 highly correlated factors and use them. This simple pre-processing step has two advantages – it will simplify our model and, make our model more interpretable. To measure correlation, we will use the Pearson correlation coefficient [9].

5.3.1 Linear regression

We can use these 30 features to build a linear model that would predict the opioid problem in year t with socioeconomic factors of each county in year t. Let us denote the matrix containing the top 30 socioeconomic factors of all counties in year t as S_t. Each row of S_t corresponds to a single county and each column corresponds to a single socioeconomic factor. Then, our problem is that of estimating parameters $\beta_{\mathbf{1}}$ such that:

$$S_t\beta_{\mathbf{1}} = \mathbf{x_t} \tag{4}$$

This is once again a least squares estimation problem to find the best $\beta_{\mathbf{1}}$

5.3.2 Logistic regression

If we define a threshold of drug reports per capita, δ, for a county to be *infected* with the opioid problem, we can also use the socioeconomic factors, S_t to predict whether a county is infected or not. We can do this by constructing a logistic regression model that gives the probability that a county is infected [10]. The probability that county i is infected in year t is given by:

$$p_t^{(i)} = \frac{1}{1 + e^{-S_t^{(i)}\beta_{\mathbf{2}}}} \tag{5}$$

where $\beta_{\mathbf{2}}$ are the parameters of the model that brings $p_t^{(i)} \to 1$ for infected counties and $p_t^{(i)} \to 0$ for non-infected states.

5.4 Combining the two models

So far, we have decoupled the problem into external and internal, and we have developed models that solve both problems. Now, we are ready to combine the models to complete our model development phase. Here, we propose two different ways to combine the models.

5.4.1 A convex combination

We can combine the external propagation model in Equation 2 and the internal linear regression model in Equation 4 in a convex fashion to predict the opioid problem in year t as:

$$\mathbf{x_{t+1}} = \theta \cdot QA\mathbf{x_t} + (1-\theta) \cdot S_{t+1}\beta_{\mathbf{1}} \tag{6}$$

where $\theta \in [0,1]$ is a parameter that controls how much each model contributes to the final prediction. We can choose a θ that minimizes the error on our prediction.

5.4.2 A two stage model

Instead of directly combining the results from both models, we can use the results independently to determine if a county is infected or not. Using the external propagation model from Equation 2 and applying the threshold δ, we can identify infected counties. Then, independently, we can use the internal logistic regression model from Equation 5 to identify infected counties. Then, from both results, we can infer potential risk for infection for counties.

A county identified as infected by both models is labeled high risk, while a county identified by only one model is labelled medium risk, and a county identified by neither models is labelled low risk. Note that the infection threshold, δ, is kept constant over the time span of both models.

6 Solutions

We estimated parameters to the models developed in Section 5 described in Equations 2,4,5. In this section, we will discuss the results we get from our models.

6.1 Estimating δ

To fix the threshold to declare a county as infected, we took the first year in our dataset i.e., 2010 and calculated the threshold to be one standard deviation greater than the mean drug reports per capita. This allows us to pick the top 16% of counties that are most infected, assuming the the drug reports per capita to be normally distributed.

$$\delta = \text{mean}(\mathbf{n_{2010}}) + \text{std}(\mathbf{n_{2010}})$$

This yields, $\delta \approx 0.0038$ as the threshold to declare a county as infected. Figure 1 shows the spread of infected counties from year 2010 to 2016 using this threshold. Alternatively, in Appendix C, Figure 3 shows the spread of opioid crisis from year 2010 to 2016 in terms of the actual number of drug reports per 100 people.

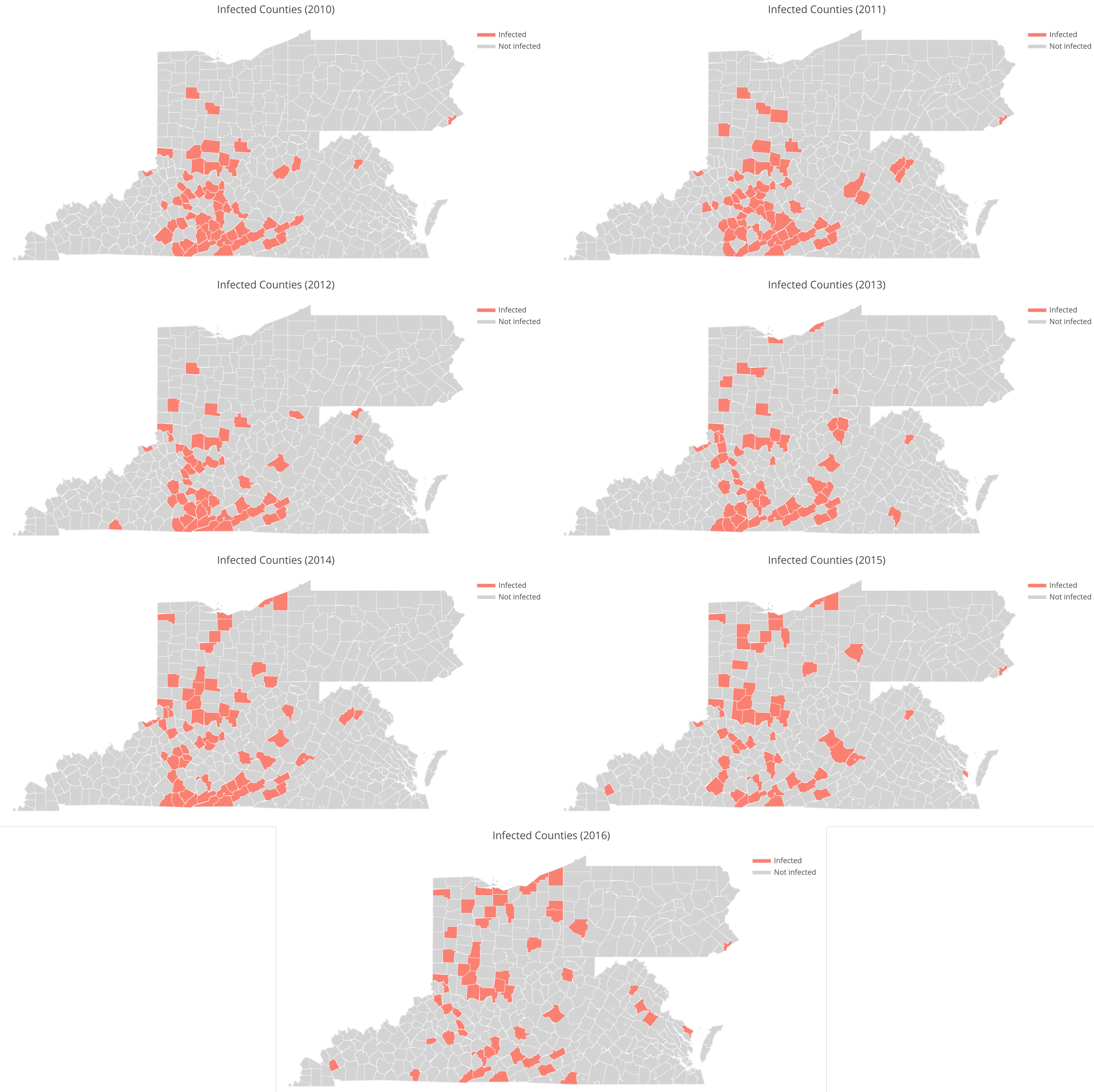

Figure 1: *Spread of the opioid crisis from year 2010 to 2016. The threshold for infection is* $\delta = 0.0038$ *drug reports per capita. Counties colored red are infected.*

6.2 Estimating the influence of external factors

We estimated Q with the help of Equation 3 using data from years 2010-2015. Then, we tried to predict the number of drug reports per capita in year 2016, $\mathbf{n_{2016}}$ using this Q.[1] Our prediction has a mean squared error of 2×10^{-5} drug reports per capita across all counties.

Such a small error might be because we have overfit our model. So, instead, we also estimated Q using data from years 2010, 2011. Then, we predicted the number of drug reports for years 2012-2016. Here, we observed a mean squared error of 8.42×10^{-5} for year 2016. This shows that our model for external factors does a good job for predicting well into the future.

Then, we estimated Q from all the data given to us i.e., from years 2010-2016. Now, we propagated the model forward in time and found that, all else constant, by year 2025, all counties are *cured*. Further, by year 2038, the drug reports per capita reaches a steady state. In 2038, $\text{mean}(\mathbf{n_{2038}}) = 1.8 \times 10^{-3}$ and $\text{std}(\mathbf{n_{2038}}) = 0.002$. Further, $\text{mean}(\mathbf{n_{2010}}) \approx 1.8 \times 10^{-3}$. This shows that as we propagate forward in time, if there are no other changes, the drug reports per capita from all states converge to the mean of the initial data.

6.3 Estimating the influence of internal factors

We were given 149 socioeconomic factors for years 2010-2012. Some of these factors were subdivided into multiple categories in subsequent years, so the number of socioeconomic factors increased to 152 in years 2013-2016. We analyzed the correlation between each socioeconomic factor and the *health* of counties for all years using the Pearson correlation coefficient [9]. Using this, we picked the top 30 highly correlated (including both positive and negative correlations) for our analysis. This simplified our model as we discarded socioeconomic factors with low magnitude correlations. The list of socioeconomic factors we chose are listed in Appendix B.

Using just these socioeconomic factors, we performed regression analysis to determine if it is practical to use a linear model described by Equation 4 in Section 5.3.1 to predict the *opioid problem*. Our regression coefficient was, $R^2 = 0.1296$. This suggests that we cannot use just the socioeconomic factors to accurately predict the *opioid problem*. In fact, after training the model on data from years 2010-2015 and attempting to predict the *health* of counties for year 2016 gives a mean squared error of 1.949×10^{-3} drug reports per capita, which is much larger than the error we got in the external model. This is not surprising because it turns out that human behavior, which is the crux of the opioid problem, is very hard to predict on socioeconomic features [11].

However, when we built the logistic model described by Equation 5 in Section 5.3.2, we got a mean accuracy of 87.98 % accuracy i.e., the logistic model seems to do better. A closer look at the results says that our true positive rate is 3.57% and our true negative rate is 99.51%. This suggests that our classifier is skewed. Taking a look at the data, we observed that only 12 % of our original data contains positive examples i.e., infected counties. This was by design when we chose the threshold, δ. Thus our classifier suffers from a class imbalance problem that skews the classifier towards predicting negative more often. We argue that we

[1] If we have $\mathbf{x_t}, \mathbf{n_{t-1}}$, we can find $\mathbf{n_t} = 0.5 \cdot (\mathbf{x_t} + \mathbf{n_{t-1}})$

care more about capturing all the infected states, even if there are some false positives, as opposed to choosing to ignore infected states. In other words, it is better to diagnose an infection and have the infection not be there than to claim a group is in good health when it is not. In that case, setting the *probability threshold* for declaring a county to be infected to 0.1 captures 70 % of all infected counties.

These analyses suggest that using socioeconomic factors **alone** to predict the *health* is neither accurate nor useful even though there is a lot of correlation. So to make this model work we will have to combine it with the external model.

6.4 Building a convex combination of models

We combined the external and internal models estimated from Sections 6.2, 6.3 to develop the convex combination of the two models as described by Equation 6 in Section 5.4.1. We varied θ from 0 to 1 in steps of 0.001 and found that we get the best prediction when $\theta = 1$ i.e., we only consider the external model developed in Section 6.2.

There are two explanations to this: The first is that the internal factors alone are not an accurate indicator of county *health*. Although unlikely, our results from Section 6.3 also suggest the same. The second, more plausible, conclusion is that the socioeconomic factors are already encoded into the *health* of each county. This explains why there is high correlation, but not enough evidence to make accurate predictions.

Since the convex combination of the two models is effectively just the external model, the future predictions from this will be identical to the predictions described in Section 6.2.

6.5 Building the two stage model

Next, we built the two stage model described in Section 5.4.2. Here, we analyzed the spread of the opioid crisis by comparing the infected states predicted, independently, by the external model with internal logistic regression model. A county predicted as infected by both models is determined to be high risk; predicted as infected by only one model is low risk and; predicted by neither is not at risk.

Since we do not have socioeconomic data after year 2016, it will not be possible for us to make predictions past this using the two stage model. Instead, to measure the performance of this model, we trained our model on years 2010-2011 and made predictions for years 2012-2016. The results are summarized in Table 1. Our models suggest that the number of high risk counties decreases as we move forward in time while the number of low risk counties increases. This means that the opioid crisis somehow *stabilizes* across all counties with the infection decreasing in counties with a high opioid problem and increasing in counties with low opioid problem. This is similar to the results we got in Section 6.2 and suggests that the external model once again dominates over the internal model.

6.6 Epicenters and vulnerabilities

We identified two different classes of counties that are important in the opioid crisis - epicenters and vulnerabilities. We define an *epicenter* as an infected county from where the infection

Year	Number of high risk counties	Number of low risk counties
2012	6	60
2013	3	66
2014	0	70
2015	0	70
2016	0	76

Table 1: *Predictions from two stage model*

spreads outward to surrounding counties in subsequent years and around which there are a lot of other infected counties. And, we define a county to be *vulnerable* if its socioeconomic (internal) factors give enough evidence to suggest that it will become infected in the near future, due to low internal immunity to infection.

6.6.1 Identifying epicenters

From our definition of epicenter, it makes sense to use the external model to determine the epicenters. To do this, we need to take into account both how much influence a county has on other counties and how far away the affected counties are. This information is encoded in the transition matrix Q, and the adjacency matrix A respectively. To combine these, we performed the elementwise matrix multiplication operation, $\circ$,[2] on Q and A. So, each element $(Q \circ A)_{i,j}$ tells us the influence of county j on county i after factoring in the distance. Then, we sum the columns to tell the overall influence of all counties. Through this method, we chose the 10 counties that had the greatest influence as epicenters.

We predicted that the following counties, in no particular order, are the epicenters:[3]

- Doddridge, Harrison, Marion, and Morgan Counties in West Virginia.
- Bracken, Carroll, Harlan, Gallatin, and Campbell Counties in Kentucky.
- Mecklenburg County in Virgina

Figure 1 shows how the infection spreads from year 2010 to 2016. Here, the threshold for infection is $\delta = 0.0038$ drug reports per capita. Figure 2 shows the epicenters identified by our model. From both these figures, we can see that our model was able to accurately identify all the counties from where the infection spreads outward. This gives us evidence that our models are working the way we intended them to.

The epicenters in Figure 2 share a common characteristic – they are centrally located with respect to other infected counties, which means the average location influence would be fairly high. As an outlier, we found that Mecklenburg County, Virginia is an epicenter. We can see that this county does not have a great influence on surrounding counties. Our model does not distinguish between incoming and outgoing drug reports per capita. So, a plausible explanation is that this county is being infected by neighboring counties outside our data set.

[2]This is often described as the Hadamard product
[3]The ranking for these counties by strength of influence can be found in table 4 in Appendix A

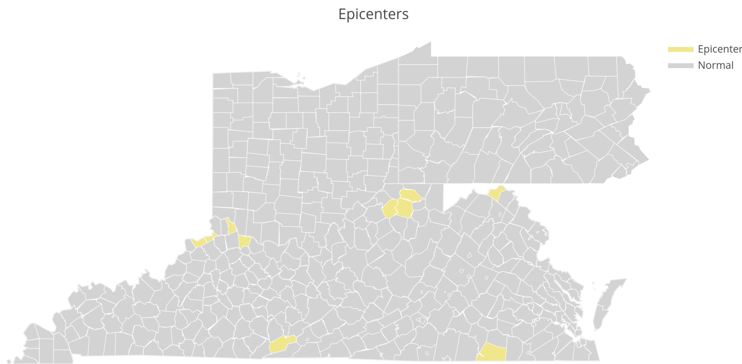

Figure 2: *Epicenters of the opioid crisis as identified by our model. The epicenters are colored yellow.*

6.6.2 Identifying vulnerable counties

By definition, vulnerable counties are determined by the socioeconomic factors. So, we used the logistic regression model and chose 10 counties whose average probability of being infected were the highest across years 2010-2016. These counties, in no particular order, are as follows:[4]

- Scioto, Montgomery, and Hamilton Counties in Ohio
- Philadelphia, Pennsylvania
- Kanawha County, West Virginia
- Pulaski, Pike, Laurel, Bell, and Jefferson Counties in Kentucky

We also predicted 10 counties that would be vulnerable in 2017, if their demographic composition stayed constant. To see these counties, consult Table 3 in Appendix A

6.6.3 Understanding these counties

Although the epicenters in Figure 2 appear to be isolated, these counties are adjacent to four major roadways – (1) I-75 that cuts across Harlan, Kentucky through Bracken, Campbell and Carroll, Kentucky to Hamilton, Ohio and further north; (2) I-64 that runs west to east across Kentucky, West Virginia and Virginia; (3) I-79 from Marion, West Virginia northwards into Pennsylvania and; (4) U.S. Route 23 from the bordering counties of Kentucky, West Virginia, and Virginia northwards into Ohio.

Based on these results, it is clear that our models were able to inherently learn the means of transit of opioid drug reports across major roads within the region. We hypothesize that if neighboring states were included in the analysis, such as Maryland or Delaware, other major roads would connect the remaining vulnerable counties and epicenters. So while many vulnerable counties are not currently infected, identifying if they fall along these major roadways could help us in mitigation efforts.

[4]The ranking of these counties can be found in Table 2 in Appendix A

7 Sensitivity Analysis

In this section, we will test our model under different initial conditions to see their effects. First, we will vary the external factors and then the internal factors.

7.1 Tweaking the external factors

First, we set the *health* of all identified epicenters to 0 at year 2010. This is analogous to purging the opioid crisis at the root. Then, we propagated the model to 2016. At the end, we observed that only 7 counties were infected as opposed to 56 in the original data.

Then, we implemented a version of border control where we set the distance to and from the epicenters in the adjacency matrix, A, to 0. This is effectively preventing flow of drugs through these counties. Once again, at the end of the simulation, we observed that only 8 counties were infected, 7 of which showed up in the previous simulation. We also implemented another version of border control where we set the entries of Q matrix corresponding to the epicenters to 0. The results were identical to the previous border control experiment.

7.2 Tweaking the internal factors

We changed some of the highly correlated socioeconomic factors to see if they had any any effect on our logistic regression model. For the following simulations, the probability threshold for declaring a county as infected was set to 0.20, and data from year 2016 was used.

We decreased the fraction of children under the age of 18 living with their grandparents by half. This decreased the number of infected counties, as predicted by this model, from 56 to 6. We gathered this trend to be coming from one of two possibilities. The first is the fact that many seniors are in chronic pain as a result of age or old injuries, and thus are more likely to be prescribed opioids and become addicted. The other is that the children are stealing opioids from those grandparents and abusing them recreationally.

We then lowered the fraction of people who were never married by half. This increased the number of infected counties to 142. Correspondingly, decreasing the fraction of people currently married by half decreases the number of infected counties to 29. This trend is interesting and further research in the field of psychology in the future can help us explain this.

More interestingly, decreasing the number of veterans above the age of 18 decreases the number of infected states to 0. This trend is likely because veterans are prone to be prescribed opioids to endure the traumas of war and their injuries as a result, and simply were never able to stop upon returning home.

It also turns out that decreasing the fraction of people with a Bachelor's degree increases the number of infected states to 105. This trend is likely because people with an education are more likely to know the adverse effects of prolonged opioid use and are therefore cautious with their usage. They are also less likely to be in a career that requires high amounts of manual labor, and therefore would necessitate an opioid prescription at some time due to work.

Most of these trends can be explained intuitively. This suggests that both our models work in the manner we expect them to and are sensitive to parameters and initial conditions in the way we want them to be. This gives further evidence that our models will generalize well to data outside the data we used to develop the model. In addition, this gives us clear ideas for groups to target with new policy and possibly assistance.

8 Strengths and Weaknesses

In this section, we will discuss the strengths and weaknesses of the model we have developed.

8.1 Strengths

- Our model is nearly memoryless as it uses the Markovian assumption. This allows our model to accurately represent data without the tradeoff of computation power. In our model, we are very capable of approximating the drug use in the future based solely off of the present opioid use and the trend from the previous year. Thus our model does not need to use data that is older than two years to make new predictions.
- Our model is incredibly accurate in prediction. Training our model using data from just two years, our model was able to predict the opioid use in each county for the following five years with a mean squared error of 8.42×10^{-5} opioid reports per capita.
- Our model represents the difference in influence between counties. This is an important factor given that some counties will have greater influence on other counties, due to resources, location, and restrictions. This allows our model to be better at predicting future trends.
- Our model also gives us interpretable values for each county's influence on other counties. Our model thus identifies the most influential counties giving law enforcers tangible targets.
- Finally, our model is able to generalize to all areas and problems. We made no assumptions about the regions we are dealing with when we developed the model – all the intrinsic factors of the regions are just parameters. We also built our model representing the opioid crisis as an epidemic. Thus, we can use the same model and apply it to different regions across the world and to other problems that can be thought of as an epidemic.

8.2 Weaknesses

- In our model, epicenter calculation does not distinguish between outgoing and incoming opioid use. This means that influence of each county, in which the epicenters are determined based off, counts opioids that flow into the state in the same manner as opioids that flow out of the state. This means that our epicenter calculations do not solely predict counties that have the most influence on the outwards spread of opioids, but the total spread of opioids.

- While our model incorporates location, the finiteness of the data set does not allow the model to factor in a county's location to counties outside of the data set. Due to this, our model assigns added influence to edge counties in the data set. However, this problem could be solved very easily by adjusting the adjacency matrix, A, to reflect this.
- Our model sets a threshold for an infected county. The threshold set by our model to consider a county infected, is any county whose opioid reports per capita is one standard deviation above the mean. A weakness of this design is that values that do not fall above this threshold, even slightly are not considered infected and therefore are not priority target areas.
- Along the same lines, our model does not establish a clear ranking system of counties to target first. While our model cares about the overall drug reports per capita in each county, we do not take into account the difficulties of law enforcement in different counties. This is an important factor when looking from the policy makers' perspectives.
- According to our model, the socioeconomic factors are not helpful in making predictions alone even though the model proves that some specific socioeconomic factors directly affect opioid use when altered. Even with this knowledge, when the socioeconomic model is combined with the external model(through a convex combination), it is seen that the most effective model is a model in which the socioeconomic factors do not contribute. This is most likely due to the fact that the socioeconomic model is based on modelling the actions of individual humans, which is largely unpredictable through statistics.
- Finally, our model does not take into account the feedback loop between opioid crisis and the socioeconomic factors. While our model accounts for the effect of socioeconomic factors on opioid use, it does not factor in how opioid use will cause changes in socioeconomic factors. Thus we are ignoring the effects of this feedback loop, without knowing how strong it is. Studying this feedback loop could help answer why our socioeconomic factors fail in making accurate predictions when used alone.

9 Policy Strategies

We identified the top 30 socioeconomic factors that correlate to opioid use. We can target policy changes to affect some of these socioeconomic factors in order to indirectly tackle the opioid crisis. We can also use our knowledge about the epicenters to decrease the spread of opioid use from these counties. In this section, we discuss a few of these strategies to mitigate the opioid crisis.

Educating vulnerable groups: The most important strategy would be to educate the groups that fit into the socioeconomic factors that are deemed most vulnerable to spread of opioid use. One of the largest socioeconomic factors that play into opioid abuse is households with both children under 18 and adults over 65. This is likely attributed to teenagers using opioids prescribed to the older family members, and seniors in general having to be prescribed opioids more frequently. One example would be advertising on media frequented by seniors to tell them to lock up drugs from adolescents, and informing them of the dangers of prolonged opioid use and addiction. We also found that decreasing the number of residents with college degrees intensified the opioid crisis. Another effort that would mimic a positive change in this regard is to educate young people through school programs about the specific health effects of opioid use, and the dangers of addiction and overdose.

Educating health care providers: Another important correlation with opioid use is the fraction of disabled residents and veterans in a county, which is probably due to increased rates of chronic pain in these groups. A study shows that patients receiving 200 MME (mg Morphine equivalent) per day showed a nearly 3-fold increase in opioid-related mortality [12]. One potential solution would be educating health care providers to prescribe alternative treatments of chronic pain, such as physical therapy or nonnarcotic drugs. Reduction within these two groups resulted in almost complete elimination of the opioid problem in our simulations.

Mental health care for people experiencing high life uncertainty: Our model also predicts that those experiencing high uncertainty in their life, like enduring divorce or housing uncertainty, are more likely to abuse opioids [13]. This can be attributed to those experiencing uncertainty being vulnerable to impulsive decisions as a solution to their troubles. A possible solution for these factors would be to encourage mental health care for those experiencing high life uncertainty, and make it more accessible to groups who would have difficulty affording it.

Law enforcement: One way in which to target the overall flow of opioids is to reduce the access to opioids in the counties that our model deems epicenters. These counties are those that have the largest impact on the spread of opioid use across the data set. Allocating opioid preventative resources towards these counties would allow law enforcement to target suppliers who increase opioid use in their own counties, and prevent the supplies from these counties to spread to other populations. Alternatively, more closely monitoring potential drug traffickers through checkpoints at state borders would greatly reduce the threat of those opioids traveling to vulnerable counties along the four major roadways mentioned in Section 6.6.3.

10 Conclusion

The use and abuse of opioids in the United States has become a major problem in the past two decades. Between 1999 and 2014, over 165,000 people died from prescription opioid overdoses alone [2]. The deaths in that time span most likely could have been prevented, with proper strategies employed. We examined the drug reporting data in five states to develop a model which would inform us as to which specific strategies would be most effective in preventing future prescription opioid and heroin overdoses, and other medical complications arising from those addictions.

We developed our model based on the possible external and internal factors present within the counties and between them. Using the external model, we found that the epicenters for the spread of opioids are the following counties: West Virginia: Doddridge, Harrison, Marion, Morgan; Kentucky: Bracken, Carroll, Harlan, Gallatin, Campbell; Virginia Mecklenburg .

These are therefore the most important counties in terms of law enforcement efforts and reduction of illicit supply. In addition, we predicted the counties that were most probable in becoming infected in subsequent years, based on their socioeconomic profile and the correlations between those demographics and opioid use. These counties were: Ohio: Scioto, Montgomery, Hamilton; Pennsylvania: Philadelphia; West Virginia: Kanawha; Kentucky: Pulaski, Pike, Laurel, Bell, Jefferson.

With all of these counties on one map, we first believed them to be separate regions that each spread opioids fairly independently. Upon further analysis, they were connected by four important highways in the region, which confirms the influence values that we predicted. We also witnessed many infected counties moving up and down these roads as time progressed within our simulation. This gave us insight as to which policies with regard to law enforcement to employ to best reduce the flow of opioids around the data.

With the external model, our predicted drug reports per capita differed from the data on average by 8.42×10^{-5}. This means that this model was extremely robust in knowing where opioids will flow and where they originate. While designing the internal model, we found that there was a high correlation between socioeconomic factors and the drug abuse problem, but the socioeconomic factors alone could not accurately predict the drug abuse problem. Further, when we combined the external and internal models, the external models always dominated. Therefore, we concluded that the socioeconomic factors are important, but are already encoded in the drug reports per capita of each county.

By combining the two models and running simulations under various conditions, we were able to identify specific strategies to combat the drug problem in these five states.

For future development, more data on hospital reports in addition to police reports, can give a better picture of how many people have a problem with opioids. Similar data from more neighboring states can help us understand the problem in a broader context, and possibly connect the isolated epicenter points to show their influence more effectively. More data on people traveling between these counties can provide additional insights into how the drug problem in each county affects every other county. Overall, however, this model very closely matches the data provided, and we hope that the strategies can be reasonably employed to help those who desperately need it.

11 Policy Letter Regarding Opioid Use in the United States

For the attention of the Chief Administrator of the DEA/NFLIS

Dear Chief Administrator,

One of the most pressing problems in the United States is the spread of opioid use across the nation. Given the severity of opioid addiction and its strong link to overdoses, it is imperative that a plan be developed to stop the rapid dispersion of opioids across the U.S. In an effort to reverse the spread of opioid use, we have developed a predictive model to extract the target areas and demographics to address the opioid crisis across five neighboring states.

Our model identified the most influential counties across the states of Kentucky, Virginia, West Virginia, Ohio, and Pennsylvania in the spread of opioid use. We identified these areas as epicenters from where the problem emanates. These are Doddridge, Harrison, Marion, Morgan, Bracken, Carroll, Harlan, Gallatin, Campbell and Mecklenburg. Our model also determined the 30 most correlated socioeconomic factors in opioid abuse. We believe that policies targeting these factors can indirectly affect opioid crisis.

With our model, we tested the effects of limiting the influence of the identified epicenters and the results were notably positive. Along with targeting these, we believe that resources should be directed to target features of the socioeconomic profile that are highly correlated with opioid use. Below, we have summarized a few strategies that we think will decrease opioid abuse by targeting the epicenters and the socioeconomic profile of these five states.

Firstly, we suggest directing resources to drug abuse prevention, education programs, and increasing rehabilitation services at the epicenters. We found that, when the prevalence of opioids were cut severely in 2010 in the 10 counties identified as epicenters, nearly 90 percent of opioid reports disappeared in 6 years. This demonstrates the effectiveness of eliminating opioid use in these epicenters. We realize that it is difficult to reduce the opioid problem in any county to zero instantly, but we believe that targeting these counties will have a ripple effect on the entire region. This will allow for the problem to be combated considerably more efficiently than before.

Secondly, we suggest increasing border control near these epicenters. When we eliminated communication between the epicenters and the other counties i.e, decreased the influence of the epicenters on other counties, the results were equally staggering. Our model showed that by eliminating drug flow in and out of these 10 counties, we were able to cut the number of infected counties by 85% in 6 years. We believe that this decrease in opioid influence could be mimicked through border control along the major roadways in these counties and at state border. This strategy is formulated based on the observation that epicenters are generally located near large highways, thus providing them more influence in the spread of opioids. Border control would decrease the ability of opioids to flow across state borders along these major roadways and elsewhere, decreasing the influence these high influence counties have on the spread of opioid use.

Thirdly, we suggest advertising targeted at seniors older than 65 about the dangers of addiction and the risk presented by teenagers consuming opioids. In our model, one of the highest correlated factors in opioid use is households where seniors are taking care of children under 18. This is likely due to the fact that seniors are more likely than other groups to experience chronic pain, and thus receive a prescription of opioids, which comes with a possibility of addiction. By extension, many of the children in these households have greater access to opioids due to their presence in the household, and consume these opioids for recreational purposes. We tested this hypothesis, and found that when we decreased the proportion of these households by half, the number of infected counties decreases by nearly 90 percent. We understand that while it is very difficult to reduce the number of seniors older than 65 taking care of children, it is possible to create informative advertising targeted to this group to make them more aware of the problem. This advertising, placed on channels with high viewership by seniors, would warn them of the dangers of addiction.

Fourthly, we suggest encouraging health plans for veterans to include physical therapy, in addition to rehabilitation for those already addicted, as opposed to prescribed opioids. Our model identified the veteran population as a highly correlated group. We attribute this to the fact that veterans are frequently prescribed opioids to deal with the traumas of war and chronic pain. When we lowered the fraction of veterans in each county to half in our model, the results were shockingly effective with no infected counties. Therefore, we believe that by reducing the destructive relationship between veterans and opioids, we can dramatically reverse the opioid problem. We suggest that veterans be presented better access to physical therapy or other alternative treatments rather than the prescription of opioids. This could be done by encouraging health plans for veterans to include physical therapy. Not only would this limit the connection between this group and opioid use, but it would also decrease potential access to opioids by other groups.

Finally, we suggest making education more accessible to everyone. Our model showed a negative correlation between people with a bachelor's degree and opioid abuse. This demonstrates that people who are educated enough to know the potential drawbacks of long term opioid use will be at lower risk to try them for recreational purposes. When we reduced the number of people with a bachelor's degree by half, the number of infected counties nearly doubled its previous value. With this in mind, general education of all people on the dangers of opioids will achieve a similar effect. Apart from making education more accessible, we also suggest using television and other media to advertise the overall harm that can be done by opioid use, the overall education of the population with regard to prescription opioids and heroin will increase, resulting in less impulsive decisions to use them without need. While this will not raise the percentage of formally educated people in the region, we believe that it will exhibit a similar change to that predicted in our model.

We wholeheartedly hope you consider our suggestions, which will help this country decrease the influence of an epidemic that has claimed thousands of lives, and millions more currently under the grip of prescription opioids and heroin.

Thank You,
Team #1901213

References

[1] A. S. of Addiction Medicine, "Opioid addiction, 2016 facts & figures." https://www.asam.org/docs/default-source/advocacy/opioid-addiction-disease-facts-figures.pdf, 2016.

[2] "Cdc guideline for prescribing opioids for chronic pain united states, 2016 — mmwr." https://www.cdc.gov/mmwr/volumes/65/rr/rr6501e1.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fmmwr%2Fvolumes%2F65%2Frr%2Frr6501e1er.htm, March 2016.

[3] N. Volkow, "Addressing the opioid crisis means confronting socioeconomic disparities," 2017.

[4] J. Roth, "County distance database." https://www.nber.org/data/county-distance-database.html, 2016.

[5] J. Kingman, "Markov population processes," *Journal of Applied Probability*, vol. 6, no. 1, pp. 1–18, 1969.

[6] A. D. Barbour, "On a functional central limit theorem for markov population processes," *Advances in Applied Probability*, vol. 6, no. 1, pp. 21–39, 1974.

[7] G. J. Gibson, "Markov chain monte carlo methods for fitting spatiotemporal stochastic models in plant epidemiology," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 46, no. 2, pp. 215–233, 1997.

[8] S. Yakowitz, R. Hayes, and J. Gani, "Automatic learning for dynamic markov fields with application to epidemiology," *Operations research*, vol. 40, no. 5, pp. 867–876, 1992.

[9] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*, pp. 1–4, Springer, 2009.

[10] C. Robert, *Machine learning, a probabilistic perspective.* Taylor & Francis, 2014.

[11] J. Frost, "How high does r-squared need to be?," 2018.

[12] A. Baldini, "A review of potential adverse effects of long-term opioid therapy: A practitioners guide," 2012.

[13] J. Garami, "Intolerance of uncertainty in opioid dependency relationship with trait anxiety and impulsivity," 2017.

[14] S. Spampinato, V. Locatelli, D. Cocchi, L. Vicentini, S. Bajusz, S. Ferri, and E. E. Müller, "Involvement of brain serotonin in the prolactin-releasing effect of opioid peptides," *Endocrinology*, vol. 105, no. 1, pp. 163–170, 1979.

[15] N. A. Battista, L. B. Pearcy, and W. C. Strickland, "Modeling the opioid epidemic," *arXiv preprint arXiv:1711.03658*, 2017.

[16] "Preventing drug misuse and addiction: The best strategy — national institute on drug abuse (nida)." https://www.drugabuse.

gov/publications/drugs-brains-behavior-science-addiction/preventing-drug-misuse-addiction-best-strategy.

[17] "Preventing drug abuse and excessive alcohol use — surgeongeneral.gov." https://www.surgeongeneral.gov/priorities/prevention/strategy/preventing-drug-abuse-excessive-alcohol-use.html. (Accessed on 01/28/2019).

[18] "Curbing opioid addiction needs more than new drugs," 2018.

[19] S. Pei, S. Kandula, W. Yang, and J. Shaman, "Forecasting the spatial transmission of influenza in the united states," *Proceedings of the National Academy of Sciences*, p. 201708856, 2018.

[20] A. L. Pitt, K. Humphreys, and M. L. Brandeau, "Modeling health benefits and harms of public policy responses to the us opioid epidemic," *American journal of public health*, vol. 108, no. 10, pp. 1394–1400, 2018.

A Appendix: Epicenters and vulnerable counties

Vulnerable Location	County	State
1	Scioto	Ohio
2	Montgomery	Ohio
3	Philadelphia	Pennsylvania
4	Kanawha	West Virginia
5	Pulaski	Kentucky
6	Pike	Kentucky
7	Laurel	Kentucky
8	Hamilton	Ohio
9	Bell	Kentucky
10	Jefferson	Kentucky

Table 2: *Vulnerable Counties by Demographic Model from 2010-2016*

Vulnerable Location	**County**	**State**
1	Kanawha	West Virginia
2	Scioto	Ohio
3	Montgomery	Ohio
4	Pike	Kentucky
5	Whitley	Kentucky
6	Lincoln	Kentucky
7	Jackson	Kentucky
8	Marion	Ohio
9	Richmond	Virginia
10	Pulaski	Kentucky

Table 3: *Vulnerable Counties by Demographic Model in 2017*

Epicenter	County	State	Border
1	Carroll	Kentucky	A
2	Bracken	Kentucky	D
3	Morgan	West Virginia	B
4	Harlan	Kentucky	C
5	Marion	West Virginia	E
6	Doddridge	West Virginia	E
7	Gallatin	Kentucky	A
8	Harrison	West Virginia	E
9	Mecklenburg	Virginia	K
10	Campbell	Kentucky	D

Table 4: *Epicenters Ranked by QA Element-wise Product*

Epicenter	County	State	Border
1	Carroll	Kentucky	A
2	Morgan	West Virginia	B
3	Harlan	Kentucky	C
4	Campbell	Kentucky	D
5	Harrison	West Virginia	E
6	Marion	West Virginia	E
7	Gallatin	Kentucky	A
8	Hampshire	West Virginia	B
9	Bracken	Kentucky	D
10	Doddridge	West Virginia	E

Table 5: *Epicenters Ranked by QA Product*

Epicenter	County	State	Border
1	Carroll	Kentucky	A
2	Harlan	Kentucky	C
3	Morgan	West Virginia	B
4	Lee	Kentucky	G
5	Bracken	Kentucky	D
6	Campbell	Kentucky	D
7	Harrison	West Virginia	E
8	Marion	West Virginia	E
9	Allen	Kentucky	F
10	Nicholas	West Virginia	H

Table 6: *Epicenters Ranked BY Q Alone*

Epicenter	County	State	Border
1	Elliot	Kentucky	I
2	Harlan	Kentucky	C
3	Bell	Kentucky	C
4	Lewis	West Virginia	E
5	Owsley	Kentucky	G
6	Powell	Kentucky	G
7	Carroll	Kentucky	A
8	Lee	Kentucky	G
9	Bristol	Virginia	J
10	Morgan	West Virginia	B

Table 7: *Epicenters Ranked by* $|Q|$

B Appendix: Top 30 highly correlated socioeconomic factors

Here are the top 30 highly correlated socioeconomic factors, in no particular order:

- Number of grandparents responsible for own grandchildren under 18 years
- Number of grandparents living with own grandchildren under 18 years
- Number of females 15 years and over
- Number of males 15 years and over
- Number of people ever married
- Number of people now married
- Civilian population 18 years and over
- Population 3 years and over enrolled in school
- Total population
- Number of women 15 to 50 years old who had a birth in the past 12 months
- Number of people with a bachelor's degree
- Number of people with Ukrainian ancestors
- Number of people with Russian ancestors
- Number of people with Italian ancestors
- Number of people with Arab ancestors
- Number of people who speak Asian and Pacific Islander languages
- Number of people who speak English and Asian and Pacific Islander languages
- Number of people who lived in a different house, but same county
- Number of natives who entered the state before 2010
- Number of people born outside the United States who entered the state before 2010
- Number of foreign born who entered before 2010
- The world region of foreign born population
- Citizenship status of foreign-born population
- Disability status of 18 to 64 years
- Disability status of all civilians
- Disability status of 65 years and over
- Disability status of under 18 years
- Households with a spouse
- Family households with a married couple
- Houses with computers and broadband use

C Appendix: Supplementary images

Figure 3: *Spread of the opioid crisis from year 2010 to 2016. The legend tells the number of drug reports per 100 people in each county. A darker green indicates a higher number of drug reports per 100 people.*