For office use only	Team Control Number	For office use only
T1 ________________	**1901679**	F1 ________________
T2 ________________		F2 ________________
T3 ________________	Problem Chosen	F3 ________________
T4 ________________	**C**	F4 ________________

2019
MCM/ICM
Summary Sheet

After exploring the data provided to our team describing drug use and socioeconomic factors between 2010 and 2016 inclusive, we sorted the 69 types of opioid substances into four drug categories based on their synthesis and availability. Plotting use rates of each category of drugs over time revealed that use of mild painkillers and natural alkaloids have stayed relatively constant over time, semi-synthetic drugs have declined slightly, and synthetic drugs such as fentanyl and heroin have increased dramatically. These findings align with reports from the CDC. We also selected 54 out of 149 socioeconomic variables based on their variance inflation factor score (a common measure of multicollinearity) as well as their relevance based on the public health literature. To model the spread of the opioid crisis across Kentucky, Ohio, Pennsylvania, West Virginia, and Virginia, we took a dual-pronged approach, developing two completely different models and then comparing them at the end.

Our first model is founded on common modeling approaches in epidemiology: SIR/SIS models and stochastic simulation. We designed an algorithm from scratch which simulates a random walk between six discrete classes, each of which represent a different stage of the opioid crisis using thresholds for opioid abuse prevalence and rate of change. We penalize transitions between certain classes differentially based on realistic expectations. Optimization of parameters and coefficients for the model was guided by an error function which we also designed from scratch, and was inspired by the global spatial autocorrelation statistic Moran's I. Testing our model via both error calculation and visual mapping illustrated high accuracy over many hundreds of trials. However, this model did not provide much insight into the influence of socioeconomic factors on opioid abuse rates, because incorporating socioeconomic factors did not significantly change the model results.

Our second model made up for this deficiency in socioeconomic factor analysis. By running a collection of spatial regression models on our final collection of socioeconomic predictors (including total drug use rate), we explored characterising the spatial patterning of the opioid crisis as the result of a spillover effect, as the result of spatially-correlated risk factors, and as a combination of the two, using spatial lag, spatial error, and spatial Durbin models respectively. While all models confirmed significant spatial signals, the spatial Durbin model always performed the best. We also calculated the direct, indirect, and total impacts of each predictor variable on opioid abuse rate. Far and away the most important variable in all models was the total drug use rate in each county. The average result (across all seven years) was that, all else equal, a unit increase in total illicit drug use rate would raise the opioid abuse rate by 52%. This is quite realistic given a CDC statistic that in 2014, 61% of drug overdose deaths involved some type of opioid. By contrast, an ordinary linear regression reported only a 37% increase in opioid abuse rate per unit increase in total drug use rate. Statistical measures such as the Akaike Information Criterion and Likelihood Ratio Test verify the superiority of our spatial models.

To predict possible locations of origin of the opioid epidemic in each of the five states, we ran a Monte Carlo simulation of our random walk model from 2000-2010. We map out these counties and discuss their arrangement in the context of our other findings. The random walk finds that the opioid crises most likely starts in Montegomery, KY which aligns with our research that opioid abuse is more prevalent in rural communities than urban [10].

To forecast spread of the opioid crisis from 2017-2020, we used both our random walk and spatial regression models. The two models display surprisingly minimal deviance from one another, especially in 2019 and 2020. The random walk predicts that the number of counties above the illicit opioid use threshold will go down naturally within the next 7 years which also aligns with the idea that the opioid epidemic follows the spillover effect seen in epidemiology.

Due to the assumption that the SES indicators will change linearly, the second model's error will significantly increase after about 4-5 years. The random walk on the other hand, operates on a healthy tension between wanting to cluster together and randomly assigning classes. Near the initial date, it clusters more but the randomness starts to compound rather quickly. For this reason, the random walk has lowest errors near the 4-7 year mark. This means that the best strategy to predict the near future would be the spatial regression and the random walk for the 4-7 year range. Predicting anything beyond this point will have high error. Afterwards, we provide the the suggestion to the government that reducing general drug usage will help decrease illicit opioid usage.

Random Walks and Rehab:
Analyzing the Spread of the Opioid Crisis

Control #1901679

January 29, 2019

Contents

1 Introduction and Problem Statement

The deadly consequences of abusing prescription narcotic pain relief medications, heroin, and synthetic opioids are affecting people in all 50 states and across all socioeconomic classes. The opioid epidemic claims the lives of 115 people in the United States every day [46]. Through healthcare costs, rehabilitation treatment, lost productivity, and criminal justice involvement, the opioid crisis is costing the US federal government an estimated 78.5 billion dollars each year [47]. Our team was presented with the following modeling tasks: firstly, to characterize the spread of the opioid epidemic throughout Kentucky, Ohio, Pennsylvania, Virginia, and West Virginia and analyze resulting patterns; secondly, to incorporate socioeconomic factors in our model and analyze the associations, if any, between them and opioid abuse rates; finally, to use results from these models to recommend public policy strategies to combat the opioid epidemic. To perform these analyses, we were limited to sets of data covering the years 2010-2016 from the American Community Survey (ACS), which provided our socioeconomic indicators, and from the National Forensic Laboratory Information System (NFLIS) on illicit drug use. All data was provided at the county level in the aforementioned five states.

To characterize the spread of the opioid crisis throughout these five states, we developed two models. The first, of our own design, simulates a random walk through stable, endemic, and epidemic stages. The second is a more standard collection of spatial regression models. In this paper, after describing these two modeling approaches and their results, we report forecasts from both models on the future spread of the opioid epidemic, and compare the results. This dual-pronged approach provides diverse insights into the nature of the opioid crisis, and helps us to identify strategies for government intervention.

2 Etiology of the Opioid Crisis

2.1 A Brief Timeline of the Epidemic

Previously restricted to use for chronic pain due to cancer, opioids emerged into the non-cancer pain market following a series of studies published in the early 1990s indicating that pain was inadequately treated, with up to 42% of sample populations reporting mismanaged pain. Around this same time, pharmaceutical companies and medical societies began to relax their informal ban on opioids as they were reassured by somewhat erroneous studies and articles that opioids were not at all addictive. Compounding the problem was the release in 1996 of OxyContin by Purdue Pharma, along with a targeted marketing campaign that encouraged ongoing treatment for chronic pain and implied minimal side effects for the semi-synthetic drug [24].

Thus the **first great wave** of opioid prescriptions began. The number of prescription opioids in the hands of consumers continued to rise throughout the 1990s by as much as 2-3 million per year and into the early 2000s as the government (Joint Commission) released new standards surrounding the monitoring and treatment of pain.

Experts believe the **second wave** of the epidemic occurred around 2010 as the true nature of the epidemic began to surface. In an effort to curb addiction, government organizations placed limits on opioid prescriptions. Many of those who were already addicted turned to heroin instead, as it was cheaper and could provide a more intense high than standard prescription medication. However, due to illicit manufacturing, heroin was often impure or mixed with other drugs without consumers' knowledge, leading to increased deaths related to drug abuse in the United States [48].

The **third and hopefully final wave** of the epidemic came in 2013 with the rise of illegally manufactured synthetic alternatives to prescription opioids, such as fentanyl. However, efforts to combat the epidemic also intensified. The US Centers for Disease Control declared the opioid problem an official epidemic in 2011 [49] and the US federal government followed suit in 2017. Legislation calling the epidemic a "public health emergency," freed the Department of Health and Human Services to channel additional resources into treatment, distribution of overdose antidotes, and prevention research [59].

2.2 The Science of Addiction

Addiction is a chronic biophysical disease with genetic predispositions and long-term consequences. As a chronic disease, there is no cure and usually treatment must be continued throughout the life of the recovering addict. Note that addiction to and physical dependence upon a substance are quite different. Physical dependence is a common phenomenon and involves the biology of the body rather than the behavior of the individual. Addiction, on the other hand, can interfere with a person's self-control and challenge their ability to resist intense urges to use drugs [45]. This is why it is so common for addicted individuals to relapse, even years after "getting clean."

Addiction begins simply: a person participates in a behavior or has an experience that triggers the production of dopamine in their brain, making the individual feel happy. Usually these experiences are healthy and the release of dopamine is intended to make the person repeat the behavior; natural causes of this include connecting with other people in a community, sleeping, and eating. However, addictive substances like opioids and alcohol also activate dopamine production, usually at an elevated level, causing an individual to feel elated. More often than not, the individual will attempt to repeat this behavior to feel similar results [55].

Over time, the brain attempts to combat the addiction by reducing its number of dopamine receptors or producing less dopamine. Thus a larger amount of the substance must be consumed to repeat the high; this is known as *tolerance*. The decrease in dopamine receptors or in dopamine production can also lead to the individual finding little joy in other activities they previously enjoyed. Craving elation, the person will turn again to the substance. Since dopamine is also involved in the learning process, the brain will associate the drug with the high. Eventually, the desire for the drug becomes more important than the high itself. And for some substances, such as cocaine and opioids, addicted brains show fewer neurons in the frontal cortex, which means that cognitive reasoning and impulse control have been impaired [55].

2.2.1 Risk Factors

Genetics and environmental factors both play large roles in determining whether a person will become addicted to a substance. Some people are more susceptible to neurobiological changes that allow an addiction to take hold, while others might have a genetic predisposition to addictive behavior. Additionally, those with poor social support networks, individuals who have experienced trauma or abuse (especially at a young age), and people with some mental illnesses are more likely to become addicted to a substance than members of the general US population.

Currently, age appears to have the largest impact on susceptibility to addiction. The younger a person is, the more vulnerable she is to addiction. In fact, a federal study found that 74% of individuals admitted to treatment programs between 18 and 30 years of age had started abusing drugs before the age of seventeen. However, this same study discovered that the majority of people who were admitted for heroin and prescription painkiller addictions started using drugs after the age of twenty-five [3].

This is only one of many ways in which opioid addiction is different from other kinds of drug addiction.

2.3 Common Epidemiological Models

We now switch gears to the mathematical modeling side of epidemiology. Several different types of models and overarching principles in public health research helped to guide our approach to the given tasks.

2.3.1 Compartmental Models

A *compartmental model* is commonly used to simplify mathematical modeling of infectious diseases. Populations are divided into compartments with assumptions about the nature of each compartment and the time rate of transfer between them [27]. One such model is known as the SIR model, where the population is partitioned into three groups: susceptible (S), infected (I), and removed (R). S's are individuals who have not been infected, but who are susceptible to infection; I's are those who are infected and are capable of transmitting the disease; R's are people who can no longer contract the disease because they have recovered with immunity, been quarantined, or died [20]. In the simplest form of SIR model, individuals move directly from susceptible to infected to removed. Adaptations of this model include the SIS, in which there is no removed category and individuals move from infected back to susceptible, and the SIRS, in which immunity to the disease decays over time, eventually sending individuals in this category back to the susceptible category.

While there is some contention that the spread of opioid abuse can be characterized by infectious disease diffusion, dynamic compartmental models have already been applied to this topic. A 2018 study by researchers at Stanford University used an SIR model to predict opioid overdose deaths as far into the future as 2026 and then to assess the effects of different policies on the epidemic [2]. These investigators determined that reducing opioid prescriptions by 25% results in 2,500 fewer overdoses in the next ten years across the country. But this alone will not solve the problem, as opioid abusers will turn to heroin or illicitly manufactured synthetics. "It's like squeezing a balloon," commented addiction expert Keith Humphreys. "When you touch one aspect of the situation, an unexpected consequence often pops up somewhere else."

2.3.2 Stochastic Simulation

A stochastic simulation is one in which variables may change randomly based on certain probabilities. Because the spread of a disease (or any social phenomenon) is likely to be influenced by randomness, we decided to do further research on stochastic simulation applied to epidemiology. It turns out that many forms of stochastic epidemic models have been developed, which are often closely related to deterministic counterparts [34].

In 1965, a researcher named Thorston Hagerstrand changed the field of spatial diffusion models utilized in social science by incorporating an essential but overlooked factor: time. His *Monte-Carlo Approach to Diffusion* looked at the spread of an idea through a social network [22]. Seven years later in 1972, A.A. Brownlea adapted and applied this model to the spread of infectious hepatitis in Wollongong, Australia. He assumed a closed population, an equal chance of diffusion in all directions, and community immunity as the epidemic passed and simluated a random diffusion that advanced as a ring from the origin of infection. Of course, the actual advance of the disease was not a perfect ring, but any irregular bulges in its shape were attributed to ecological parameters, such as the locations of geographic barriers, which Brownlea had previously identified but not included in his model [20].

Another stochastic model that made headway around the same time as Brownlea's paper is called the Agent-Based Model (ABM). An ABM simulates a system as a collection of decision-making entities called agents. Each individually assesses its situation and makes decisions on the basis of a set of pre-programmed rules. Even in their most basic form, ABMs can exhibit complex behavioral patterns and are used to model the stock market, traffic, and diffusion within a population. In more complicated ABMs, agents can evolve and portray entirely new behaviors [19].

2.3.3 The Greenwood Model

Based on our research, we decided to combine the approaches of dynamic compartmental and stochastic simulation to describe the spread of the opioid crisis. This combination in itself is not a novel idea. One specific model called the Greenwood Model combines Markov chains (one of the simplest stochastic models) and the SIR model, with the additional assumptions that the total population size is constant, the number of infected individuals in a generation is a *binomial random variable*, and the probability of becoming infected is a constant not depending on the number of infected individuals.

Our needs diverge from the Greenwood model in that both our population size and the probability of "becoming infected" are dynamic, especially once we include socioeconomic factors as predictors– these predictors change over time. Thus, the foundation of our first model is a time-inhomogeneous random walk, where each county is acting as an "agent".

3 Foundations of the Models

3.1 Terminology

- In epidemiology and health statistics, **prevalence** is the percentage of people in a population who are sick with a disease at a particular point in time [20].
- An **epidemic** burdens a disproportionately large number of individuals within a population, region, or community at the same time; another definition of this term is a disease that occurs at levels clearly beyond normal expectation [42].
- A disease that is constantly present in a given area is called **endemic** [20].
- A disease that occurs with intense transmission, exhibiting a high and continued incidence is called **hyperendemic** [42].
- A phenomenon exhibits **spatial autocorrelation** if the presence of some factor in a sampling unit makes that factor's presence in neighbouring sampling units more or less likely [38].

3.2 Assumptions

1. County boundaries do not change significantly between 2010 and 2016.
2. All counties for which we have no drug data have low or no opioid abuse rates. This is reasonable because we were told that all counties with illicit opioid cases were reported in our data set.

3. We assume that research statistics which externally validate our model results are not included in the no external data ban.
4. Three months is the minimum period of time in which a county can transition between distinct stages in the opioid epidemic (discussed in detail in section 4.2, Model Testing).
5. Education level is a proxy for income and healthcare status, ancestry and language together indicate race, and veteran status can stand in for disability (discussed in detail in section 5.2, American Community Survey Indicators).
6. Linear extrapolation of socioeconomic indicators is an acceptable estimate within five years beyond the time span for which we have data (2010-2016).

3.3 Overarching Concepts

A significant peril in using aggregated data to characterize social or ecological phenomena is the temptation to commit the Ecological Fallacy. One commits this fallacy by asserting that associations statistically identified at one scale of analysis are valid at or can be generalized to either larger or smaller scales [20]. In this paper, we make every attempt to avoid committing this fallacy. However, in general we consider that attempting to draw large-scale conclusions based on many observations at smaller scales is more acceptable than attempting to draw fine-scale conclusions based on associations at large scales. In other words, we consider it acceptable to incorporate some logic about individuals' opioid abuse into our model for county opioid rates, but unacceptable to go the other way and claim that the patterns we observe at the county level represent the experiences of individuals, or even necessarily of municipalities (depending on the size of a municipality relative to its county).

Another major theme in spatial analysis for public health is the need to determine whether a spatial autocorrelation of a health outcome is due to a spillover (diffusion) effect or is simply explained by regions near each other having similar social, economic, and environmental characteristics, which result in the development of similar health outcomes. In this paper, we explore the possibilities of characterizing the spatial patterning of the opioid crisis as the result of a spillover effect, as the result of spatially-correlated risk factors, and as a combination of the two. This kind of analysis, coupled with intentional avoidance of the ecological fallacy, helps to inform future modeling approaches as well as our recommendation of policy interventions.

3.4 Exploring the Drug Report Data

Tracking the 69 unique drugs tagged in the National Forensic Laboratory Information System (NFLIS) data is a burdensome task and one that would probably not produce statistically significant results, due both to the sheer variety of opioids and the labelling discrepancies from the different crime laboratories. For instance, some labs will label a drug generically as fentanyl, while others will note the specific kind of fentanyl. Thus, to diagnose meaningful trends in the NFLIS data, we divided opioids into four categories based on their chemical synthesis and availability:

[noitemsep]*Methadone, buprenorphine, etc.*: these are mild painkillers and in some cases are used to treat opioid addiction. As such, they are easily available in clinics and are not as intensely regulated as other opioids [5], most of which are controlled substances. *Hydrocodone, oxycodone, etc.*: these are semi-synthetic opioids and are among the most addictive and deadly, contributing to the most overdose deaths of any prescription opioid in 2017 [26]. Their semi-synthetic nature makes them more difficult to produce outside of a laboratory; they are likely abused as prescriptions or through overlooked distribution leaks [60]. *Fentanyl, heroin, U-48800, etc.*: this is the largest category and the one with the highest increases in abuse and overdose deaths in recent years. It includes mainly synthetic opioids, of which more are introduced every year. For instance, U-48800 made its debut in American cryptomarkets as recently as 2017 [39]. Heroin is included in this category despite its semi-synthetic nature because it is illegal, and thus like other synthetic drugs, it can/must be made entirely without the assistance of a professional laboratory. *Morphine, codeine, etc.*: these are the natural alkaloids of the opium poppy and are among the oldest opioids. They are less potent than their semi-synthetic and synthetic cousins; codeine can be found in varieties of Tylenol painkillers [4]. Medicines of this variety are easier to find in medicine cabinets and in some cases can be gateways to more potent opioids.

As seen in Figure 1, the prevalence of the opioids in each category (with the exception of synthetics) remained relatively constant over the six years for which we have data. The synthetic opioids became increasingly prevalent following 2011; this aligns with reports from the Centers for Disease Control [52].

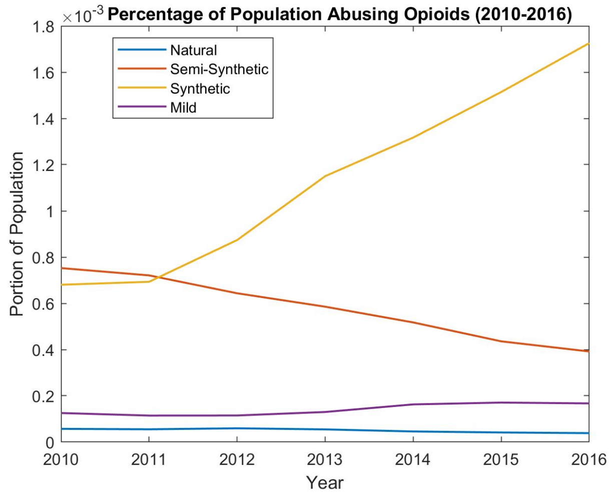


Figure 1: Drug Trends for Each Category from 2010-2016

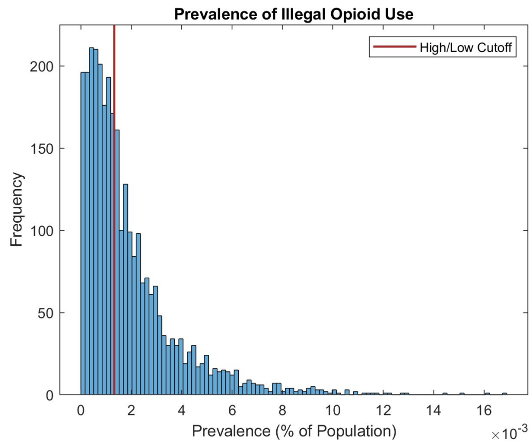
4 Building a Model, Part I: Characteristics of the Epidemic

One important aspect of our problem is the discrete nature of both annual and county-aggregated data. This led us to consider discrete time and discrete space (DTDS) Markov models. A stochastic process has the Markov property if the conditional probability of future states in the process depends only upon the current state, not the sequence of past states [30]. In formulating our first model, a random walk, we decided early on to incorporate this property. Our intuition was that just like an individual recovering from addiction, a county may regress if it does not make consistent efforts to address drug abuse. Thus, the situation of each county in past time steps does not have direct influence on the situation of that county in future time steps.

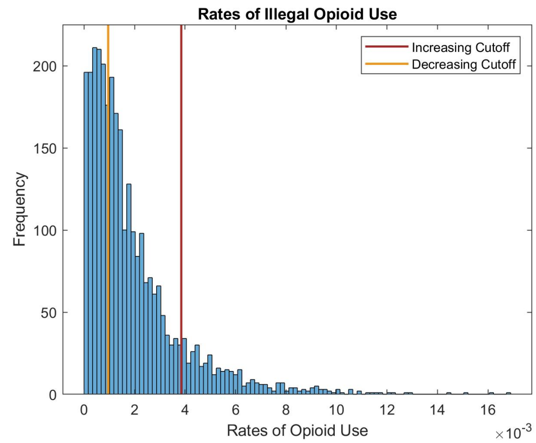
4.0.1 Discrete Classes for Prevalence and Rate of Change

Based on our research of SIR and other compartmental models, our immediate intuition was to organize data into "endemic", "epidemic", and "recovered" categories instead of investigating the exact prevalence of opioid abuse in each county. This would allow us to better detect large-scale trends in the epidemic over time. Further development of this idea led us to a county categorization based on "low" or "high" opioid abuse prevalence in combination with relative changes in prevalence.

To determine the cutoff points for low, high, increasing and decreasing, we inspected histograms of both prevalence and rates of change in prevalence (from year to year). For prevalence, we chose the median. *Note: in Figure 2a, the red line represents the median. In Figure 2b, the orange and red lines represent the decreasing and increasing cutoffs, respectively. The x-axes have a scale of* 10^{-3}.



(a) Prevalence of Opioid Abuse



(b) Rates of Change of Opioid Abuse

For rates of change, we classified all values greater than one standard deviation above the mean to be increasing, and all values lower than one-half standard deviation below the mean to be decreasing. Our reasoning for the different magnitudes in the positive and negative directions is (1) that our data is all from years when the opioid crisis was in full swing, so normal in our data set is not the normal in general, and (2) that it is harder for a county to rehabilitate than to develop an opioid abuse problem, so smaller changes in the negative direction represent larger changes in the county. All values that were between our increasing and decreasing cutoffs were classified as stable, then classified based on absolute level of prevalence.

Initially, we devised four categories: low and stable (LS), high and stable (HS), increasing (I), and decreasing (D). These were meant to represent counties being in endemic, hyper-endemic, epidemic, and recovering stages regarding opioid abuse. After significant development of our model, we decided to use six categories: low and stable (LS), high and stable (HS), low and increasing (LI), high and increasing (HI), low and decreasing (LD), and high and decreasing (HD). We will use our model with six classes to show the results of our model, and discuss the results from our model with four classes in sensitivity analysis.

Note: any counties on our map for which we had no data were classified as "low stable". This was a reasonable decision because we were informed that any county for which no illicit opioid cases were reported in a year was simply not included in the NFLIS data for that year.

After classifying each of the counties in the NFLIS data, we made histograms of the proportions of counties in each class in 2010 and 2016. To our surprise, the histograms looked nearly identical. Fortunately, we had the wherewithal to map the results before despairing and giving up on this approach. While the proportions of each class stayed constant between 2010 and 2016, the spatial distribution of the classes did not. In fact, the pattern reminded us of the description of Brownlea's hepatitis model as a ring-shaped clinical front, expanding radially from the city of Wollongong [20]. The maps below depict the prevalence of opioid abuse in each county after sorting into our six classes. On the left we represent the epidemic in 2010 and on the right, 2016. Note the "ring-like" expansion of the darker-red sections (which indicate regions of high and stable opioid abuse). *Note: in the following sections, we will depict a number of similar maps. The numbers in the legend correspond to each of our six classes: 1 – Low Stable, 2 – High Stable, 3 – Low Decreasing, 4 – High Decreasing, 5 – Low Increasing, 6 – High Increasing.*

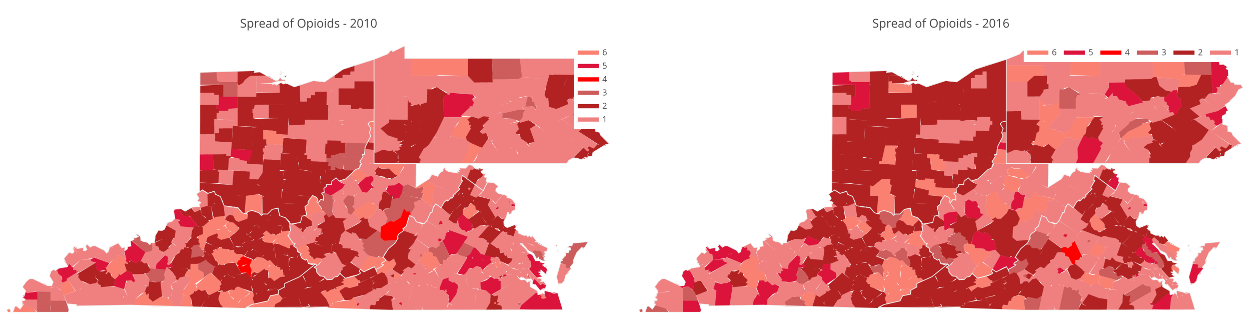


Figure 3: Spread of Opioid Epidemic from 2010 (left) to 2016 (right)

This discovery gave us both confidence to continue, and a suspicion that the rough origin of the opioid epidemic in this region of the US was southwestern West Virginia, based on visual inspection. Although our team could not formally quantify our visual stimulation by these maps, these observations held some weight given how adept our eyes and brains are at pattern recognition [23].

4.1 How the Model Works

In basic terms, the classification of each county in our model at the next time step is influenced by its current class, the classes of its neighboring counties, a noise parameter, and random selection from a probability vector. To run the model, we wrote code in Python which performs the following tasks:

1. Initialize the probability vector by adding noise uniformly to each class, so the probability of a county transitioning from one class to any other is greater than zero. The level of noise added to each class was optimized before running the model to minimize the error (discussed below).
2. Each neighboring county adds a tally to the index of its class in the probability vector to increase the probability of the original county being in that class in the next time step. This tally incorporates inverse distance weighting. In other words, a county directly adjacent adds a tally of 1, a county with one county in between it and the origin county adds a tally of 1/4, etc.
3. Divide each entry in the probability vector by the relevant transition score in the penalty matrix below. This matrix reflects our expectations about how difficult it should be to transition between each class.

$$\begin{array}{cc} & \begin{array}{cccccc} LS & HS & LI & HI & LD & HD \end{array} \\ \begin{array}{c} LS \\ HS \\ LI \\ HI \\ LD \\ HD \end{array} & \left[\begin{array}{cccccc} 1 & 4 & 2 & 3 & 2 & 1 \\ 4 & 1 & 3 & 2 & 3 & 2 \\ 2 & 3 & 1 & 2 & 2 & 2 \\ 3 & 2 & 2 & 1 & 4 & 2 \\ 2 & 3 & 2 & 4 & 1 & 2 \\ 3 & 2 & 2 & 2 & 2 & 1 \end{array}\right] \end{array}$$

4. Scale the probability vector by a coefficient vector which was previously optimized to account for the overall importance of each class, in the same manner as the noise was optimized.
5. Normalize the probability vector so it is actually a discrete probability distribution.
6. The final probability vector defines the distribution for a stochastic simulation which selects a class for the county in the next time period.
7. Repeat the steps above for some number of time steps and then for some number of trials so each trial walks through the specified number of time steps.

After running all of the trials, pick the majority vote for each county at the point in time we are interested in analyzing.

4.1.1 Evaluation of the Model

Once we created the model's structure, we realized we needed an error function to evaluate model performance and help us optimize the levels of noise and the values of coefficients for each class. Error analysis is tricky for this model because a good result does not necessarily mean a perfect match to the data. If our algorithm classifies a county as high stable, and the data says it should be high increasing, we want the error to reflect the fact that our model was close, and not so wrong as to classify that county as having low prevalence of opioid abuse. This need to distinguish the probabilities of certain transitions gave us the idea for the penalty matrix, displayed in the previous section.

However, after imposing the penalties, the model results started clustering a lot more than the real data, as seen in Figure 4, below.

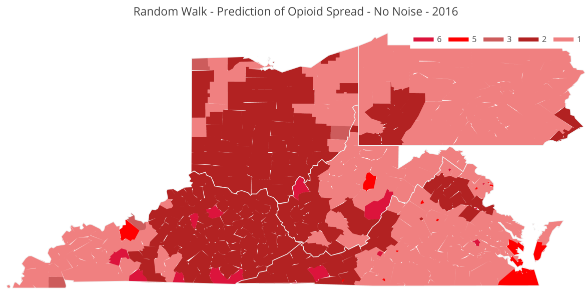


Figure 4: Clustering (2016)

Instead of trying to characterize clustering at each specific location, we had the idea to devise an error function based on overall spatial autocorrelation in the study area. We looked to the Moran's I statistic [20] (a standard measure of spatial autocorrelation) for inspiration and came up with the following formula:

$$\mathrm{A}_{\text{class}} = \frac{1}{\text{number of counties of that class}} * \sum^{\text{counties of that class}} \left(\frac{\#\text{neighbors with same class}}{\#\text{neighbors}} \right) \tag{1}$$

$$\mathrm{Err}_{\text{total}} = \sum^{\text{classes}} \left| \mathrm{A}_{\text{predicted}} - \mathrm{A}_{\text{real}} \right| \tag{2}$$

In plain English: for each class, for all counties in that class, our spatial autocorrelation measure is the mean proportion of neighbors with the same class. Then we take the sum over the classes of the difference between the simulated and real spatial autocorrelation measures.

Upon using this error function to optimize levels of noise and other parameters in our model and mapping the results, we found that our clustering problem had been all but eradicated, leaving us with results akin to the those depicted below. *Note: each time we run the model, we get slightly different results due to the random nature of added noise.*

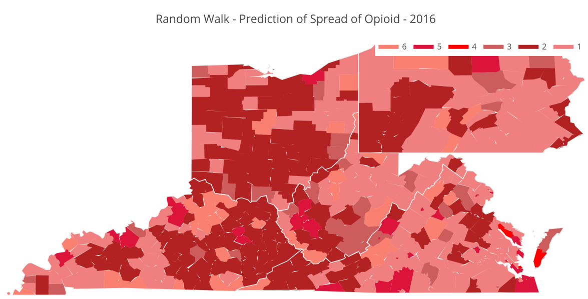


Figure 5: Best Prediction of Class Distribution (2016), Naive Model

It is notable that we are able to achieve such similar patterning to the real data by focusing only on the penalty matrix and overall spatial autocorrelation.

4.2 Model Testing

During model testing, we observed that the histograms of the six classes stayed more or less constant over time, depending on the levels of noise that we added to the initial probability distribution. They all start off close to the distribution of the data. Then, models with more noise start to diverge. However, by 2016, the noise has compounded sufficiently so that the simulations with noise have distributions which match the data closer than the simulations without noise. Using our new error function, we determined that the optimal level of noise was 0.3 (shown in the histogram below).

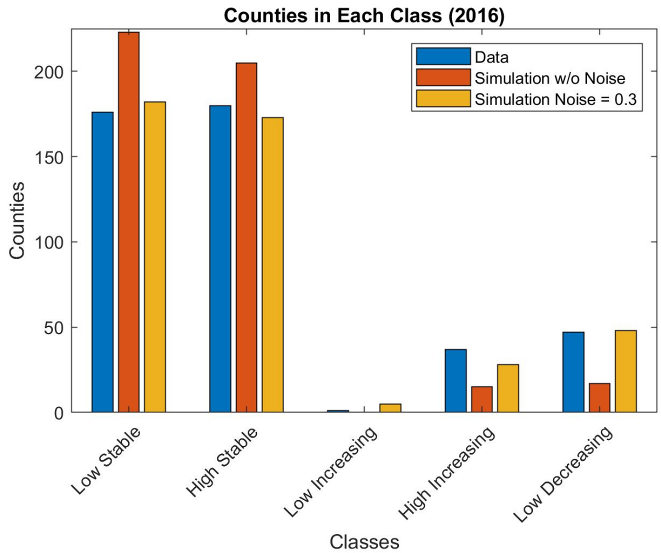


Figure 6: Class Prevalence for Data and Simulations (2016)

Another aspect of model testing was deciding what time step to assume. Because the random walk is on its own after being initialized by the 2010 data, there was no obvious reference point from which to infer optimal time step size. Eventually, we realized that later optimization of coefficients would fit the model regardless of step size, so we chose the step size of three months (four between each year) for ease of comparison to the annual data and with the thought that three months is about the minimum time frame in which a county would switch between one of our classes. The final optimization determined the coefficients on each class. *Note: Formal error and sensitivity analyses are discussed in later sections, after incorporating socioeconomic factors. Future trends of this model can be found in section 5.6, where they are compared to those of the model with socioeconomic data.*

5 Building a Model, Part II: Socioeconomic Factors

5.1 In the Literature

The opioid epidemic is especially notable for its range not only geographically but across socioeconomic classes. Upon initial examination, there appears to be little to no socioeconomic pattern in its spread throughout the country. However, after extensive analysis of opioid abuse data, researchers have discovered much about the communities which are currently affected most dramatically by the epidemic. Three major findings are explored in detail below.

Race: Non-Hispanic white Americans have been disproportionately affected by the epidemic, with whites roughly 50% and 167% more likely to overdose from opioids than blacks and Hispanics, respectively [16]. According to the Substance Abuse and Mental Health Services Administration (SAMHSA), over 7 million white Americans over the age of 12 were abusing opioids in 2016 and 2017, compared to 1.2 million black Americans and 1.9 Hispanic or Latino Americans [51]. The epidemic first besieged whites due to an increase in opioid prescriptions in the 1990s; at the time, white Americans were much more likely than any other race to be prescribed opioids by a physician. However, the racial gap is quickly narrowing as heroin and synthetic drugs such as fentanyl become more popular. Opioid prescription rates for both black and white Americans were the same (23%) in 2015 [9], but opioid abuse has increased by nearly 107% annually for blacks, compared with only 79% for whites [15].

Education: While there does appear to be a correlation between level of education and opioid abuse, researchers are split on its direction. According to SAMHSA's annual drug abuse survey, education level and opioid abuse is directly correlated to a point. High school graduates, those with some college, and people with a four-year college degree reported the highest rates of opioid abuse. People with advanced degrees and without high school diplomas reported much lower rates of abuse [51]. On the other hand, a few private studies indicate opioid abuse is negatively correlated with education level, finding overdose deaths among people with less than a high school education are more than double those among people with a college degree [53].

Veteran Status: A majority of veterans (64%) are prescribed opioids upon return to civilian life [12]. Coupled with high levels of severe mental illness (that increases susceptibility to addiction), it comes as no surprise that opioid abuse rates among veterans are the highest of any population (4.5%) [50]. In the years 2010 to 2015, the number of veterans addicted to opioids rose 55%. Thankfully, the Veterans Affairs Department (VA) has started to cut back on opioid prescriptions and defer their patients to other forms of pain management, including physical therapy [36].

Other risk factors for opioid abuse include gender (women, despite recent increases, are less likely to abuse opioids than men [29]), age (younger people are more at risk [28]), and population size (opioid abuse is more prevalent in rural communities than in urban settings [10]). Initially, we considered disability as a possible risk factor because of the higher rates of opioid prescriptions associated with chronic pain, but very little data on this topic is available; in fact, the National Institute on Disability, Independent Living, and Rehabilitation Research recently put out a Request for Information regarding the subject [44].

5.2 American Community Survey Indicators

The subset of the American Community Survey (ACS) data from 2010-2016 provided to us for this problem included variables about household size and family structure, age and gender distribution, educational enrollment and attainment, veteran status, disability status, residential mobility, place of birth, language spoken at home, and ancestry [43]. It did not include certain statistics which our research indicated would be useful in our analysis of socioeconomic factors' association with opioid abuse, such as income, unemployment rate, health care coverage, and race. Also, for the sake of consistency across years, we decided to remove any variables for which one or more of the years was missing data. This process removed fertility statistics, disability status, citizenship status, world region of birth for foreign born, and several miscellaneous household/family structure variables– in total, 27 out of 149 socioeconomic factor percentages. Because of these discrepancies between the data we had and the data we wanted, we had to assume that education level was a proxy for income and healthcare status, ancestry and language indicated race, and veteran status stood in for disability.

Because there was so much information in the survey data regarding household size and family structure, we focused subsequent research on any ties between household/family structure and opioid addiction. While the epidemic does seem to affect families of all kinds, we did notice an increase in grandparents raising their grandchildren in areas where the opioid epidemic has hit hardest [54]. This occurs as parents are separated from their children, both voluntarily and not, due to their addiction. Data on grandparents responsible for their grandchildren was available from the survey data, so we included it in our model.

5.3 Model Optimization with Socioecononomic Status

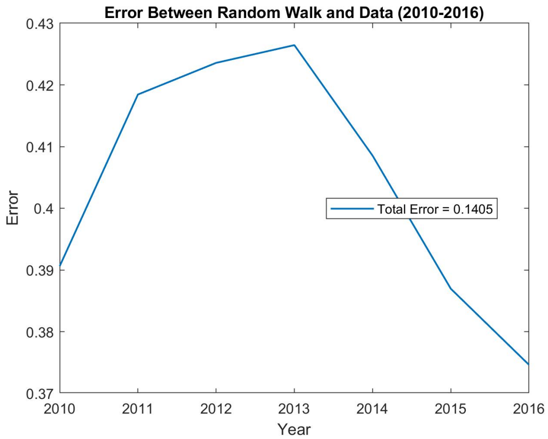
To incorporate socioeconomic factors into our random walk model, we first ran a random forest algorithm in **sklearn** that would classify each county into one of our six classes based on 23 socioeconomic predictors. (A more detailed explanation of our selection of these predictors is in section 6.3, below.) We also experimented with logistic regression, but random forest always performed better, achieving a test accuracy of 90%. Using the feature importance attribute in **sklearn**, we found that the top ten most important socioeconomic factors (according to the classifier) were the total illicit drug use rate, total population, people born in the US, American ancestry, Irish ancestry, only English spoken at home, people with some college but no degree, high school graduation rate, Polish ancestry, and people with a graduate or professional degree. Unfortunately, these feature rankings are based on absolute magnitude, so they do not provide insight into the direction of influence on opioid abuse rates. Our adapted model uses probabilities generated by this random forest classifier to initialize the probability vector for our random walk. The algorithm then proceeds as before.

After making this adaptation, we used our error function to compare the performance of the new model to the old. When we do not optimize the coefficients on each of the classes, the new model performs better than the old model. However, once we optimize the coefficients, both models become more accurate, and their performance becomes comparable.

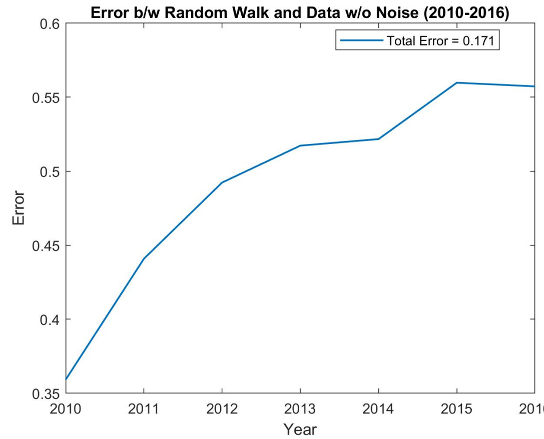
As seen in Figure **??** below, the per year error of the model increases a but but in the middle but decreases again after 4 years. This information along with the fact that there is a lack of testing data would usually suggest that the model is over fitting too the specific scenario between 2010 and 2015 via the coefficients optimization. The coefficient optimization was done to reduce the total error between 2010 and 2015 but the error still decreases in 2016. This lack of convergence informs us that over fitting should not be a problem. Thus, we rely on our second model (spatial regression) to provide further insight into the importance of socioeconomic factors in characterizing the spread of the opioid epidemic.

5.4 Error and Sensitivity Analysis

The random walk model takes on high error immediately after the start of the simulation. With noise, the error peaks around 2013 and then begins a steep descent; without noise, the error continues on an upward path. These trends are depicted in the Figures 7a and 7b, below.



(a) Error b/w Random Walk and Data (Noise)



(b) Error b/w Random Walk and Data (w/o Noise)

Our model supports a healthy tension between the clustering we noticed early on and the randomness added later with the inclusion of noise. However, this balance takes about 24 time steps or 6 "years" to reach, which is why the error is much higher towards the beginning of the simulation than at the end.

A second form of error comes from the lack of decimal points in the optimization of our coefficients. The cost of running this simulation for long periods of time or with extended trials is high, which increases the cost of optimizing the coefficients by several degrees. Due to the time constraints of the competition, we chose to limit the accuracy of our coefficients and reduce the time required to optimize them. Given more time, the error would have been lower.

5.5 Possible Origin Locations

To avoid adding bias to our model through reliance on linearly-extrapolated socioeconomic factors, we decided to run the origin identification analysis with the old model, which did not rely on socioeconomic factors. We ran a Monte Carlo simulation to find the possible origin locations, starting the epidemic in each county in 2000, during the height of the first wave of the opioid crisis. The simulation then propagated forward in time until 2010, at which time we compared the results to the given data. The simulations with the lowest amount of dissimilarity to the 2010 data represented the counties in which the epidemic likely began. The figure below highlights the counties in which the epidemic could have started (purple) and the counties that register as the first to contract the epidemic in their respective states (blue). The most likely origin of the crisis according to the random walk model is Montgomery, Kentucky (shown in gold in the map below).

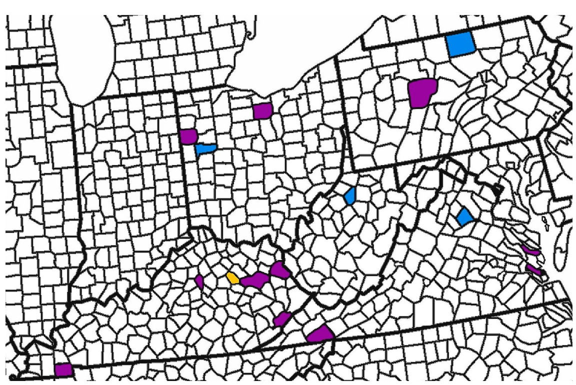


Figure 8: Origin Locations of the Epidemic

When analyzing the characteristics of the epidemic in the previous section, we mentioned our belief that the crisis had originated somewhere in southwestern West Virginia, based on visual inspection of our results and the Brownlea ring effect discussed in section 2.3.2. While this was not entirely correct, the actual origin according to our model is not far from this region, reinforcing the importance of the ring effect. *Note: We discuss future trends and drug identification thresholds in comparison with the second model. This discussion is located in section 6.4.*

6 A Second Modeling Approach: Spatial Regression

To further our understanding of the opioid epidemic, we explored the application of spatial regression models to the NFLIS and ACS data. Spatial regression is commonly used to model social and biological phenomenon [58, 13]. Taking into account spatial relationships is important because spatial correlation dramatically reduces the information contained in a sample of independent data. A conservative rule of thumb is that it reduces the data set information by a factor of 2 [33].

Within spatial regression, there are three common types of models. Spatial Autoregressive models (SARs), also known as spatial lag models, quantify the spatial dependence of the dependent variable y among neighboring regions. In other words, they quantify the diffusion or "spillover" effect of y [57]. The formal model is:

$$y = \rho W y + X\beta + u$$

where W represents the spatial weights (adjacency or distance between regions), ρ is the spatial autoregressive coefficient, and the error u is assumed to be classical (independent of y) [62].

Spatial Error Models (SEMs) quantify spatial dependence of the residuals. Instead of a spillover effect, we can conceptualize spatial error as spatial correlation in one or more unidentified predictor variables. The formal model is:

$$y = X\beta + u$$

$$u = \lambda W u + \nu$$

where in addition to the variables defined above, λ is the spatial autoregressive coefficient, and $\nu \sim N(0, \sigma^2)$ [62].

Spatial Durbin Models (SDMs) are a combination of spatial lag and spatial error models [14]. By lagging the predictor variables in the model using W, we can get a collection of spatial predictor variables in addition to the regular predictors. The formal model is:

$$y = \rho W y + X\beta + W X\theta + e$$

where in addition to the variables defined above, θ is a vector of the regression coefficients for the lagged predictor variables and e is the error. [57]. It is also possible to include the lagged predictors in a spatial error model, resulting in a Durbin Error Model (DEM), but that model appears to be less common and we do not explore it further in this paper.

The question of model specification depends on whether we think the spatial patterning of the dependent variable (in our case, opioid abuse) results from diffusion or is simply influenced by spatially-varying risk factors. Fortunately, we can rely on statistical tests as well as the combination of our research and intuition to select a model.

6.1 Model Fitting

To perform statistical regression, we used an R package called **spdep**. This package facilitates model specification and interpretation of results in addition to the regression procedure. To quantify the spatial weights (W in the model), we used a shapefile of all US counties [21], subsetted it to include only our five states of interest, and converted it to a spatial weights object, illustrated in Figure 9.

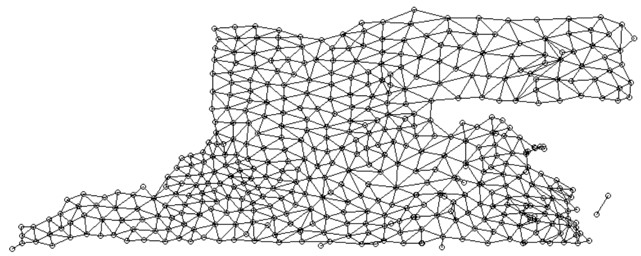


Figure 9: Visualization of Spatial Weights Between Counties

Because our ACS data set did not include all the counties in the shapefile, we spent a while figuring out how to run the regressions with missing data. Then, we experienced trouble getting the models to converge. It turns out that having highly correlated variables and predictors with lots of digits (i.e. number $\times 10^{\pm \text{high power}}$) cause problems during the numerical solution of matrices. After reading lots of documentation and diagnosing these issues, we adjusted the numerical tolerance levels, rescaled the predictor variables, and removed highly-correlated variables. We identified the highly-correlated variables using the method of variance inflation factors (VIF). The formula for VIF is as follows:

$$VIF_k = \frac{1}{1 - R_k^2}$$

where R_k^2 is the R^2-value obtained by regressing the k^{th} predictor on the remaining predictors [25]. Thus, if a predictor is highly-correlated with the other predictors, its R_k^2 value will be large, the denominator of the VIF expression will be small, and the VIF score will be large. A common heuristic for addressing multi-collinearity is to remove predictors with VIF scores over 10. In our data, the group of variables with VIF scores over 10 included many overlapping measures of household/family structure, educational enrollment and attainment, residential mobility, place of birth, and language. Removing that group of variables left us with 54 predictors, mostly from the ACS but also including total illicit drug use rate (TotalDrugReportsCounty from NFLIS divided by total population from ACS), plus our dependent variable: illicit opioid use rate (NFLIS DrugReports divided by total population). Note: even though several of the educational attainment categories has VIFs around 8 or 9, we decided to keep those in the model because we read so much about education and income level (for which education is a proxy) in our research about socioeconomic factors. We later confirmed that this was an acceptable decision when we ran the models with only the variables whose VIF scores were less than 4 and got extremely similar results (slightly different coefficients but the same statistical significance).

In the end, all the work paid off. We successfully ran regressions for spatial lag, spatial error, and spatial Durbin models, using the functions **lagsarlm**, **errorsarlm**, and **lagsarlm** with *type = "mixed"*, respectively. Note: the way these models are coded, it is not possible to have more than one value per spatial region (county). Thus, we had to run each year of data separately, and consider the results in aggregate. Although it did add some complexity to our workflow, this allowed us to analyze temporal non-stationarity, the ways that the models reflected differences in the data across the years.

6.2 Statistical Results

After running all the models, we analyzed the results. Every single model (across the seven years) confirms a highly-significant spatial signal in the data. The Likelihood Ratio Test (LRT) is a ratio of the likelihood functions from two nested models, one with more parameters than the other[41]. In this case, the LRT compares spatial linear regression with regular linear regression. All p-values on Likelihood Ratio tests and asymptotic t-tests for the spatial autoregressive coefficients ρ and λ (for SAR and SEM respectively) were $\leq 2^{-3}$. This means that spatial regression gives us significantly more information than regular regression would have.

Regarding spatial model specification, the most direct comparison can be based on the maximized log-likelihood [14]. Across all years, the spatial Durbin model had the highest Log Likelihood. This makes sense because we expect the spread of the opioid epidemic to display both spatial lag and spatial error patterning, based on both human interactions and socioeconomic and/or regulatory patterning. *Note: the Likelihood Ratio Test cannot be used to compare spatial lag and spatial error models directly because they are not nested [14].*

We report the overall results in Table 1 below.

Year	SAR LogLik	SEM LogLik	Durbin LogLik	Durbin ρ
2010	2480	2479	2544	0.299
2011	2382	2378	2442	0.205
2012	2405	2400	2454	0.128
2013	2462	2456	2510	0.209
2014	2442	2442	2496	0.188
2015	2408	2408	2457	0.208
2016	2475	2485	2529	0.214

Table 1: Log Likelihood of each Spatial Model and Spatial Autoregressive Coefficient for Best Model

Each SAR and Durbin model also performs a Lagrange Multiplier (LM) test for residual autocorrelation. About half of the SAR and Durbin models showed significant residual autocorrelation (p-value < 0.01). This indicates (a) there may be other spatially autocorrelated variables which could improve our model and/or (b) our error term is heteroskedastic (has non-uniform variance) [58]. To test the latter, we perform a Breusch-Pagan (BP) test on each spatial Durbin model. Sure enough, all BP tests showed highly-significant heteroskedasticity in the residuals. In spatial models, this is often due to spatial units having different population sizes [58]. We include more discussion about potentially important other variables in section 9.1 (Future Exploration).

Unlike the coefficients produced by ordinary linear regression, the coefficients from a spatial lag model do not facilitate interpretation, because a change in one predictor in one region influences response in other regions, which in turn influence the response in the region where the initial change occurred [57]. To account for both direct (local) and indirect (spillover) effects, we used the **impacts** function in **spdep** to calculate the global average impact of a unit increase in each predictor variable. Our results indicated that total illicit drug use rate in a county was by far the strongest predictor of opioid abuse levels in that county. Below is a subset of the output from **impacts**, showing direct, indirect, and total influence of a unit increase in total illicit drug use rate on opioid use rate.

Year	Direct	Indirect	Total
2010	0.388	0.236	0.624
2011	0.384	0.178	0.562
2012	0.353	0.126	0.478
2013	0.356	0.149	0.505
2014	0.409	0.123	0.532
2015	0.387	0.118	0.505
2016	0.332	0.080	0.411
Average	0.372	0.145	0.517

The Centers for Disease Control reported that in 2014, 61% of drug overdose deaths involved some type of opioid, including heroin [32]. By 2017, this statistic was 67% [26]. Because we know that opioid use rates have increased dramatically since the 1990s, our finding (averaging the estimates from 2010 through 2016) that, ceteris paribus, a one-unit increase of total illicit drug use rate would raise the opioid abuse rate by 52% seems very realistic. For comparison, we ran an ordinary linear regression on the same predictor set and found that while the R^2 value was 0.74 (relatively high), the coefficient on total illicit drug use rate was only 0.374. It is notable that our average direct impact rate was 0.372. This result confirms that our spatial model is far superior to a regular regression model in predicting opioid abuse, because it takes into account the indirect impacts of spatial diffusion. The Akaike Information Criterion (AIC), another output from the **sarlm** models, formally quantifies this comparison for all the models.

10 illustrates the most important variables in 2012, a fairly representative sample of the seven years. A full listing of the impact coefficients (direct, indirect, and total) for 2014, another fairly representative model, is in the Appendix.

Predictor Variable	Total Impact
All Illicit Drugs Rate	0.4784479
Ancestry: Norwegian	-0.0005062
Ancestry: Danish	-0.0004275
Ancestry: Portuguese	-0.0003794
Ancestry: West Indian, Excluding Hispanic Origin	0.0003416
Ancestry: Arab	0.0002836
Lived Abroad Last Year	-0.0002458
Ancestry: Ukrainian	-0.0002386
Single Male with Children Under 18	0.0001958
Ancestry: French Canadian	0.0001874

Figure 10: Top Ten Predictor Variables with the Largest Impacts in the 2012 Spatial Lag Model

Note: this table was scanned in from R, hence its label as a Figure, not a Table.

A union of the seven lag models' top ten most important variables (other than total illicit drug use) shows:

- Ancestry: overall, northern and eastern European ancestry was a negative predictor, while French Canadian, Arab, West Indian, and Slovak were positive predictors
- Housing/family structure: unmarried partner was a negative predictor; single male with children under 18 was a positive predictor
- Miscellaneous: Living abroad last year was a negative predictor

None of these variables seem particularly helpful in informing policy to help address the opioid crisis. Also, because there is such a discrepancy in impact coefficient size between the total illicit drug rate and all other predictors (in every model), focusing on the total illicit drug rate would have much more influence. We discuss strategies for reducing the total illicit drug rate in section 7 (Strategies to Combat the Opioid Epidemic).

6.3 Sensitivity Analysis

In earlier phases of testing the spatial regression models, we used a smaller subset of the socioeconomic variables, manually selected based on our research. (These 23 variables emphasized educational attainment and ancestry, and also included percentages of grandparents responsible for grandchildren, civilian veterans, those who moved in the last year, birthplace, and speaking only English at home.) The results were not significantly different than those from the all-encompassing dataset. The spatial signal was highly significant across all years and models; the spatial Durbin models all performed the best. The average direct, indirect, and total impacts of a unit increase in total illicit drug use rate on opioid use rate were 0.377, 0.154, and 0.531 respectively, each of which are less than 0.01 different from the corresponding average impacts from the larger model. We hypothesize that the values from the larger model are slightly smaller because the more variables there are to inform spatial patterning, the less the model relies on any one variable. Another interesting observation is that there was a small drop in ρ and spatial regression significance values for the larger models. We hypothesize this is because there were so many variables informing the response that the model had plenty of information despite spatial correlation (and the associated information loss we mentioned at the beginning of the spatial regression section).

In the earlier model (with 23 hand-selected predictors), the consensus among the models was the following:

- *Education:* percent with a Bachelor's degree was all positively associated with illegal opioid use; percent high school graduates and percent with some college but no degree was mostly negative
- *Family:* percent of grandparents responsible for grandchildren was all positive
- *Language:* percent with only English spoken at home was mostly positive
- *Ancestry:* percent Dutch and Irish were all positive; percent Norwegian, Ukrainian, and Hungarian were mostly negative

Note: "mostly" positive/negative indicates that there was at least a 5:2 majority in that category among the seven years' models.

6.4 Future Trends and Comparison to First Model

The **predict.sarlm** function in R allows us to apply trends from our initial results to new data. We used it to predict illicit opioid use up to five years in the future for both models. Because we are trying to forecast on the same predictor variables but different observations of the variables in the model (the extrapolated socioeconomic factors), we use the "trend" prediction type [17].

After making predictions for opioid abuse prevalence, we ran these numbers through our six-stage classifier, to make the results comparable to those from our random walk model. The figure below illustrates the deviation between the forecasts of our first model (random walk) and second model (spatial lag).

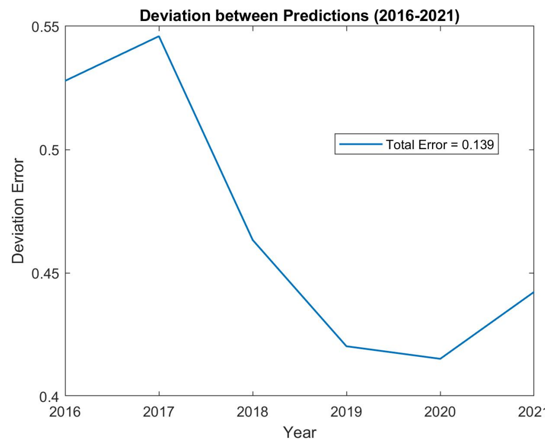


Figure 11: Error between Random Walk and Spatial Regression Predictions

The spatial regression model is heavily dependent on the accuracy of the socioeconomic factors. Because we assume that the SES indicators change linearly over time, the extrapolated indicators are most accurate nearest to the initial data, which means the spatial regression model is most accurate in years not long after 2016. As discussed in previous sections, the random walk has significant error for a few years right after the initial point. Toward the beginning of the simulation, the second model is more accurate while the random walk fails, explaining the discrepancy between them in 2017 and 2018. After a few years, our previous assumptions regarding the socioeconomic indicators make the spatial regression more and more inaccurate; the accuracy of the first model increases as tension between randomness and clustering reaches a balance.

6.4.1 Drug Identification Thresholds

Drug use has been a near constant in the past millenia and will likely continue in the future. With this in mind, we focused on high or increasing levels of drug usage in a county when attempting to identify it as "problematic." However, we could not accurately determine an empirical measure for "high stable" or "increasing" that would hold over time. Instead, we decided to classify a county as problematic if it was in a concering state, i.e. High and Stable, High and Increasing, or Low and Increasing for longer than one and a half years or 24 time steps. Figure 12, below, depicts the predicted number of problematic counties in the next decade.

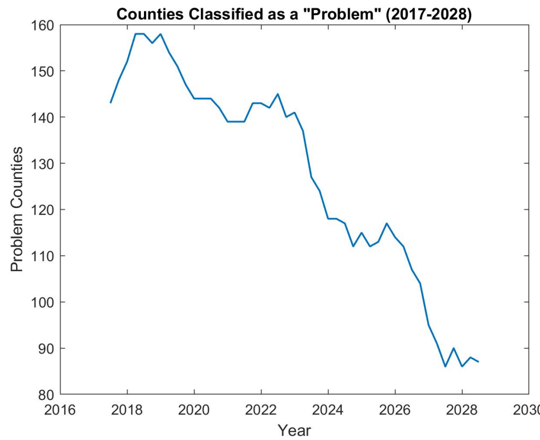


Figure 12: Counties Classified as a ”Problem” (2017-2028)

In a hopeful twist, our model indicates that the number of problematic counties will decrease fairly steadily over the next ten years. However, this may be due to the aforementioned spillover effect. As the crisis diffuses to other areas of the country, the magnitude of the problem may decrease within our study region.

6.4.2 Concerns for the U.S. Government

While continued rise of opioid abuse levels could truly devastate the socioeconomic stability of the nation, some of the more concerning consequences of the epidemic are already occurring. Incarceration levels in the United States are among the highest in the world [40] and with the continuation of the drug war, will continue to climb. Drug addicts are more likely to be arrested than their “clean” counterparts, and most will return to prison within five years of initial release [56]. The opioid epidemic is also forcing many children out of their homes and into the foster care system after losing their parents to addiction. For states low on child welfare money, this means children sleeping in state buildings when there are no more foster homes available [61]. This, lost productivity, and healthcare costs all contribute to the opioid epidemic’s staggering pricetag [47]. Opioid overdose deaths are soaring, and without intervention, they’re predicted to reach 93,000 per year by 2027 [18]. A comparable increase in fiscal cost would mean an annual price of 143 billion dollars.

7 Strategies to Combat the Opioid Epidemic

7.1 Reducing Illicit Drug Use

Results from both our spatial regression and random forest models show that total illicit drug use rate is the most important predictor of opioid use rate in each county. The average results across the seven years of spatial regression models are that, all else constant, a 0.1% reduction in total illicit drug use would reduce total illicit opioid use by 0.052%, which is almost half of the median opioid abuse prevalence rate across counties in this region.

Further reading on our part revealed that reducing overall usage of illicit drugs significantly impacting usage of non-prescribed opioids is documented in the literature. In 1993, the National Research Council (US) Committee on Substance Abuse Prevention Research conducted a landmark review of studies on illicit drug use in the US and identified that very few people try highly addictive illicit substances without first using other less powerful substances, such as marijuana [31]. This and similar findings in the literature indicate that reducing the total amount of drug use in an area would reduce the amount of opioid use, because fewer people would progress through the drug hierarchy.

Recommended ways to address usage of illicit drug use rate include investing in rehabilitation infrastructure in and beyond the criminal justice system [63], reaching out and providing resources to communities who are willing to cooperate with police efforts [37], and improving adolescent education about the addictive properties of illicit substances [31]. The results of our models with regard to ancestry and education indicate that aiming education initiatives at people of all races and socioeconomic statuses is necessary.

7.2 Direct Opioid Interventions

In the early 1970s, the United States declared a war on drugs, increasing the size and presence of federal drug enforcement agencies while halting investigation into the medical efficacy of these substances. Rates of criminal drug abuse have remained constant or increased in the decades since [1]. Launching another war on drugs will not address the opioid epidemic. Instead, we must understand it and treat it accordingly: as a public health crisis, requiring science and health-based solutions, rather than the combative approach used in years past. The first steps in this direction are occurring with limitations placed on opioid prescriptions, but we still have a ways to go. The next steps in this process involve prevention, reducing overdose deaths, and improving addiction care in and beyond the criminal justice system [11].

Implement comprehensive public awareness and education campaigns. A series of campaigns should be addressed to those already abusing opioids to communicate urgent safety concerns, such as the dangers of blood-borne illnesses from sharing needles. Prevention measures directed at adolescents and young adults are particularly important because of their increased susceptibility to addiction [6]. Instruct educators in effective prevention and intervention strategies based in public health science rather than punishment that takes students away from resources, which can inadvertently reinforce the behavior. Use school websites to distribute information about opioid abuse to parents and children.

Reduce availability of and accessibility to opioids. States should expand and encourage the use of Prescription Drug Monitoring Programs (PDMP) to track the prescribing and dispensing of controlled prescription drugs to identify suspected misuse, doctor shopping, or diversion. Programs should be put in place to inform medical professionals about safe prescribing practices for pain management and promote adherance to CDC prescribing guidelines. Around 61% of the drugs that are prescribed in the US each year are not consumed [7]. The remaining medication can be misused or diverted, and flushing them can contaminate water supply. Beginning and expanding medicine take-back programs would decrease the number of these drugs that make it to the streets [11].

Curtail overdose deaths. Get lifesaving opioid overdose antidotes like *naloxene* into the hands of first responders. Administration of the drug has already saved tens of thousands of lives from overdoses [8]. In addition, expedite the distribution of new information about the crisis to stakeholders. This includes news about emerging synthetics and how to treat someone who has been exposed to them.

Improve addiction treatment in and beyond the criminal justice system. Increasing treatment capacity, expanding availability of medication-assisted treatment (MAT) programs, and extending insurance coverage for addiction treatment programs will help people outside of the criminal justice system get the help they need to overcome addiction. Providing adequate recovery support systems will encourage recovering addicts to remain clean. Improving treatment in the criminal justice system will decrease recidivism rates among addicts, three quarters of whom are arrested for another crime within five years of release [56]. To do this, states are encouraged to educate law enforcement officers about addiction as a chronic health condition and to implement and support diversion programs such as Alternative to Incarceration (ATI), which gives individuals in the criminal justice system (CJS) greater access to treatment options. In addition, states should support treatment options for those in the CJS upon reentry to civilian life.

8 Limitations and Sources of Error

No model is without error, least of all models that were devised and implemented in less than 99 hours. What follows are possible sources of error within or weaknesses of our modelling approaches that could be addressed through future exploration.

8.1 Overall

Data Restrictions: A limitation of this report is the data constraint, which reduced our available information to the ACS and NFLIS data provided. While we were able to derive additional variables from the initial data and include map data from outside sources, our exploration was hindered by the restriction to given socioeconomic factors.

Modifiable Areal Unit Problem: The MAUP describes discrepancies in spatial analyses performed at different scales arising because the scale at which one chooses to analyze information or grouping schemes can produce different results [20]. Had we been provided with data at different scales, for instance at the individual level, or aggregated at the census tract or ZIP code levels, our investigation may have produced significantly different results.

Adjacency vs. Distance: In both of our models, we incorporated spatial weighting based on county adjacency rather than physical distance. To the models, it appears that all counties are equally spaced, despite some counties being dramatically different in size and shape. For instance, two counties separated by a tiny county probably share more characteristics than two counties separated by a large county. Changing the type of spatial weighting could significantly alter our results [33].

Linear extrapolation of SES variables: Forecasting opioid abuse rates for 2017-2021 required the extrapolation of socioeconomic indicators beyond the time frame of given data. For simplicity, we assumed a linear trend for each variable. In reality, these trends likely are/will be more varied. Note: Incorporating socioeconomic indicators for future prediction was necessary to compare results between our spatial regression and random walk models. However, for our identification of possible points of origin (in the past), we used the version of our random walk model without socioeconomic factors, because biased socioeconomic factors would have introduced more bias into our model. Because the origin study constituted going back a decade from our first data, we assumed that the spatial regression models would not hold up. The Monte-Carlo aspect of the random walk facilitated this task.

8.2 Model I: The Random Walk

The Markov Property: the defining characteristic of a Markov chain or process is the future states' dependence *only* on the current state. We nominally assumed the Markov property to simulate change in each county's prevalence of opioid abuse. The "nominally" here applies because the selection of a county's class depends on the present class of both it and its neighbors, which in turn depend on the past state(s) of that county. In either case, making this semi-Markovian assumption may be unrealistic in the long-term because longer-term changes, such as policy implementation, are very much indicative of past drug abuse issues in that region, and would drive a county to have more consistent drug abuse prevalence in the long run. This assumption likely reduces some of the validity of our longer-term forecasts.

Linear interpolation of variables in Markov model: Similarly to the extrapolation discussed above, linear interpolation of socioeconomic variables for each "three month" time step in between our years of given data was a convenient but likely inaccurate shortcut. However, the error introduced by the interpolation is much less significant than that introduced by the extrapolation, because we are dealing with time steps in between given data points, and because socioeconomic factors such as percentage of people with certain household structure or ancestry are unlikely to change drastically in the span of one year, whereas they may change noticeably over, say, five years.

Identifying Potential Origin Locations: For year 2000, we initialized all counties to the low stable class. Given the randomness inherent in this model, however, there is a small likelihood of a low stable county jumping to a high increasing, decreasing, or stable class in one time step. This is somewhat unrealistic, and probably contributed to the error in our approximation of the 2010 spatial distribution when we ran these models forward in time.

8.3 Model II: Spatial Regression

Prediction: When forecasting opioid abuse rates for 2017-2021, we applied our spatial lag model from 2016. Although using the most recent model seemed like the best choice given that we had to run a separate spatial regression for each year, the 2016 model was not necessarily representative for 2010-2016. This means that our forecasts were almost certainly biased by the socioeconomic factor data from 2016.

Spatial model tests: Spatial autoregressive and spatial error regressions are asymptotic, that is to say, they give approximately valid inference only for large number of regions [33]. We did not have time to learn about and run the appropriate Monte Carlo hypothesis tests to confirm that the number of counties in our study region was large enough, but having data for approximately 460 counties in each year seemed good enough given other resources discussing statistical significance. If we did not have enough observations given the number of predictor variables, then our regression coefficient estimates were likely inaccurate.

9 Conclusion

The opioid epidemic is a crisis of epic proportions that requires immediate attention from both the public and policymakers. In this report, we characterized the spread of this crisis in the Appalachian region of the United States, in which the epidemic has been most prevalent, using an original random walk model and a suite of spatial regression models. Our models performed well given the data provided and complemented each other. The combination of these models provided useful insights regarding socioeconomic variables associated with the opioid epidemic and future spatial dynamics in this region, which allowed us to make informed recommendations for public policy interventions.

9.1 Future Exploration

With more time and resources, we would like to extend our analysis in the following ways:

- Incorporate more data into the analysis, both variables directly related to drug and opioid use, as well as socioeconomic factors such as income, health care coverage, and disability status.
- Incorporate geography (for example, the locations of major cities and highways) into our random walk model.
- Compare analyses with data at different scales (to test for MAUP), and using distance weighting rather than adjacency between counties. Possible distance weightings to try out include distances between county geographic centroids and between population-weighted centroids. Common approaches in geography include distance as the crow flies and distance on major roads.
- Use actual socioeconomic data from the past to help guide our identification of possible counties of origin, rather than excluding it or linearly extrapolating from our 2010-2016 data.
- Validate our forecasts with 2017, 2018, and future data as it comes along.
- Explore ways to incorporate the influence of long-term trends, for instance public policy, in our model.
- Learn about and try running: a) spatial regression panel models which would allow us to analyze data from all years simultaneously and make predictions based upon more than just one year's worth of data, and b) Monte Carlo hypothesis tests to confirm that the spatial regression models return valid inference given the number of counties and predictor variables we use..
- Investigate the possibility of spatial non-stationarity (different distributions/probabilistic behavior for different sub-regions, such as states or other clusters of counties). Different relationships for different regions might inform different public health intervention strategies in each region.
- Learn about and try running Generalized Linear Models (GLMs) on this data. Linear models are most widely used regression models in spatial statistics. However, Poisson and logistic regression models are most widely used in public health applications, specifically those dealing with count data (e.g. drug arrests). Spatial lag and spatial error models allowed us to confirm that spatial auto-correlation was significant in our data. Knowing that, however, we would like explore the application of spatially-adapted GLMs, which are able to combine the benefits of spatial linear regression and traditional count data models [33]. One promising option is to use the **geoRglm** package in R [35].
- Research more government strategies to holistically address the opioid crisis, emphasizing but not confined to those which reduce overall illicit drug use rates at the county level.

10 Memo to the Chief Administrator

January 29, 2019
To: Uttam Dhillon – Acting Administrator, Drug Enforcement Agency
cc: Dr. Nora D. Volkow – Director, National Institute on Drug Abuse
From: MCM Team #1901679
Subject: Strategies to Combat the Opioid Epidemic

The opioid epidemic claimed 47,000 lives in 2017 [26] and could take 500,000 more before 2027 if not addressed [18]. Our team has analyzed the given data and conceived two models that accurately characterize and predict the spread of the opioid epidemic in Kentucky, Ohio, Pennsylvania, Virginia, and West Virginia. To better understand the crisis and any chances we have of diminishing its effects, we explored current research and investigated the impact of socioeconomic status on opioid abuse within our model.

Results

We utilized two different models to gain a more comprehensive picture of the opioid epidemic. Both characterized addiction as a contagion, which spread outward from an initial source. We determined the likely point of origin for this region is Montgomery County, Kentucky, a rural region that does not have much treatment support. Thus the problem only continued to grow and spilled over into neighboring counties. In the future, we predict the epidemic will continue to spread across the nation, affecting in particular counties that already have a high incidence of illicit drug use. Slowing the spread of this epidemic is of the utmost importance and should be addressed immediately.

Proposal

To mitigate the effects of the opioid epidemic and reduce the number of overdose deaths in the coming years, we propose the following policies:

1. Place restrictions on the prescription of opioids for acute pain without limiting access to those with disabilities or chronic, severe pain. Educate medical professionals on safe prescribing practices and encourage adherence to the CDC's opioid prescription guidelines.

2. Implement targeted education and public awareness campaigns. One set should be aimed at adolescents, who are among the most susceptible to addition, and executed through school systems and websites. Another should address those already misusing opioids, warning them of potential dangers of continued use and promoting treatment programs and clinics near them.

3. Distribute the lifesaving opioid overdose antidote *naloxene* to first responders and potentially to family and friends of those with addictions, and inform them on its use.

4. Extend insurance coverage and treatment capacity for those with addiction issues. Expanding the availability of medication-assisted treatment (MAT) programs that utilize medications like methadone and buprenorphine, especially in rural areas, will also allow struggling individuals to overcome their addiction.

5. Improve addiction treatment in the criminal justice system. This will decrease recidivism rates among addicts, who are more likely to get arrested and three quarters of whom will be arrested for another crime within five years of their release [56]. Educate law enforcement officers about addiction as a chronic health condition, and support continued treatment upon return to civilian life.

References

[1] *A Brief History of the Drug War*. Online. 2018. URL: drugpolicy.org/issues/brief-history-drug-war.

[2] Tom Abate. "Researcher's model could help stem opioid crisis". In: *Stanford Medicine News Center* (2018). Online. URL: med.stanford.edu/news/all-news/2018/08/researchers-model-could-help-stem-opioid-crisis.html.

[3] Substance Abuse, Center for Behavioral Health Statistics Mental Health Services Administration, and Quality. *The TEDS Report: Age of Substance Use Initiation among Treatment Admissions Aged 18 to 30*. Online. Rockville, MD: SAMHSA, 2014. URL: samhsa.gov/data/sites/default/files/WebFiles_TEDS_SR142_AgeatInit_07-10-14/TEDS-SR142-AgeatInit-2014.htm.

[4] AddictionCenter, ed. *Codeine Addiction and Abuse*. Online. 2018. URL: addictioncenter.com/opiates/codeine/.

[5] AddictionCenter, ed. *Methadone Addiction and Abuse*. Online. 2018. URL: addictioncenter.com/opiates/methadone/.

[6] *Adolescent Substance Use: America's #1 Public Health Problem*. Online. 2011. URL: eric.ed.gov/?id=ED521379.

[7] A. Kennedy-Hendricks et al. "Medication Sharing, Storage, and Disposal Practices for Opioid Medications Among U.S. Adults". In: *Journal of the American Medical Association* 176.7 (2016). Online., pp. 1027–1029. URL: jamanetwork.com/journals/jamainternalmedicine/fullarticle/2527388.

[8] E. Wheeler et al. "Opioid Overdose Prevention Programs Providing Naloxene to Layperons: United states, 2014". In: *Morbidity and Mortality Weekly Report* 64.23 (2015). Online., pp. 631–635. URL: cdc.gov/mmwr/preview/mmwrhtml/mm6423a2.htm.

[9] Jordan M. Harrison et al. "Trends in Prescription Pain Medication Use by Race/Ethnicity Among US Adults with Noncancer Pain, 2000-2015". In: *American Journal of Public Health* (2018). DOI: 10.2105/AJPH.2018.304349.

[10] Katherine M. Keyes et al. "Understanding the Rural–Urban Differences in Nonmedical Prescription Opioid Use and Abuse in the United States". In: *American Journal of Public Health* 104.2 (). Online., pp. 52–59. URL: ncbi.nlm.nih.gov/pmc/articles/PMC3935688/.

[11] Lee C. Bollinger et al. *Ending the Opioid Crisis: A Practical Guide for Policymakers*. Online. Center on Addiction, 2017. URL: centeronaddiction.org/addiction-research/reports/ending-opioid-crisis-practical-guide-state-policymakers.

[12] Tara A. Macey et al. "Patterns and Correlates of Prescription Opioid Use in OEF/OIF Veterans with Chronic Noncancer Pain". In: *Pain Medicine* 12.10 (2011). Online., pp. 1502–1509. URL: academic.oup.com/painmedicine/article/12/10/1502/1890236.

[13] Ver Hoef et al. "Spatial autoregressive models for statistical inference from ecological data". In: *Ecological Monographs (Ecological Society of America)* (2017). Online. URL: esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ecm.1283.

[14] Luc Anselin. *An Introduction to Spatial Regression Analysis in R*. Technical report. Online. University of Illinois, Urbana-Champaign, 2003. URL: labs.bio.unc.edu/buckley/documents/anselinintrospatregres.pdf.

[15] Monica J. Alexander Mathew V. Jiang Magali Barbieri. "Trends in Black and White Opioid Mortality in the United States, 1979–2015". In: *Epidemiology* 29.5 (2018). Online., pp. 707–715. URL: journals.lww.com/epidem/Fulltext/2018/09000/Trends_in_Black_and_White_Opioid_Mortality_in_the.16.aspx.

[16] Alex Berezow. *White Overdose Deaths 50% Higher Than Blacks, 167% Higher Than Hispanics*. Online. 2018. URL: acsh.org/news/2018/04/05/white-overdose-deaths-50-higher-blacks-167-higher-hispanics-12804.

[17] Roger Bivand. *predict.sarlm: Prediction For Spatial Simultaneous Autoregressive Linear Model Objects, From spdep v0.8-1*. Online. URL: https://www.rdocumentation.org/packages/spdep/versions/0.8-1/topics/predict.sarlm.

[18] Max Blau. "STAT forecast: Opioids could kill nearly 500,000 Americans in the next decade". In: *STAT News* (2017). Online. URL: statnews.com/2017/06/27/opioid-deaths-forecast/.

[19] Eric Bonabeau. "Agent-based Modeling: Methods and Techniques for Simulating Human Systems". In: *Proceedings of the National Academy of Sciences of the United States of America* 99.3 (2002). Online., pp. 7280–7287. URL: pnas.org/content/99/suppl_3/7280.

[20] Michael Emch Elisabeth Dowling Root Margaret Carrel. "Health and Medical Geography". In: 4th ed. New York: The Guilford Press, 2017.

[21] *Cartographic Boundary Shapefiles - Counties 20m*. Online. 2017. URL: census.gov/geo/maps-data/data/cbf/cbf_counties.html.

[22] John Corbett. *Torsten Hägerstrand: Time Geography*. Online. Masaryk University, 2006. URL: is.muni.cz/el/1431/jaro2006/Z0147/time_geography.doc.

[23] Ruth C. Fong Walter J. Scheirer David D. Cox. "Using Human Brain Activity to Guide Machine Learning". In: *Scientific Reports* 8.5397 (2018). Online. URL: nature.com/articles/s41598-018-23618-6.

[24] Teresa A. Rummins M. Caroline Burton Nancy L. Dawson. "How Good Intentions Contributed to Bad Outcomes". In: *Mayo Clinic Proceedings* 93.3 (2018). Online., pp. 344–350. URL: mayoclinicproceedings.org/article/S0025-6196(17)30923-0/fulltext.

[25] *Detecting Multicollinearity Using Variance Inflation Factors*. Online. 2018. URL: https://newonlinecourses.science.psu.edu/stat501/node/347/.

[26] Centers for Disease Control, ed. *Prescription Opioid Data*. Online. The United States Government, 2019. URL: cdc.gov/drugoverdose/data.

[27] Fred Brauer Pauline van den Driessche J. Wu. "Compartmental Models in Epidemiology". In: *Mathematical Epidemiology*. Online. Springer-Verlag Berlin Heidelberg, 2008. URL: springer.com/cda/content/document/cda_downloaddocument/9783540789109-c1.pdf?SGWID=0-0-45-532715-p173817706.

[28] National Insitute on Drug Abuse, ed. *Abuse of Prescription (Rx) Drugs Affects Young Adults Most*. Online. 2018. URL: drugabuse.gov/related-topics/trends-statistics/infographics/abuse-prescription-rx-drugs-affects-young-adults-most.

[29] National Insitute on Drug Abuse, ed. *Sex and Gender Differences in Substance Use*. Online. 2018. URL: drugabuse.gov/publications/research-reports/substance-use-in-women/sex-gender-differences-in-substance-use.

[30] Richard Durrett. "Essentials of Stochastic Processes". In: 3rd ed. Switzerland: Springer International Publishing, 2016.

[31] Gerstein DR Green LW editors. *Preventing Drug Abuse: What do we know?* Online. National Research Council (US) Committee on Substance Abuse Prevention Research, 1993. URL: ncbi.nlm.nih.gov/books/NBK234579/.

[32] Rose A. Rudd Noah Aleshire Jon E. Zibbell R. Matthew Gladden. "Increases in Drug and Opioid Overdose Deaths — United States, 2000–2014". In: *Morbidity and Mortality Weekly Report (MMWR)* 64.50 (2016). Online., pp. 1378–1382. URL: cdc.gov/mmwr/preview/mmwrhtml/mm6450a3.htm.

[33] Lance A. Waller Carol A. Gotway. *Applied Spatial Statistics for Public Health Data*. Online. John Wiley & Sons, Inc., 2004. Chap. 9: Linking Spatial Exposure Data to Health Events. URL: onlinelibrary.wiley.com/doi/book/10.1002/0471662682.

[34] M. Greenwood. "On the Statistical Measure of Infectiousness". In: *Epidemiology and Infection* 31.3 (1931). Online., pp. 336–351. URL: cambridge.org/core/journals/epidemiology-and-infection.

[35] Ole F. Christensen Paulo J. Ribeiro Jr. *geoRglm: a package for generalised linear spatial models introductory session*. Online. 2017. URL: cran.r-project.org/web/packages/geoRglm/vignettes/geoRglmintro.pdf.

[36] Tamarra Kemsley. *Opioid Addiction: One More Way We're Failing our Veterans*. Talbott Recovery, 2018. URL: talbottcampus.com.

[37] Mark H. Moore Mark A.R. Kleiman. *The Police and Drugs*. Online. U.S. Department of Justice, Office of Justice Programs, 1989. URL: ncjrs.gov/pdffiles1/nij/117447.pdf.

[38] Brian Klinkenberg. *Spatial Autocorrelation*. Online. 2019. URL: ibis.geog.ubc.ca/courses/geob479/notes/spatial_analysis/spatial_autocorrelation.htm.

[39] Francois Lamy. *eDarkTrends update: New Synthetic Opioids ... Identified on Two Cryptomarkets*. Online. 2017. URL: ndews.umd.edu/sites/ndews.umd.edu/files/pubs/.

[40] Michelle Ye Hee Lee. "Yes, U.S.locks up people at a higher rate than any other country". In: *The Washington Post* (2015). Online. URL: washingtonpost.com/news/fact-checker/wp/2015/07/07/yes-u-s-locks-people-up-at-a-higher-rate-than-any-other-country/?noredirect=on&utm_term=.202a0bf0c748.

[41] *Likelihood Ratio Tests*. Tech. rep. Online. Colorado State University. URL: https://sites.warnercnr.colostate.edu/gwhite/wp-content/uploads/sites/73/2017/04/LikelihoodRatioTests.pdf.

[42] *Merriam-Webster Dictionary*. Online. Merriam-Webster, Inc, 2016. URL: merriam-webster.com/dictionary/.

[43] *National Forensic Laboratory Information System Subset*. Online. 2010-2017. URL: comap.com/undergraduate/contests/mcm/contests/2019/problems/.

[44] Independent Living National Institute on Disability and Rehabilitation Research. *Request for Information: Opioid Use Disorder Among People with Disabilities*. Online. 2018. URL: acl.gov/sites/default/files/about-acl/2018-01/Final_RFI_Opioid_Use_Disorder_PwD_Jan2018.pdf.

[45] NIDA. *Understanding Drug Use and Addiction*. Online. The United States Government, 2018. URL: drugabuse.gov/publications/drugfacts/understanding-drug-use-addiction.

[46] *Opioid Crisis*. Online. Health Resources & Services Administration, 2019. URL: hrsa.gov/opioids.

[47] *Opioid Overdose Crisis*. Online. National Institute on Drug Abuse, 2019. URL: drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis.

[48] *Overdose Death Rates*. Online. 2018. URL: drugabuse.gov/related-topics/trends-statistics/overdose-death-rates.

[49] "Prescription painkiller overdoses at epidemic levels". In: *CDC Newsroom* (2011). Online. URL: cdc.gov/media/releases/2011/p1101_flu_pain_killer_overdose.html.

[50] RecoveryFirst, ed. *Why Veterans are Prone to Opioid Abuse?* Online. 2018. URL: recoveryfirst.org/veterans/opioid-abuse/.

[51] *Results from the 2017 National Survey on Drug Use and Health*. Substance Abuse, Mental Health Services Administration Center for Behavioral Health Statistics, and Quality, 2018.

[52] Josh Katz Margot Sanger-Katz. "'The Numbers are So Staggering.' Overdose Deaths Set a Record Last Year". In: *New York Times* (2018). Online. URL: nytimes.com/interactive/2018/11/29/upshot/fentanyl-drug-overdose-deaths.html.

[53] Paola Scommegna. *Opioid Overdose Epidemic Hits Hardest for the Least Educated*. Online. 2018. URL: prb.org/people-and-places-hardest-hit-by-the-drug-overdose-epidemic.

[54] BayMark Health Services. *Vermont's Opioid Addiction: A Family Crisis*. Online. 2018. URL: baartprograms.com/vermonts-opioid-addiction-a-family-crisis/.

[55] Knvul Sheikh. "Why Do We Get Addicted to Things?" In: *LiveScience* (2017). Online. URL: livescience.com/60694-why-do-we-get-addicted.html.

[56] M.R. Durose A.D. Cooper H.N. Snyder. *Recidivism of Prisoners Released in 30 States in 2005: Patterns from 2005 to 2010*. Online. U.S. Department of Justice, 2014. URL: bjs.gov/content/pub/pdf/rprts05p0510.pdf.

[57] Corey S. Sparks. *Spatial Regression Models*. Online. 2015. URL: rpubs.com/corey_sparks/109650.

[58] Seth Spielman. *Spatial Analysis and Regression*. Online. 2015. URL: sethspielman.org/courses/geog5023/lectures/Lecture_8_2015.pdf.

[59] *The Opioid Crisis*. Online. The United States Government, 2018. URL: whitehouse.gov/opioids/.

[60] *The Opioid Crisis*. Online. 2018. URL: theopioidcrisis.com.

[61] *The Opioid Epidemic's Effect on Children*. Online. 2016. URL: goodencenter.org/opioid-addiction-treatment/.

[62] Philip A. Viton. *Notes on Spatial Econometric Models*. Online. 2010. URL: pdfs.semanticscholar.org/64ab/4ec3a6cb25cb191818c5d65400e6c3697082.pdf.

[63] Redonna K. Chandler Bennett W. Fletcher Nora D. Volkow. "Treating Drug Abuse and Addiction in the Criminal Justice System: Improving Public Health and Safety". In: *Journal of the American Medical Association* (2009). Online. URL: ncbi.nlm.nih.gov/pmc/articles/PMC2681083/.

A Appendix

A.1 Predictor Variables for the 2014 Spatial Lag Model

Predictor Variable	Total Impact
All Illicit Drugs Rate	0.5323845
Ancestry: French Canadian	0.0006989
Lived Abroad Last Year	-0.0004824
Ancestry: Danish	-0.0003967
Ancestry: Hungarian	-0.0003788
Ancestry: Greek	-0.0002901
Ancestry: Welsh	-0.0002611
Ancestry: Czech	0.0002339
Ancestry: Swedish	-0.0001862
Ancestry: Ukrainian	-0.0001733
Ancestry: Scottish	0.0001691
Ancestry: Slovak	0.0001519
Asian/Pacific Language(s) at Home	0.0001320
Separated (from Marriage)	0.0001237
Ancestry: Scotch-Irish	-0.0001182
Ancestry: Swiss	-0.0001170
Ancestry: Arab	0.0000909
Ancestry: Irish	0.0000820
Ancestry: Subsaharan African	-0.0000766
Ancestry: West Indian, Excluding Hispanic Origin	0.0000736
Ancestry: French-Basque	-0.0000716
Other Indo-Euro Language(s) at Home	0.0000627
Households With Nonrelatives	-0.0000615
Ancestry: Dutch	-0.0000588
Bachelor's Degree	0.0000535
Ancestry: Polish	0.0000515
Ancestry: Norwegian	-0.0000493
Widow	0.0000448
Enrolled in Kindergarten	0.0000347
Living with Unmarried Partner	0.0000342

Predictor Variable	Total Impact
Lived in Different House in U.S. Last Year	0.0000323
Ancestry: Portuguese	-0.0000318
High School Graduate (or Equivalent)	0.0000316
Ancestry: Russian	-0.0000291
Enrolled in Grades 9-12	-0.0000283
Never Married	0.0000250
Ancestry: American	0.0000237
Ancestry: English	0.0000224
Spanish Spoken at Home	-0.0000224
Grandparents Responsible for Grandchildren	0.0000211
Single Male with Children Under 18	-0.0000203
Enrolled in Nursery/Preschool	0.0000180
Civilian Veterans	0.0000143
Graduate/Professional Degree	0.0000125
Ancestry: Lithuanian	-0.0000112
Ancestry: Italian	-0.0000083
Ancestry: German	0.0000074
Single Female with Children Under 18	-0.0000058
Divorced	0.0000024
Enrolled in Grades 1-8	-0.0000015
Person 65+ Living Alone	-0.0000010
FIPS Code	0.0000000
Total Population	0.0000000