Modeling the Opioid Epidemic

Andrew Mashhadi Melissa Lee Rishab Rajagopal Bryce Palmer-Toy

December 16, 2019

Summary

The United States is facing a severe opioid crisis that must be reined in before it is too late. If left unchecked, it could have a lasting impact on the country's economy and society. We modelled the spread of synthetic and non-synthetic opioid use in and between 5 different states — West Virginia, Virginia, Kentucky, Ohio, and Pennsylvania — using recurrence equations for each county in the state using data from the National Forensic Laboratory Information System (NFLIS) for the years 2010-2016. This condenses each county's opioid problem into a simple model that can be studied to come up with countermeasures and predict future epidemics. Additionally, the statewide recurrence models can be obtained by averaging the recurrence coefficients of all counties contained within the state.

This recurrence equation takes into account the influence of other counties, influence within the county, the county's mortality rate, and possible recovery rates for addicted individuals. The spherical distances between counties (computed using the Haversine function) and these values, along with the provided opioid reports and the area of counties were all used to calculate influence values that eventually interpreted potential growth factors within a county from one year to the next. Additionally, the growth factor of each county's recurrence equation led us to find which counties are self-correcting and which counties will have an increase in future opioid use. In this case, we define a self-correcting county as a county that possesses a growth factor under 1, which implies that the opioid count will reach zero over time. We selected counties with the highest recurrence coefficient and analysed the growth of opioid reports in the surrounding counties to determine possible origins of the heroin epidemic. Using this method, we singled out Perry, Pennsylvania and and Fairfax, Virginia as potential epicenters of the heroin epidemic. Based on our model's projections, we define a state/county with one addiction per mile^2 to be in a "state of emergency" and we calculated exactly how long it would take each state to reach this critical level. Based on our model, it was determined that Pennsylvania is the most alarming, as our calculations show it is already in a state of emergency. However, we found that the other four states have at least 1-2 decades before reaching this critical level of opioid abuse.

To analyze the effects of various socioeconomic factors on the drug use within the county, we obtained the coefficient of determination for each factor. We found that high numbers of veterans and associate degree holders are positively correlated with a higher opioid reports in each county, and therefore our model suggests to us that we should be concerned about the number of associates degree holders and the number of veterans within a county. Under the assumption that a larger population of veterans and associates degree holders implies a greater amount of opioid addictions, we modified our original countywide recurrence equations to also consider the effects from the veteran and associate degree holders population by incorporating a logistic function that increases/decreases the central influence value accordingly. We reconciled the correlations between the various socioeconomic factors and the drug counts with credible sources, which partially justifies our assumption that the correlations imply causation in this case. For possible countermeasures to this crisis, we connected our model with outside sources and research from those in other fields and found potential strategies that involve discounted housing for veterans, marriage counselling for couples experiencing marital strife, and affordable rehabilitation for addicted individuals.

Introduction

Opioid abuse is a scourge that affects the entire world. It is a matter of foremost concern for our country since it can affect every one of its citizens, taking the lives of over 47,000 people every year in the United States alone [Nida]. Further, the spread of opioid abuse through the population can have an adverse impact on the country's workforce and society. The Centers for Disease Control and Prevention report that prescription opioid misuse costs the United States $78.5 billion every year [Nida]. Tackling this problem is an intimidating task; however, it can be made simpler using mathematical modeling.

Problem Statement

Enforcing rules and laws regarding opioid use is difficult in a country that is as large and populous as the United States. Therefore, it should be broken down into more manageable pieces. The spread of opioid use in the five states must be modeled and then analyzed to predict how long each state has until it reaches a critical level of addiction. Further, using the data from the U.S. Census Bureau, we have to determine which socioeconomic factors are correlated with drug use. Lastly, based on our findings and outside research, we suggest solutions to the burgeoning opioid problem, and alter the model accordingly.

Part I

Data Sources

Before describing our unadjusted model, it is important to note that in its creation and derivation, we make use of only the data provided to us by the DEA/National Forensic Laboratory Information System (NFLIS), and additional data from the U.S. Census Bureau from the years 2010-2016. Most importantly, we use the NFLIS data to gather the drug identification counts in years 2010-2016 for narcotic analgesics (synthetic opioids) and heroin in each of the counties from the the five U.S. states Ohio, Kentucky, West Virginia, Virginia, and Pennsylvania and we use the U.S. Census Bureau data to gather information about each county's population size and other socioeconomic factors for us to use later in our adjusted model.

We also used the county names provided in the data to match each county with a latitude, longitude coordinate and to gather county sizes (area in mi^2), from the U.S. Census Bureau.

Model Derivation

When we began discussing potential ideas for a mathematical model to describe the spread and characteristics of the reported synthetic opioid and heroin incidents, we knew we needed a mathematical way to quantify and measure the influence an entire county might have on an individual to encourage them to start using opioids. This measure must take into account many potential forms of influence. For example, we might consider the influence from advertisements, current opioid addicts, and even some medical professionals. Age of the population also plays a huge role in our influence factor since teenagers and young adults could be more impressionable than other age groups. Clearly, if we can find a method for quantifying this "influence factor", we would have a value to help construct our model of the geographical spread in and between the five states.

Influence Factor Assumptions

Before we find the influence values to assign to each county, we must make the following assumptions based on the data we are given, and for simplicity:

- The reported opioid counts from the NFLIS data make up all of the opioids in the county. So opioids that have not been found yet are disregarded.
- A single opioid count represents one addict in the county's population. Therefore, if multiple counts of opioids are found and reported, we will assume there were multiple addicts at that location.
- We will assume that any county with a missing opioid entry has had zero counts of that drug for that year.
- The influence from each county will be considered uniformly distributed across the entire county. That is, the influence "value" at different points within a county should have the same magnitude, even if the county is large and the points are at the maximum distance away from each other in the same county.
- We will assume that any decrease in population is a result of opioid related deaths. That is, if there are n deaths, we will assume that n opioid addicts have died.
- We will make the very safe assumption that all newborn infants are not addicted to opioids. Therefore, an increase of n people in the population should not affect the number of opioid addictions.
- We do not consider immigration and emigration between counties. Note, we do consider the effect of other county's population, we just assume they do not make up the living population in that county.
- When we refer to the "center" or a county or state, what we are really looking at is the centroid. So given a planar figure, this is defined to be the point where you would have to place a fulcrum so that the flat plate made by the shape would balance perfectly. We assume that this definition of center allows for the best approximations of distances between counties and states.
- We assume that the more opioid users currently in the county or neighboring counties, the more susceptible an unaffected individual is to opioid abuse. However, we only consider the influence from the neighboring counties that exist within Ohio, Kentucky, West Virginia, Virginia, and Pennsylvania.

Now we begin deriving each county's "influence values". Say we want to find the influence values for the i^{th} county. We first find the influence that county i has on its own population. Based on our last assumption written above, in a given county, we know that the risk of an individual developing an addiction to opioids increases with the proportion of opioid use in that county, so we begin using our data to find an approximate proportion of opioid use in each county. To do this, we use NFLIS data to find the sum of all the reported synthetic opioid and heroin counts for the year 2010, c_i^{2010}, then we use the U.S. Census Bureau data to find the population in the year 2010 by multiplying the i^{th} county's total number of households, n_i^{2010}, by the average household size, h_i^{2010}. We then continue this same process for each subsequent year up to and including the year 2016. Next we obtained the average of the i^{th} county's drug counts per year, c_i, and population level, P_i, for all seven years. That is,

$$c_i = \frac{\sum_{j=2010}^{2016} c_i^j}{7 \text{ years}} \tag{1}$$

$$P_i = \frac{\sum_{j=2010}^{2016} h_i^j n_i^j}{7 \text{ years}} \tag{2}$$

So we define the opioid use influence factor that county i has on itself, or the *central influence value*, I_i, as

$$(\text{central influence value}) = I_i = \frac{c_i}{P_i} \tag{3}$$

Notice equation (3) is a unitless quantity because we have made the assumption that a single drug count is equivalent to one opioid addiction and therefore the units of population over time will be canceled out. This will eventually play a large role in our model, as each county's influence value will act as a substantial part of a county's growth factor.

Now we shift our focus on the opioid use influence from the outside counties of county i, the *outside influence value*. To make use of all of our NFLIS data, we want to account for every other county's opioid use influence on county i. However, we know a county that is closer to the i^{th} county will have a stronger influence than a county that is farther away. After discussing many methods to account for this issue, we decided to use a familiar idea brought to us from the natural world. In physics, there are many cases where a point source which spreads its influence equally in all directions without a limit to its range will obey the *inverse square law* [Phys]. Point sources of gravitational forces, electric fields, light, sounds, and even radiation obey this inverse square law. So if we consider the i^{th} county as a "point source" we can account for this diminishing influence value as a function of actual distance from each county. Therefore, we define the j^{th} county's influence value on county i, I_i^j,to be inversely proportional to the distance from the center of county i to the center of county j, d_i^j, squared. That is,

$$I_i^j \propto \frac{1}{(d_i^j)^2} \qquad \text{for } i \neq j \tag{4}$$

In addition to this proportionality we have placed, we should account for two more very important quantities that county j possesses. We know that the j^{th} county has its own *central influence value*, so we multiply the right hand side of equation (4) by the j^{th} county's central influence value, I_j. This way, when the j^{th} county's central influence value is very large, even if it is further away from the i^{th} county, it may still have a larger impact on the i^{th} county's opioid use than a closer county with a smaller central influence value. Conversely, if two outside counties have a similar central influence value, the closer of the two counties will have a substantially larger effect on the central county's opioid use. In addition to multiplying the j^{th} county's central influence value, we also multiply the j^{th} county's total area, A_j. This does two things:

1. Since our model will likely dilute the influence values from larger counties (because their centroids will be farther from the other centroids by virtue of the county's size), it helps us

to quantify any additional drug influences from the actual geographical size of county j. For example, now consider having two counties surrounding the i^{th} county with equal central influence values. If one of those two counties is much larger than county i, and the other outside county is considerably smaller than county i, then the outside influence from the larger county would be much lower than the outside influence from the smaller county due to its center being further away. To help with this issue, we multiply the outside influence value by the outside county's corresponding area. This way, we can counteract the dilution of the larger neighboring county's influence values by greatly increasing the outside influence values for larger neighboring counties and only mildly increasing the outside influence values for smaller neighboring counties.

2. For convenience, we want the j^{th} county's influence value on county i, I_i^j, to also be unitless. This is because our model will eventually be using both the outside and central influence values as part of a growth factor. Since the squared distance from the center of county i to the center of county j, $(d_i^j)^2$ has units of $(\text{length})^2$ and the j^{th} county's central influence value, I_j, is already unitless, multiplying by the area of the j^{th} county cancels out the $(\text{length})^2$ units in the denominator from the $(d_i^j)^2$ term, and makes sure that I_i^j also remains unitless.

We are now left with the following definition of the j^{th} county's influence value on county i, I_i^j,

$$I_i^j = \frac{I_j A_j}{(d_i^j)^2} \qquad \text{for } i \neq j \tag{5}$$

Recall that the I_j is the j^{th} county's central influence value shown in equation (3) above. Since we want to consider *every* other county's effective opioid use influence on county i, for county i we sum up every outside county's influence value to obtain,

$$\text{total outside influence value} = I_i^O = \sum_{\substack{j=1 \\ j \neq i}}^{n} I_i^j \tag{6}$$

Using our new central and outside opioid use influence values, we can now begin developing our model. Our model will be using a recurrence equation to describe the spread and characteristics of the reported synthetic opioid and heroin incidents in and between the five states mentioned above. For each individual county, our recurrence equations $R_i(t)$ will tell us the number of opioid addictions per area, $(\frac{\text{addicts}}{\text{area}})$, for the specified year. Therefore, we will be using a discrete time step of $\Delta t = 1$ year in our actual calculations. For the i^{th} county, the recurrence equation will look like the following,

$$R_i(t + \Delta t) = (1 + (kI_i + I_i^O))R_i(t) - (m_i + r)\Delta t R_i(t) \tag{7}$$

where

- m_i is the mortality rate
- r is the opioid addiction rehabilitation/recovery rate
- k is our internal influence proportionality constant
- $R_i(2010)$ is our initial value starting at the year $t = 2010$

It is important to note that the unique influence values used in each county's corresponding recurrence equation is time independent. That is, the influence value and the growth factor for a county will remain fixed for the entire simulation and will *not* be altered when the drug counts in neighboring cities change each year.

A Brief Mention of Data Simplification

Before describing the process of finding the above values in our recurrence equation, we must briefly mention what type of data cleaning methods and simplifications we have made to the NFLIS data in order for us to accurately measure these quantities. First, for every city reported in the data, we added its opioid counts to its corresponding and/or surrounding county, but did not add its population to the county unless the city entirely makes up its own county, or if the city was independent. The reasoning of this is to make sure we do not add a city's population if it is already part of a corresponding county (which most are), as that would double count the population for everyone living in the cities. Second, for each county, we have summed up *all* of the synthetic, semi-synthetic, and non-synthetic opioids mentioned in the data, and have avoided using the total drugs given in a column provided, as we are unsure of what other drugs have been included in those drug counts. Third, we set the opioid count values equal to zero for any opioids listed, but not reported.

Finding m_i, r, k, and $R_i(2010)$

The methods used to find the mortality factor/rate, m_i, for each county were very straight forward. However, much of the following calculations are based under the assumption that all deaths in a county are from opioid use. Thus, for each individual county we find the difference in population levels between the years 2010-2011 and divide by the earlier population level, then we repeat the same process for the years 2011-2012, and so on. After determining all six of these values, we then average them up to obtain our mortality rate, m_i (note the dimensions of m_i are $(\text{time})^{-1}$ because it is a per year rate). Recall we are working under the assumptions that no immigration occurs and that newborns cannot be addicted to opioids, so if the average is negative ($m_i < 0$) then we set $m_i = 0$, since an increase in the population is assumed to have no effect on the number of opioid addictions per area.

The value r is a parameter we will use to factor in the effects of opioid addiction recovery. That is, for all counties, r will represent the per year recovery rate. Based on the data we have, we cannot include an addiction recovery rate, so all of the proceeding calculations in Part I of this paper will be using a recovery rate of zero $r = 0$. However, it should be noted that a recovery rate can greatly affect the spread of opioid use in a county. Since we are using a recurrence equation, a slight increase in r could greatly reduce projected drug use levels. We will examine the effects of modifying r later on in Part III. For now, it is simply enough to notice that a larger r could make our growth factor small enough to result in an elimination of opioid abuse throughout a given county.

Since each county's recurrence equation models its own county's opioid addictions per area, it is important to place an emphasis on the central influence value for that individual county, so we multiply only the central influence value by an internal influence proportionality constant, $k = 10$. We chose to make $k = 10$ here by comparing our model's 2011-2016 predicted opioid use per area values with the actual 2011-2016 values opioid use per area values (from the NFLIS data) with different constants $k > 1$ values. We eventually found that $k = 10$ gave us the most accurate projections in those years. Now the largest increase in the coefficient of the i^{th} county's recurrence equation will be made up of mostly its own influence value, and only slightly modified by the outside influences. This is what we want, as outside influence should be less of a factor than internal influence. It is important to note that although k is set to a constant here, in Part II, k will be modified to become an actual function of other variables introduced later on.

Lastly, we find $R_i(2010)$ using the 2010 data provided in NFLIS. For the i^{th} county, we divide the total opioid related drug counts by the i^{th} county's total area.

Computational Methods

In order to proceed with our recurrence equation model, (7), we must compute our influence values, I_i and I_i^O, for every city, and then combine them with the other necessary values gathered from the data (k and m_i). Recall from equation (5) that in order to find the j^{th} county's influence on the i^{th} county, we must divide the area of county j by the distance from county i to county j. With only the provided data, we were unable to efficiently record the standard planer Euclidean distances between every pair of counties. Therefore, to find the distances between every county, we first had to assign each county with its corresponding latitude and longitude coordinates using the county's FIPS code from our data. For the counties listed in the simplified data, we used a MapBox Geocoding API [Mpbx] to gather the approximate latitude and longitude coordinates. Using these coordinates, we are essentially offered two ways to approximate the distances between the counties. First, we could use the standard Euclidean distance formula in Equation (11), also referred to as Chordal distance,

$$\Delta X = cos\phi_2 cos\lambda_2 - cos\phi_1 cos\lambda_1 \tag{8}$$

$$\Delta Y = cos\phi_2 sin\lambda_2 - cos\phi_1 sin\lambda_1 \tag{9}$$

$$\Delta Z = sin\phi_2 - sin\phi_1 \tag{10}$$

$$C = \sqrt{(\Delta X)^2 + (\Delta Y)^2 + (\Delta Z)^2} \tag{11}$$

where λ_1, ϕ_1 and λ_2, ϕ_2, are the geographical longitude and latitude coordinates in radians. This first method is relatively simple and familiar; however, this could lead to some very troubling

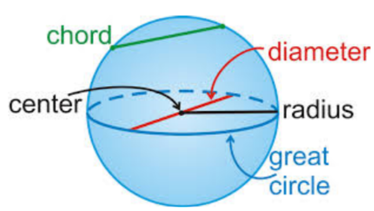

Figure 1: *Chordal Distance Example*

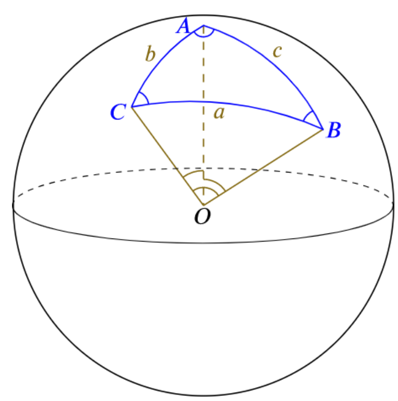

Figure 2: *Spherical Distance Example*

inaccuracies. The problem is that the chordal distance C represents the distance between two counties only if we were to tunnel through the earths surface (see Figure 1), so it should really only be used to approximate distances less than a couple hundred miles apart.

Since we are dealing with some relatively large distances we decided to use the Haversine formula for the most accurate computations. This formula calculates spherical distance between the two locations using the great circle passing through both points on the sphere. To see why the distances are more accurate, see Figure 2 above. Using simple trigonometric manipulations and identities along with our original Chordal Distance formula, we can derive the following Haversine formula for the spherical distance between two points on a sphere,

$$D = 2R\arcsin\sqrt{(\sin^2(\frac{\phi_2-\phi_1}{2})) + cos\phi_1 cos\phi_2(\sin^2(\frac{\lambda_2-\lambda_1}{2}))} \tag{12}$$

where R is the sphere's radius and λ_1, ϕ_1 and λ_2, ϕ_2 are still the geographical longitude and latitude coordinates in radians.

Therefore, we have found an accurate method to calculate the distances between all of the counties. So we used this formula to create a function in C++ to call whenever we need a quick distance between two different counties. The code for this function can be seen in Figure 3.

```cpp
double haversine(double lat1, double lon1, double lat2, double lon2)
{
    // distance between latitudes
    // and longitudes
    double dLat = (lat2 - lat1) * 3.1415 / 180.0;
    double dLon = (lon2 - lon1) * 3.1415 / 180.0;

    // convert to radians
    lat1 = (lat1) * 3.1415 / 180.0;
    lat2 = (lat2) * 3.1415 / 180.0;

    // apply formulae
    double a = pow(sin(dLat / 2), 2) + pow(sin(dLon / 2), 2) * cos(lat1) * cos(lat2);
    double R = 3958.8;
    double c = 2 * asin(sqrt(a));
    return R * c;
}
```

Figure 3: *C++ Haversine Function*

```cpp
void calc_out_influence(std::vector<County>& Counties)
{
    for (size_t j = 0; j < Counties.size(); ++j)
    {
        double jth_lat = Counties[j].lat;
        double jth_lon = Counties[j].lon;
        for (size_t k = 0; k < Counties.size(); ++k)
        {
            if (k == j)
                continue;

            double num = (Counties[k].I * Counties[k].area);
            double den = (haversine(Counties[k].lat, Counties[k].lon, jth_lat, jth_lon));
            Counties[j].I_o += (num / (den * den));

        }
    }
}
```

Figure 4: *Outside Influence Value Function*

After quickly gathering each county's area from an ArcMap Geocoding API [Armp] (also used to check our longitude and latitude coordinates), we then created an organized independent file containing all of the necessary simplified pieces of data. That is, for each county we had the name, location coordinates, area A_i, mortality rate m_i, and the average opioid counts per population (central influence values I_j). Using C++, we read all of this data from the single organized file, filled up a vector holding each county's independent information, and calculated our outside influence values using the function shown in Figure 4. Notice the code in Figure 4 calls the Haversine function mentioned earlier.

Using this complete vector of counties (holding each county's important values), we finally have the necessary values to model the growth and spread of opioid use per area in each county. So we then printed out a list of each county's individual recurrence model.

Since we also want to model the level of opioid addictions per area throughout each state, we decided to find a statewide recurrence model as well. To do this, we grouped each countywide recurrence model together with only other counties from the same state, and averaged all of the coefficients in front of the $R(t)$ term in the countywide recurrence equations. This way, the statewide model will have a recurrence equation representing the average growth rate of opioid use per area

```
********************* StateWide Recurrence Equations *********************

     State:                 Model:
     Kentucky               R_1(t + 1) = (1.051662675)R_1(t) - rR_1(t)
     Ohio                   R_2(t + 1) = (1.059860708)R_2(t) - rR_2(t)
     Pennsylvania           R_3(t + 1) = (1.029827955)R_3(t) - rR_3(t)
     Virginia               R_4(t + 1) = (1.043530144)R_4(t) - rR_4(t)
     West Virginia          R_5(t + 1) = (1.041328182)R_5(t) - rR_5(t)
```

Figure 5: *Recurrence equations for each state*

```
************** Average Recurrence Model **************


          R(t + 1) = (1.046756651)R(t) - rR(t)
```

Figure 6: *Recurrence equation for all states*

from all of its own counties, but still considers the influence effects from counties of other states (because the countywide models already incorporated the outside influence factors).

As there are several hundred counties, we cannot display all of the countywide recurrence models, however, in the Figure 5 above we display the statewide recurrence equations, meant to predict the opioid addictions per area for each listed state.

In Figure 6, we use the average growth factor from all of the countywide models to display a recurrence equation for all the states. This represents the model for the average opioid use per area between Ohio, Kentucky, West Virginia, Virginia, and Pennsylvania.

Testing and Validation

Before addressing other aspects of what our models can give us, we need to test its accuracy. To do this, we compare the results of our model to the NFLIS data provided for each county in the years 2011-2016. In particular, for the years 2012, 2014, and 2016 we will compare the drug counts per area in each county to the opioid use per area given to us by our models.

```cpp
for (size_t n = 0; n < Counties.size(); ++n)
{
    out << std::fixed << std::setprecision(8);
    out << std::setw(50) << Counties[n].name;
    double total_coef;;
    if (Counties[n].gf > 0)
        total_coef = 1 + 10 * Counties[n].I + Counties[n].I_o;
    else
        total_coef = 1 + 10 * Counties[n].I + Counties[n].I_o + Counties[n].gf;

    double rec_values = initial[n];
    rec_values = total_coef * rec_values;
    out << std::setw(30) << rec_values;
    for (size_t i = 2; i < 7; i++)
    {
        rec_values = total_coef * rec_values;
        out << std::setw(30) << rec_values;
    }
    out << std::endl;
}
```

Figure 7: *Code for Recurrence Values*

To do this for the i^{th} county, we start our model off with its initial drug counts per area from the 2010 NFLIS data ($R_i(2010) = \frac{\text{drug counts in 2010}}{\text{area of county } i}$) and use programming software to recursively run through and store the $R_i(t)$ values for $t = 2011, 2012, ..., 2016$. Using the C++ code in Figure 7, we performed this task and stored these values for every county $i \in \{1, 2, 3, ..., n\}$ where n is the total number of counties, using the vector of county objects with corresponding county information.

The program in Figure 7 created a new file filled with each county's model values for the years 2011 through 2016. That is, we created a file with predicted opioid addictions per area for each county, starting with the initial data from 2010. We then used Tableau, a data visualization tool [Tblu], to overlay each county's predicted opioid addictions per area values onto a map of Ohio, Kentucky, West Virginia, Virginia, and Pennsylvania for the years 2012, 2014, and 2016 [Mpbx]. Next, we overlaid the drug counts per area values found from the actual NFLIS data onto the same

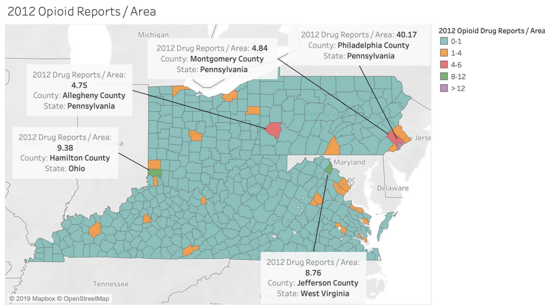

Figure 8: *2012 Opioid Counts per Area*

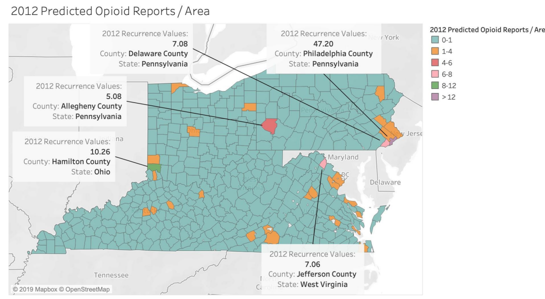

Figure 9: *2012 Predicted Opioid Use per Area*

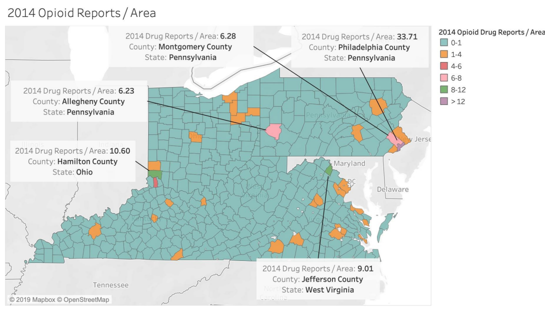

Figure 10: *2014 Opioid Counts per Area*

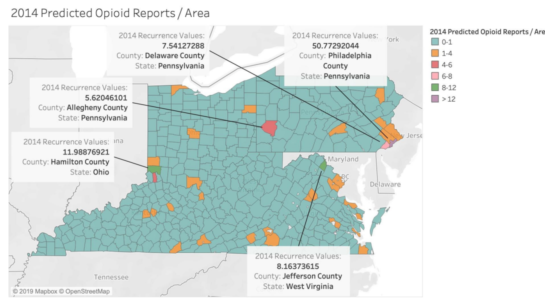

Figure 11: *2014 Predicted Opioid Use per Area*

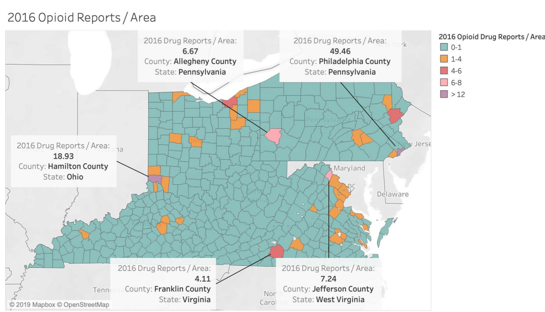

Figure 12: *2016 Opioid Counts per Area*

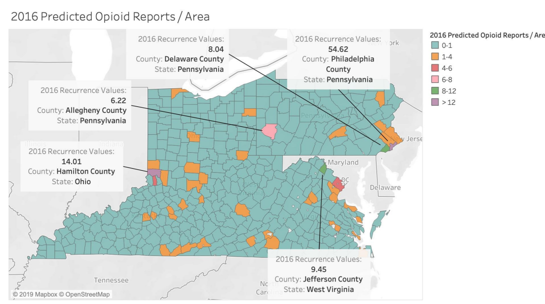

Figure 13: *2016 Predicted Opioid Use per Area*

map for the same years. We display and compare these plots in Figures 8-13 above. Notice, the darker purple shaded counties on the maps represent the highest opioid use/counts per area, and the lighter mint shaded counties on the maps represent the lowest opioid use/counts per area. For your convenience, we labeled a few counties with the largest values.

For each year shown in the above figures, we can see that many of the predicted opioid use per area values are similar, even for years as far out as 2016. For example, in 2016, the NFLIS data shows that the four highest drug counts per area were from Philadelphia County, Hamilton County, Jefferson County, and Allegheny County, which are all within the five highest counties predicted from our model, as shown in Table 1 below.

Reported 2016 Values	Predicted 2016 Values
Philadelphia County 49.46	Philadelphia County 54.62
Hamilton County 18.93	Hamilton County 14.01
Jefferson County 7.24	Jefferson County 9.45
Allegheny County 6.67	Allegheny County 6.22
Monroe County 5.04	Delaware County 8.04

Table 1: *Top five highest opioid use per area values*

Reported Change	Predicted Change
Hamilton County ≈ 9	Hamilton County ≈ 4
Allegheny County ≈ 2	Allegheny County ≈ 1.3

Table 2: *Change in opioid use from 2012 to 2016*

Looking at the trends of opioid use from the figures, it's important to notice that the values of opioid use per area also seem to grow at a similar rate. For example, notice in Table 2 above that Hamilton county seems to have had a significant jump in drug use per area between the years 2012 and 2016, and our model approximates a jump in opioid use values as well. And for a county with only a slight increase in drug use per area between 2012 and 2016, like Allegheny county, our model seems to only increase its values moderately. Therefore, with a reasonable measure of error, our model seems to predict countywide opioid use per area very well.

Identifying Opioid Origins

The coefficient in front of the $R(t)$ term of the recurrence equations calculated for all counties and states represents the predicted growth factor of opioid use per county. However, a better way to think of these "growth factors" is to consider them as values that represent each state's (or county's) potential for opioid use. That is, states with high growth factors in their recurrence equation will demonstrate a large scale and quick spread of opioid use through the county or state. It is important to note that this is *not* the same as predicting that the counties and states with the highest growth factors will have the largest amount of opioid addictions in every subsequent year; however, it does mean that if there is even the smallest amount of opioid use in the initial value (in this case $R_i(2010)$), then as time goes on ($t \to \infty$), these counties or states will have the largest amount of opioid use per area. Using Tableau again, we displayed each county's growth factors, or more appropriately, each county's "potential for an opioid epidemic" values on a map of the the five states in Figure 14 below. Notice that the darker blue shaded counties represent the higher drug use potential, and lighter green shaded counties represent lower drug use potential. We circled the four darkest shaded counties (counties with the highest growth factors) for your convenience.

We decided to identify the origins of the non-synthetic opioid, heroin, as it was reported as the most used opioid in the NFLIS data provided. To visualize this, see Figure 15 below, where each blue blob represents the relative size of individual drug counts for every drug included in the

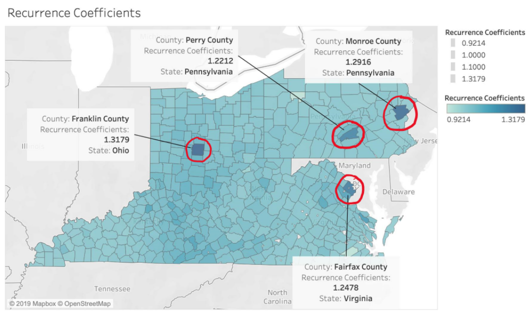

Figure 14: *Countywide Growth Factors (Potential Values)*

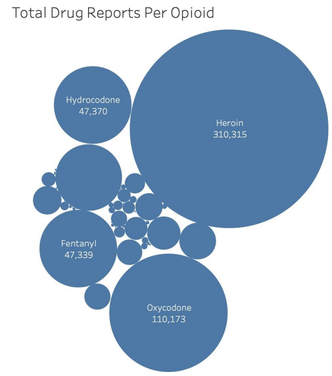

Figure 15: *Total Reports for each Opioid from NFLIS Data*

data. As you can see, heroin clearly had the most reported counts out of all the synthetic and non-synthetic opioids listed in the data set provided.

We use a county's growth factor (the "potential" value) to identify any possible locations where heroin use might have started because if a county has a high growth factor, even a small initial amount of heroin use will exponentially grow in a short period of time. Not only does this mean heroin use in the county itself will increase, but it also means that its influence on other counties will increase as well, and will subsequently impact its surrounding counties and their corresponding growth factors. As this process continues, the surrounding counties with a now higher growth rate will go through the same process, causing a spread of heroin use and heroin use influence. Using Figure 14, we know that Franklin County, Monroe County, Fairfax County, and Perry County have the largest potential values and growth factors, in that order. So we analyze these counties individually.

Given this reasoning, when we look through these four counties to find locations where heroin use may have started, we are looking for a county that does not only show a rapid increase in its relative heroin use, but *also* shows a substantial relative increase of heroin use in its neighboring counties.

Therefore, out of the four counties with the highest calculated growth factors, we are looking to

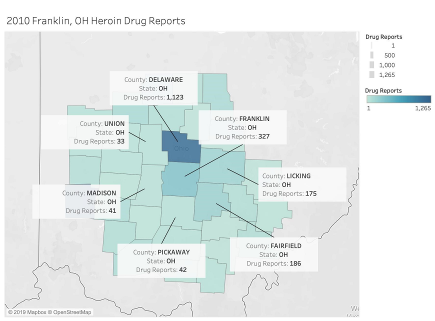

Figure 16

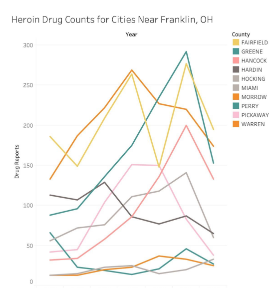

Figure 17

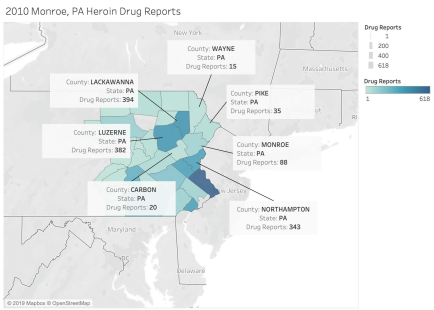

Figure 18

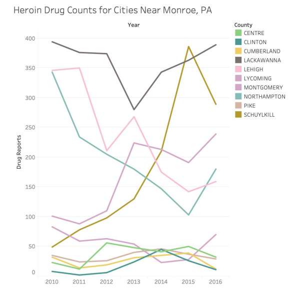

Figure 19

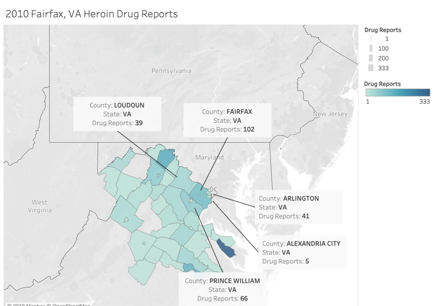

Figure 20

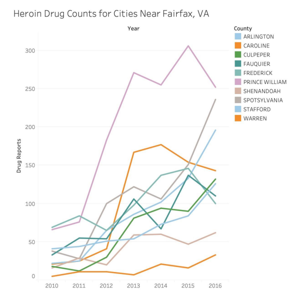

Figure 21

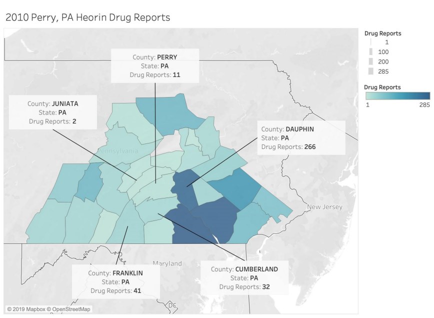

Figure 22

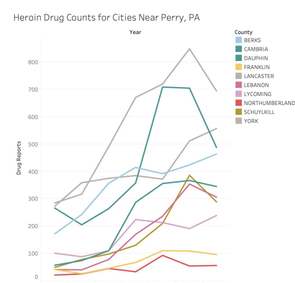

Figure 23

find the counties that show similar heroin use growth in their neighboring counties as well.

Using the data provided, we can use visualization tools to find which of these four counties are suitable choices for a possible heroin use origin. On the left side of the Figures 16-23, we display each of the four counties, along with its closest neighboring counties. For each of the neighboring counties, we also include the number of heroin counts. On the right side of the figures above, we plot heroin counts with time for the ten closest neighboring counties surrounding the specified county. In these figures, it is important to note that each color represents a different neighboring county.

Using Figure 17 for Franklin County and Figure 19 for Monroe County, we can see that most of their neighboring counties do not show a noticeable growth in heroin counts per year, which is not what we are looking for. However, notice Figure 21 for Fairfax County and Figure 23 for Perry County display substantial growth in the heroin counts over time in most of their neighboring counties. This means that out of the four highest potential valued counties, only Fairfax County and Perry County demonstrate the qualities we are are looking for. Therefore, Fairfax County and Perry County could be possible starting locations for heroin use within these five states.

Concerning Patterns and the Threshold Level

Given that most of our countywide recurrence models and all of our statewide recurrence models have growth factors larger than one, the U.S. government should very much be concerned for the future of these five states. We know from Figure 6 that the average growth factor (also described as potential value) for all 5 states, is approximately 1.05. This means that every year, our model predicts an average 5% increase in the number of opioid addictions per year.

Since most of the drugs use per area values predicted from our model remain relatively low (< 1) for the first 5-10 years, we will consider any state/county with an opioid use per area value greater than $1 \frac{\text{drug addictions}}{\text{miles}^2}$ to be experiencing an opioid addiction *state of emergency.* That is, $1 \frac{\text{drug addictions}}{\text{miles}^2}$ will be considered as our drug identification threshold level. Now, obviously a county/state with a high growth factor and/or high initial drug count per area will reach an opioid use *state of emergency* in less time than other counties or states. Using our model, we now calculate how long it will take for each state to reach a *state of emergency*, given each state's average 2010 opioid counts per area as its initial condition.

We rewrite our recurrence model to isolate the years from initial time variable, so we can solve for the number of years from 2010 until the corresponding state is in a *state of emergency.*

$$R(t+1) = (\alpha)R(t) \tag{13}$$

$$\implies R(2010+\tau) = (\alpha)^{\tau} R(2010) \tag{14}$$

$$\implies \frac{R(2010+\tau)}{R(2010)} = (\alpha)^{\tau} \tag{15}$$

$$\implies \tau = \log_{\alpha} \frac{R(2010+\tau)}{R(2010)} \tag{16}$$

$$\implies \tau_{\text{emg}} = \log_{\alpha} \frac{R_{\text{emg}}}{R(2010)} \tag{17}$$

For Kentucky:

$$R(2010) = 0.2367 \text{ , } \alpha = 1.05167 \text{ and } R_{\text{emg}} = 1$$

$$\implies \tau_{\text{emg}} = 28.602 \text{ years}$$

For Ohio:

$$R(2010) = 0.382 \text{ , } \alpha = 1.05986 \text{ and } R_{\text{emg}} = 1$$

$$\implies \tau_{\text{emg}} = 16.553 \text{ years}$$

For Pennsylvania:

$$R(2010) = 1.056 \ , \ \alpha = 1.0298 \text{ and } R_{\text{emg}} = 1$$

$$\implies \tau_{\text{emg}} = 0 \text{ years}$$

For Virginia:

$$R(2010) = 0.305 \ , \ \alpha = 1.0435 \text{ and } R_{\text{emg}} = 1$$

$$\implies \tau_{\text{emg}} = 27.887 \text{ years}$$

For West Virginia:

$$R(2010) = 0.248 \ , \ \alpha = 1.0413 \text{ and } R_{\text{emg}} = 1$$

$$\implies \tau_{\text{emg}} = 34.453 \text{ years}$$

For each state, τ_{emg} represents the number of years (from 2010) before reaching the critical level of $1 \frac{\text{drug addictions}}{\text{miles}^2}$. As we can see above, most of the states still have plenty of years before opioid use per area reaches a *state of emergency*, with the exception of Pennsylvania. It is clear that Pennsylvania is already in a state of emergency from opioid addictions, most likely from Philadelphia county's high density of drug reports per area.

The United States government should be highly concerned as time goes on. Since our model is predicting an inevitable annual increase in opioid use over time, the government should be considering methods to lower each state's/county's central and outside influence values, as they are currently allowing states or counties that have already surpassed the state of emergency level to largely affect states or counties that have not yet reached that critical level. The United States government should also be considering methods to actively increase the recovery rate (parameter r), as that could pull the growth factor down below 1, and consequently lead to a total opioid use of 0 as time goes on ($t \to \infty$). In Part III, we will explore the effects this recovery rate term, r, has on our model. For now, unfortunately, none of the locations currently above the threshold are predicted to go below it without substantial actions against opioid addiction.

Part II

Drug misuse can be vastly influenced by other factors. In an effort to track this relation, the coefficient of determination, R^2, was calculated for many socioeconomic factors. For instance, individuals with associate's, bachelor's and graduate degrees, veteran status, high school graduates, and those who have attended college without completion of a degree, to name a few. As we did not have the resources to examine every listed socioeconomic factor in the data provided, we broke the factors up into separate classes and we randomly chose factors within those classes to examine. The classes were educational attainment, family structure, residence status, and veteran status. The chosen socioeconomic classes, and the limited number of socioeconomic factors may have colored our results, however, in order to simplify data gathering, this was a necessary measure.

There were several steps we took to get the data into an analyzable form. First, we combined data for individual cities with their corresponding or surrounding counties. If the cities were independent, we summed the opioid counts and populations with the counties, but if the cities were officially part of the county, we only added opioid counts to avoid double counting the population. Next, we had to trim away several counties with missing opioid count values from the data, since they would make it impossible to find an R-squared value. Additionally, we pulled the data of some with suspicious or impossible trends in drug counts, and assumed there was some kind of recording error as we did not want them to distort the R-squared values.

To analyze the opioid count data, we first used Tableau, the data visualization tool [Tblu], to create the necessary scatter plots that display correlation, then we simply used the R-squared, average, and median functions in Microsoft Excel to find the mean and median R-squared values across all counties for the socioeconomic factors we were investigating. In the end, we used the mean R-squared values to determine which factors we would focus on addressing. We chose to work with the mean values instead of the median values because we would like the larger and smaller values to have a significant effect on the data, since counties with high levels of opioid abuse are especially high-priority. We say this because counties with exceptionally high levels of opioid abuse exert higher external influences on surrounding counties, worsening the spread of the addiction epidemic.

We found that high veteran populations and high populations of people with an associate's degree were the two factors that most positively correlated with opioid counts in a county, in that order. Although the divorced and residency > 1 population also showed meaningful positive correlations with opioid counts in the counties, they were not as highly correlated as the veteran and associates degree population was. Additionally, in the limited socioeconomic factors chosen, we were unable to find any factors that showed a *significant* negative correlation with opioid counts. However, there were a few factors with low negative correlation coefficients. To represent these R-squared findings visually, we have created a table and bar chart breaking down some of the socioeconomic factors with the highest correlation with opioid abuse and their respective R-squared values:

Factor	R^2
Divorces	0.705
Bachelor's Degree	0.638
Associate's Degree	0.719
Graduate Degree	0.628
Veteran Status	0.730
Residency >1 Year	0.709

Table 3

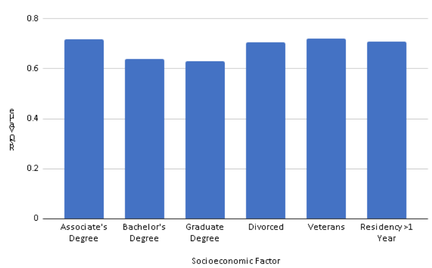

Figure 24: *Histogram of R-squared Values*

Additionally, to demonstrate our process of finding these correlation coefficients, we include several graphs in Figures 25-28 below comparing the socioeconomic factors with the two highest R-squared values, number of veterans and number of associates degree holders, with opioid counts per county in the years 2011 and 2014.

Based on the R-squared values found, our model suggests to us that we should be concerned with the associate degree holders and veteran populations, however, we have only found a corre-

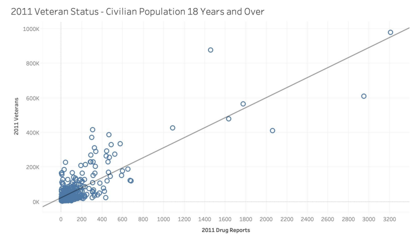

Figure 25: *Opioid Counts vs 2011 Number of Veterans*

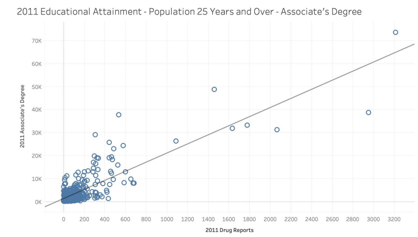

Figure 26: *Opioid Counts vs 2011 Number of Associates Degree Holders*

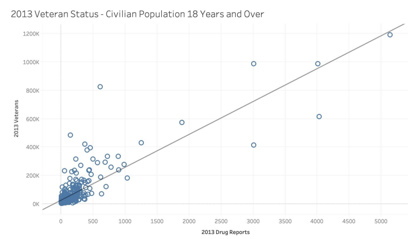

Figure 27: *Opioid Counts vs 2014 Number of Veterans*

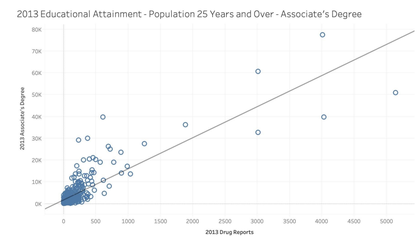

Figure 28: *Opioid Counts vs 2014 Number of Associates Degree Holders*

lated relationship between these factors and opioid abuse, not a causal relationship. Therefore, we decided to make the assumption that greater populations of associates degree holders and veterans directly implies a greater amount of opioid abuse. Using this assumption, we then chose to modify our model by changing the internal influence proportionality constant k into an internal influence proportionality function. We want to make this new function depend on the number of veterans and the number of associate's degree holders in a county so that the county's growth factor will depend on these socioeconomic factors as well. We also want high levels of factors with R-squared values close to 1 to correspond with higher values of this function. In theory, we would also want high levels of factors with R-squared values close to -1 to correspond with lower values of this function, though we did not find any socioeconomic factors (from those that were randomly chosen) with significant negative correlation with opioid abuse.

In the original model the internal influence proportionality constant was set to 10, and we would like to keep it at around that same magnitude in our modified model, since it has displayed accurate projections. Therefore, since the new function we are looking for will be replacing the internal proportionality constant, $k = 10$, we aim to keep the function value reasonably close to 10, with the median value also very close to 10. Moreover, we want to keep the value of this new function positive. Thus, we choose to write the new function, $k(a_i, v_i)$, as a logistic function of the form $\frac{20}{1+e^{-(a_i+v_i-c)}}$, where a_i is the number of associate's degree holders in the ith county, v_i is the number of veterans in the ith county, and c is a constant that will be specified later. Using a logistic function for $k(a_i, v_i)$ allows us to ensure that the function values will remain between 0 and 20, i.e. reasonably close to 10, and that the median value of $k(a_i, v_i)$ will be approximately 10. To use a function of this form for our revised model, we have to make the assumption that the internal influence proportionality function increases logistically with respect to the sum of the number of associate's degree holders and the number of veterans in a county. From the U.S. Census data provided, the median number of associate's degree holders in a county is approximately 4,200, and the median number of veterans in a county is approximately 65,000. Thus, we choose c to be 69,200, giving us the following definition for our new internal influence proportionality function: $k(a_i, v_i) = \frac{20}{1+e^{-(a_i+v_i-69200)}}$. Therefore, the modified model from Part I, now including dependence on important socioeconomic factors from the data provided, is defined as:

$$R_i(t + \Delta t) = (1 + (k(a_i, v_i)I_i + I_i^O))R_i(t) - (m_i + r)\Delta t R_i(t) \tag{18}$$

$$\text{where } k(a_i, v_i) = \frac{20}{1 + e^{-(a_i+v_i-69200)}}$$

Notice, now when a county contains a larger number of veterans and associates degree holders (greater than the median), we have a larger growth factor and potential value, and when a county contains a smaller number of veterans and associates degree holders (smaller than the median), we have a smaller growth factor and potential value. This is because $k(a_i, v_i)$ increases with unusually large numbers of veterans and associates degree holders and decreases with unusually small numbers of veterans and associates degree holders. Therefore, when we use Equation 18 to model the i^{th} county's predicted opioid use per area, we slightly modify the model's predictions depending on the number of veterans and associates degrees in the i^{th} county.

Part III

Based on our findings from Part II, our model suggests to us that the populations of individuals in certain socioeconomic classes like veterans, associates degree holders, and divorcees are correlated with higher opioid abuse within a county. It is importance to note here that the correlations highlighted by our calculated R-squared values do not directly imply causation. However, based on the research and findings from the National Bureau of Economic Research and the Institute for Family Studies, we assume for now that the correlation between the examined socioeconomic factors and the opioid reports implies a causation. [Moaa][Nber][Iffs]. More specifically, we are assuming now that a high population of veterans and divorcees implies a high number of opioid addictions in a county. With this assumption made, we suggest changes that monitor and assist individuals in the socioeconomic factors highlighted above.

First, we recommend implementing programs to help steer veterans away from abusing opioids. According to a 2011 study by the Veteran's Association, veterans are twice as likely to die from an opioid overdose as compared to civilians. There are more similarly worrying statistics, as outlined in the Military Officer's Association of America's (MOAA) article on veterans and opioid abuse [Moaa]. The article also elaborates on the addictions that arise from over-reliance of the soldiers on the pain medications they are often prescribed. Additionally, this article states that when they return from duty, a great number of veterans suffer from mental health issues like Post Traumatic Stress Disorder (PTSD) and consequently resort to substance abuse. These effects are also well documented in the National Bureau of Economic Research's paper about the war on terror igniting an opioid epidemic [Nber]. The Department of Veterans Affairs (VA) has acknowledged the adverse effect of the prescription opioid crisis on veterans and has implemented a guideline that allows doctors to determine when patients must be instructed to stop consuming their medications. They also released a study suggesting that opioid based pain medication may not actually be superior to non-opioid alternatives and may sometimes lead to stronger, more dangerous, side effects [Vhpm]. Based on the positive results from the U.S. Department of Housing and Urban Development — Veterans Association's supportive housing program, the newly implemented housing support program that grants veterans housing vouchers and other services, a potential solution to our rise in opioid use within the Veteran population could involve providing counselling and having centers akin to retirement homes for veterans who may not have homes of their own [Usva]. By guiding them away from abusing opioids, this could prove to be an effective solution to increasing our addiction recovery rate r, and therefore decreasing our models projected growth in a county. [Usva].

Another factor that our model suggests is highly correlated with drug use is divorces. A report from the Social Capital Project validates this correlation with the statistic that never-married and divorced adults account for 71% of opioid related deaths, despite making up only 28% of the adult population [Iffs]. Since we are assuming that a population with higher numbers of divorces is in fact causing higher opioid addictions, we want to provide solutions that have been shown to help prevent divorces. Marriage counselling, which employs the Emotionally Focused Therapy (EFT) strategy has a 75% success rate according to the American Psychological Association [Psyc]. Therefore, the establishment of inexpensive marriage counselling programs may help reduce the number of divorces and consequently, the substance abuse associated.

According to the PEW Research center, there is no relationship between prison terms and drug misuse [Pewt]. Based on these findings, it seems appropriate for drug users to not be punished, but rather assisted. According to the National Institute of Drug Abuse, using rehabilitation centers could save the government considerable amounts of money, as drug treatment is significantly less expensive than imprisonment costs (1 year of methadone treatment is $4,700 per patient per year as compared to $24,000 per inmate per year imprisonment cost) [Nida2]. Therefore, a potential solution to the rise in opioid abuse could be to open up more well-equiped rehabilitation center's in areas with high opioid use growth factors found in our model. For example, we could attempt to open up more rehabilition centers in counties with growth factors greater than 1.2 such as Franklin County, Monroe County, Fairfax County, and Perry county (shown in Figure 14). Unfortunately, of the 23 million people currently addicted to substances (not limited to opioids), a vast majority do not seek help. Although this can be for various reasons, the biggest is a financial barrier, since treatments are a luxury only a few can afford [Drhb]. Therefore, another strategy could be to find a way to make rehabilitation centers more affordable and consequently, more accessible. A similar drug policy emphasizing rehabilitation over punishment has shown to be very successful in Portugal, where heroin abuse rates have declined after all drugs were decriminalized and recovery programs were made freely and widely available [Ptgl]

In a large country like the United States, containing the spread of diseases and addictions can seem like a daunting task. However, counties can be prioritized based on the findings from our recurrence models. Clearly, the central and external influence values vary greatly between the counties. The coefficient of a county's recurrence model, which was interpreted as a "growth factor" or "potential value", could be used to prioritize which counties to help first. That is, the counties with relatively larger growth factors can be ranked higher since their opioid use is growing faster. On the other hand, counties with a growth factor lower than one should be considered as "model counties" and could possibly be studied by outside counties since the projected opioid reports in these counties are decreasing annually and are expected to reach zero opioid addictions per area over time.

Finally, to model the effectiveness of these strategies, we can make certain changes to the recurrence equations. For counties that set up veteran housing and establish marriage counselling centers, we can reduce the central influence value, I_i, accordingly. To complement this change in the central influence value, the recovery rate parameter, r, could also be increased for counties that establish more affordable rehabilitation centers, since our sources suggest that the affected individuals will be more likely to beat their addictions given these facilities. The reductions in the central influence value and the increase in the recovery parameter r should be made with the goal of reducing the coefficients in the countywide recurrence equations (the growth factors) to a value below one because then the opioid use per area will go to zero over time.

Conclusion and Future Work

Using only the limited data provided, we developed a validated model that can be used to project future opioid use per area values in a given county. This model includes the effects of greater populations that have higher correlations to drug use. For future work, if we were given data between each year, we might consider using a countywide continuous model (derived from our recurrence equation) which would provide projections of drug use without such a large discrete time step of 1 year. If given more time and data, we could also consider *all* the socioeconomic factors that could be correlated with drug use in a county. That is, we could look for more positively and negatively correlated socioeconomic factors from the 2010-2016 U.S. Census data provided. We did not have the resources to calculate these values for the current version of the model, however, examining other factors could certainly change our current definition of the internal influence proportionality function from Part II. Since our calculated growth factors and influence values were time-independent, future work could also involve a time-dependent growth factor. This way the change in a county's influence values could change the outside influence it has on other counties over time. Lastly, future work could also take into account a larger geographical area to get a broader picture of the epidemic as it spans the United States.

Additionally, if more data becomes available, we could modify some of our assumptions. For instance, the assumption that all deaths in a county in a year are due to opioids is probably an oversimplification, but when designing the model we did not have the data to make a more precise estimate. If we had data about cause of death, a better estimate would be possible. Also, it is plausible that there are more opioid users in counties than are caught, so if we had some kind of data about what fraction of opioids in a county are discovered by police, we could make a better estimate of the number of opioid addicts in a county as a function of police opioid counts. Furthermore, distance is not the only factor that affects how much influence two counties have on each other. For example, two counties with significant trade between them may have a large amount of influence on each other even if they are not very close geographically. If given more data about trade between counties, we could make some modifications to external influence values to account for this, though this would probably complicate the model significantly. It is difficult to say how these changes would affect the model or the relative growth factors of counties, since we do not know how different the real data would be from our assumptions. However, if only a portion of deaths are assumed to be from opioid addicts, projected opioid addictions in all counties would increase, since the number of opioid addict deaths would be less than what we currently assume it to be.

Memo to the DEA/NFLIS Chief Administrator

November 27, 2019

Dear Chief Administrator,

As you are aware, the United States is currently experiencing crisis levels of opioid abuse, an epidemic that has resulted in many hundreds of deaths, lower quality of life, economic loss, and additional strain on the health care system. This epidemic is especially severe in small towns and in rural and industrial areas. In order to help address this emergency, we have developed a model to estimate and predict the spread of opioid abuse over time. We hope that our model will be able to assist the CDC and DEA in their efforts to pinpoint locations from which opioids are spreading and to identify socioeconomic factors that are predictive of opioid abuse.

Our mathematical model focuses on opioid abuse in five states in particular: West Virginia, Virginia, Kentucky, Ohio, and Pennsylvania. From the years 2010-2016, we used initial socioeconomic data and police data about opioid counts per county to model and analyze projections of drug use per area on an annual basis. This data was extracted from the NFLIS database. Regarding the structure of our model, we used a recurrence model that accounts for possible external influence regions and internal influence regions that may affect an individual county's drug use trends. In other words, we assumed that counties and states were more likely to see an increase in opioid abuse levels if nearby counties and states experienced high levels of opioid abuse, since the drugs could easily be moved across county borders, and many people travel between nearby counties and states.

Based on our verified model projections, we chose to consider a county to be in a state of emergency if there is at least one opioid addict per square mile. We found that on a state level, Pennsylvania is predicted to have the worst levels of opioid abuse, while West Virginia is expected to suffer the effects of the crisis the least. In each state, we also found a variation in opioid abuse growth rates on a county level. The counties where opioid use showed the highest potential growth were Franklin County, Monroe County, Fairfax County, and Perry County. Additionally, out of all the socioeconomic factors randomly chosen, those with the highest correlation with opioid abuse were the number of veterans in a county, the number of associate's degree holders in a county, the number of residents in a county who have lived there for at least one year, and the number of divorcees living in a county.

Based on our findings and the previous research of other organizations, we have come up with several recommendations we believe will be effective in countering the spread of opioid abuse in these five states. First, since veterans seemed to suffer high rates of opioid abuse, we suggest increased funding for programs benefiting veterans. For example, the government could fund therapy and rehab programs within the VA. Studies have suggested that such programs may reduce opioid abuse rates within veteran populations. These programs should be implemented everywhere, but highest priority should be placed on Pennsylvania since that is where the crisis is predicted to be the worst. Though our model revealed a correlation, there is not enough information to firmly prove causation or to find a mechanism for causation. Therefore, our model only suggests that the government should be concerned with the socioeconomic factors that are highly correlated with opioid abuse, such as associate's degree holders, relatively new residents, and divorcees. However, we can also suggest that more research should be done to find any causal relationships and to implement possible solutions accordingly. For example, further research could be done to determine if having an associates degree actually leads to a higher chance of opioid use, and if it does, what kind of programs would be most effective at helping associate's degree holders.

We hope that our recommendations will be given consideration and that our model may be of further use in the fight against opioid abuse.

Best regards,

Melissa Lee
Andrew Mashhadi
Bryce Palmer-Toy
Rishab Rajagopalan

References

[Mpbx] "Mapbox." *Mapbox*, https://www.mapbox.com/. Accessed 26 November 2019.

[Tblu] "Tableau." *Tableau*, https://www.tableau.com/. Accessed 26 November 2019.

[Hrsa] "Opioid Crisis." *Health Resources & Services Administration*, https://www.hrsa.gov/opioids. Accessed 26 November 2019.

[Phys] Hugh Young. *University Physics with Modern Physics 14th Edition* Pearson, 2016.

[Nida] "Opioid Overdose Crisis." *National Institute on Drug Abuse*, https://www.drugabuse.gov/drugs abuse/opioids/opioid-overdose-crisis. Accessed 26 November 2019.

[Armp] "ArcMap." *ArcMap.* http://desktop.arcgis.com/en/arcmap/. Accessed 26 November 2019.

[Moaa] "Veterans and Opioid Addiction." *Military Officers Association of America.* https://www.moaa.org /content/publications-and-media/features-and-columns/health-features/veterans-and-opioid -addiction/. Accessed 12 December 2019.

[Nber] "Did the War on Terror Ignite an Opioid Epidemic?" *National Bureau of Economic Research.* https://www.nber.org/papers/w26264.pdf. Accessed 12 December 2019

[Iffs] "Opioid Deaths Are Surging Among Single and Divorced Americans, Especially Men." *Institute for Family Studies.* https://ifstudies.org/blog/opioid-deaths-are-surging-among-single-and-divorced-americans-especially-men. Accessed 12 December 2019.

[Usva] "Homeless Veterans." *U.S. Department of Veterans Affairs.* https://www.va.gov/homeless/hud-vash.asp. Accessed 12 December 2019.

[Vhpm] "Opioid Safety." *U.S. Department of Veterans Affairs.* https://www.va.gov/painmanageme-nt/Opioid_Safety/index.asp. Accessed 12 December 2019.

[Psyc] "Couples Therapy: Does It Really Work?" *Psychology Today.* https://www.psychologytoday.com /us/blog/in-it-together/201712/couples-therapy-does-it-really-work. Accessed 12 December 2019.

[Drhb] "Barriers to Substance Abuse Treatment." *Drug Rehab.* https://www.drugrehab.com/treatment/ barriers-in-seeking-treatment/. Accessed 12 December 2019.

[Pewt] "More Imprisonment Does Not Reduce State Drug Problems." *PEW Trust.* https://www.pewtrusts. org/en/research-and-analysis/issue-briefs/2018/03/more-imprisonment-does-not-reduce-state-drug-problems. Accessed 12 December 2019.

[Nida2] "Principles of Drug Addiction Treatment: A Research-Based Guide (Third Edition)." *Na tional Institute on Drug Abuse.* https://www.drugabuse.gov/publications/principles-drug-addiction-treatment-research-based-guide-third-edition/frequently-asked-questions/how-effective -drug-addiction-treatment. Accessed 12 December 2019.

[Ptgl] "Portugal's radical drugs policy is working. Why hasn't the world copied it?" *The Guardian.* https://www.theguardian.com/news/2017/dec/05/portugals-radical-drugs-policy-is-working-why-hasnt-the-world-copied-it. Accessed 12 December 2019.