

Summary

Problem C: The Opioid Crisis

Team #1926726

Our team constructs a model based on the SIR model for epidemics to explain the trends of opioid use in Ohio, Kentucky, West Virginia, Virginia, and Pennsylvania. This model – the SIRI model – is a system of three linear ordinary differential equations describing the transition between non-users, users, and sober individuals. We incorporate elements of the SIRS model into our model, allowing individuals to transition between a user and sober state. However, once one becomes a user, they can never transition back to a non-user state.

Following the construction of our model, we analyze the data on the drug incident reports. We compare the data on a state by state basis; for each state we look at the data county by county. From this, we hypothesize which county in each state initiated the opium epidemic. We conclude that Jefferson, Kentucky; Cuyahoga, Ohio; Pittsburgh, Pennsylvania; Richmond, Virginia; and Kanawha, West Virginia are likely origins. We also analyze socio-economic conditions—specifically level of education—and their association with opioid use. There is no apparent correlation between not having a high school diploma, having a high school diploma, having a bachelors degree, or having masters degree and using opioids.

Using our model, we attempt to predict future patterns. We note that our model – though it has good correlation with the data – does not fit the shape of the data as well as we would like. We propose several corrections to our model to increase accuracy.

Our solution to combating the opioid crisis is to increase the availability and effectiveness of government rehabilitation centers. This will decrease the length of time individuals are addicted to opioids and the probability they will relapse. Because there was no correlation between education and opioid use, we do not believe increasing education will decrease the number of users.

Unfortunately, we did not have sufficient time to test the impact of all of the socio-economic factors on opioid use. Thus, we did not alter our model to include the effects of the socio-economic factors.

Mathematical Contest in Modeling
COMAP

Chief Administrator
DEA/NFLI Database

January 28, 2019

To Whom It May Concern,

After analyzing the data on opioid use in Ohio, Kentucky, West Virginia, Virginia, and Pennsylvania, our team has noted several important trends.

First, we hypothesize the opioid epidemic originated in the following counties: Jefferson, Kentucky; Cuyahoga, Ohio; Pittsburgh, Pennsylvania; Richmond, Virginia; and Kanawha, West Virginia. We noted that the number of reported drug incidents tends to increase with the size of the population. However, there are also a variety of unknown variables causing spikes and decreases in the data that require further analysis.

Second, we noted that there is no correlation between education and drug use. Therefore, increasing ad campaigns and drug education will likely have little to no effect on the number of people who are users.

Our proposed solution to counter the opioid crisis is to increase the availability and effectiveness of government rehab centers. We could improve rehabilitation centers by having stricter regulations and guidelines. Increasing rehab availability will encourage more people to seek treatment, decreasing the length of addiction. Increasing the effectiveness of the rehabilitation centers will decrease the possibility that individuals will relapse.

We urge you take our findings into consideration when implementing new tactics to battle the opioid crisis.

Sincerely,

Team #1926726

COMAP 2019
Mathematical Contest in Modeling

Team #1926726
Problem C: The Opioid Crisis

January 28th, 2019

Contents

1	Introduction to Problem	3
2	Assumptions and Justification	3
2.1	Variables	3
2.2	Model	4
3	Model Design	5
4	Testing and Analysis	6
4.1	Analysis	6
4.1.1	Counties	7
4.1.2	Current Trends	11
4.1.3	Predicting Future Trends	11
4.2	Testing	14
5	Strengths and Weaknesses	15
5.1	Strengths	15
5.2	Weaknesses	15
6	Socio-Economic Trends	16
7	Solutions to Problem	17
8	Conclusion	17
	References	19
	Appendix	20

1 Introduction to Problem

Our task is to aid the United States in developing a better understanding of the trends in use of synthetic and non-synthetic opioids. Our team was given drug identification counts from every county in Ohio, Kentucky, West Virginia, Virginia, and Pennsylvania in the years 2010 through 2017. Although the problem description includes Tennessee, the data set includes Pennsylvania rather than Tennessee; we proceed assuming we were working with data from Pennsylvania. The drug identification counts are for both synthetic opioids and heroin.

Using the data on drug identification counts, we construct a mathematical model describing the trends of opioid and heroin incidents across the five states over eight years. Using this model, we answer several questions. First, we determine where in each state opioid use may have started. Then we determine how concerning any patterns and trends are to the United States government and at what level of drug use these patterns occur. Lastly, we use our model to predict where and when these patterns will occur in the future.

After completing the model, we use the socio-economic data of each county in the above states between 2010 and 2016 to alter our model; the goal is to include pertinent factors about the influence of socio-economic conditions on drug use. Questions to be answered include: how did so many people become addicted to or begin using opioids; what type of people are using (for medical purposes) or abusing opioids; why did opioid use and addiction increase; what caused opioid increase; why do people continue to use opioids despite the danger that continued use poses?

After completing Part 1 and Part 2, we use our revised model to identify and develop a method to counter the opioid crisis. We test how viable our solution is using our model. This will show us what parameters determine the success or failure of our method.

2 Assumptions and Justification

2.1 Variables

The variables included in our model are as follows:

- t = time (discrete)

- S = non-user of drugs
- I = drug user
- R = individual who is sober
- N = total population
- k = rate of spread from non-users to users
- d = duration of drug addiction
- r = rate at which sober individuals relapse and become drug users again

We were unable to extrapolate values for several of these variables, so initial values were determined through literature. We found a value for N through the census data provided. The initial population for the 5 states was approximately 32,000,000. Our values for d and r were determined based off previous studies [3] [5]. Our k value was determined by finding the rate of change of the number of users against the total population from 2010-2017, i.e.

$$k = \frac{\Delta I}{\Delta N}. \quad (1)$$

2.2 Model

Our underlying assumptions about the model are as follows:

- The data provided are correct.
- The total population (N) remains constant.
- The total population is $N = S + I + R$.
- There are no deaths by overdose.
- Once one becomes a user, one can only transition to sober; a user can never revert to being a non-user.
- Each drug incident represents one drug user.
- All drugs are equally likely to be counted among the total drug incidents.

We begin with the assumption that the total population N of the five states neither increases nor decreases over the 8 years – people simply change from being non-users to users to sober. Realistically people will move out of the state, be born, or die, so there are other factors that should influence the population. However, because we are limited to using the data provided with this problem, we do not have access to birth rates, immigration rates, and death rates. Therefore it was more pragmatic to describe N as constant.

3 Model Design

We propose a modification of McKendrick and Kermack’s SIR model for epidemics [6] in which an individual can go from a non-user of drugs (S) to a drug user (I). Once the individual is a drug user, they can either stay a drug user, or they can become sober (R). If a user becomes sober, there is a chance that the sober individual can become a drug user again, as depicted in the figure below.

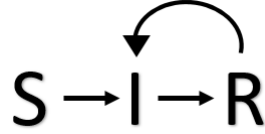


Figure 1: The SIRS Model

We assumed the spread of drugs would mimic that of disease. This hypothesis was backed up by previously developed research on the drug epidemic, which also used the SIR model [1][2].

We took inspiration from the SIRS model, which is a modification of the SIR model in which the recovered R can transition back to the susceptible population S [4]. We accounted for relapses by allowing the sober population to reenter the user population. This model, which we call the SIRS model, consists of a system of three linear ordinary differential equations:

$$\frac{dS(t)}{dt} = -\frac{k}{N}S(t)I(t) \quad (2)$$

$$\frac{dI(t)}{dt} = \frac{k}{N}S(t)I(t) - \frac{1}{d}I(t) + rR(t) \quad (3)$$

$$\frac{dR(t)}{dt} = \frac{1}{d}I(t) - rR(t). \quad (4)$$

4 Testing and Analysis

Before we could test our model, we had to analyze the data provided. Analyzing this data allowed us to notice common trends. Once we identified these trends in the data we could test how well our model fit.

4.1 Analysis

We began by analyzing the trends of drug use between the five states. Below is a graph depicting the trend in drug use over seven years. The dark blue line represents a locally weighted smoothing technique (loess) that performed kernel smoothing over the state total drug reports of each state from 2010 to 2017 with a span of 0.2. Since this data is yearly over eight years, seasonality is not considered in this summary.

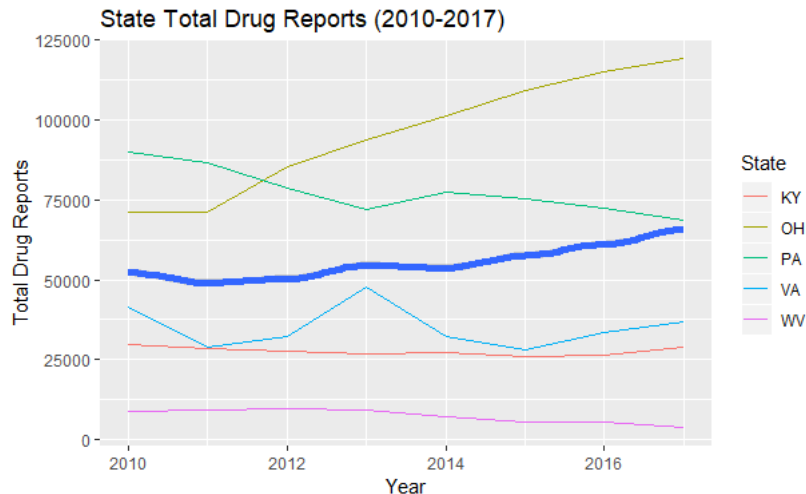


Figure 2: Total drug incidents reported for each state (2010-2017).

The overall trend of the total state drug reports – as captured by the blue line – is increasing in an upward, concave manner. The upward trend is likely influenced significantly by the increase in drug incidents in Ohio. Pennsylvania and West Virginia have slightly decreasing numbers of drug reports, and Virginia and Kentucky have slightly increasing numbers of drug reports. The cause of increasing drug reports in Virginia and Kentucky over the seven years could likely be attributed to population growth. The cause

for the increase in drug reports in Ohio beginning in 2011 is not as obvious. Interestingly, Pennsylvania has a lower number of incidents in 2017 than Ohio despite having a significantly larger population.

4.1.1 Counties

Below are graphs depicting the trends of drug use in every county for each of the five states. We hoped that analyzing the state trends based on county trends would give us more insight into why certain states had more incidents than others. Furthermore, using this data we can form hypotheses about where opioid use started in each state, what trends the United States government needs to be aware of, as well as when they need to consider them. It should be noted that the abundance of counties made a legend unreadable; lines of interest are identified in the text.

Beginning with Kentucky (Figure 3) we can see that overall, most counties have between 0 and 1000 drug incidents per year between 2010 and 2017. There are four counties that are clearly and disproportionately above that 1000 incident threshold – Jefferson (top blue), Fayette (green), Kenton (lower blue), and Campbell (orange). Furthermore, there is a clear jump between the county with the greatest number of incidents – Jefferson – and the county with the second most – Fayette. Kenton lies between Fayette and Campbell during the entirety of the seven year period. It is logical to assume that the reason Jefferson has the highest number of incidents is because it is the state capital, and therefore has a higher population.

Ohio (Figure 4) follows a similar trend as Kentucky, in that the majority of counties lie below 2,500 incidents per year between 2010 and 2017. There is more variation in which counties have higher drug use, with some overlapping and overtaking other counties. Hamilton (green) grows in a concave up manner. In 2012, Cuyahoga (gold) surpasses Hamilton's number of incidents until 2017 when they line up. Montgomery (blue) was rather stable, and was surpassed by Franklin in 2014 before it began to level off. Lastly, Lake (turquoise) has steady growth during all seven years.

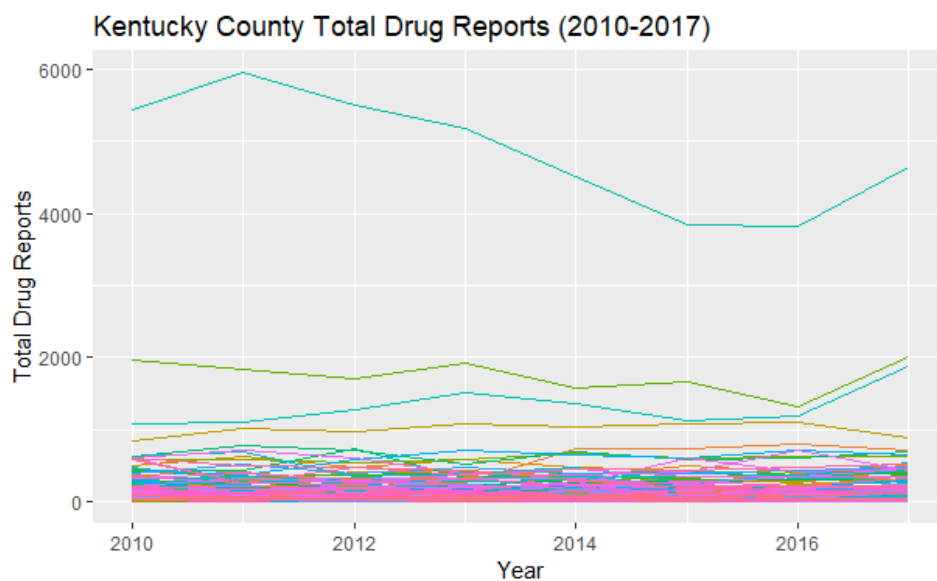


Figure 3: Total drug incidents reported for each county in Kentucky.

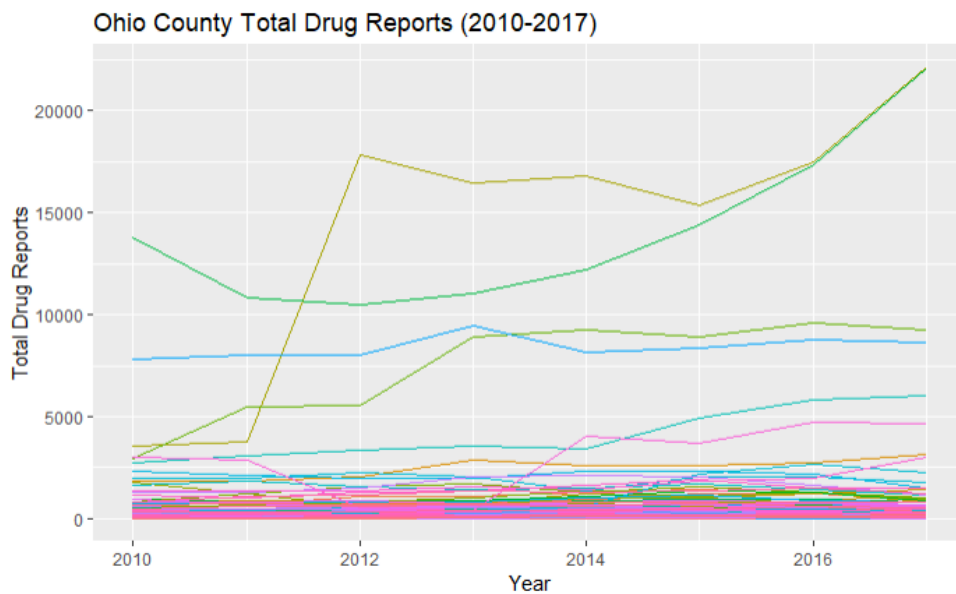


Figure 4: Total drug incidents reported for each county in Ohio.

Unlike the previous two states, Pennsylvania (Figure 5) is clearly stable or decreasing in the number of drug incidents. The threshold value is around 5,000 incidents with two counties being clear outliers. The county with the largest number of incidents, Philadelphia, has been – for the most part – decreasing over the past seven years. It is interesting to note that no other state has an incident count near 35,000. This is most likely due to the comparative size of the state Pennsylvania and the county of Philadelphia. The county with the second most incidents, Allegheny (red), is comparatively stable.

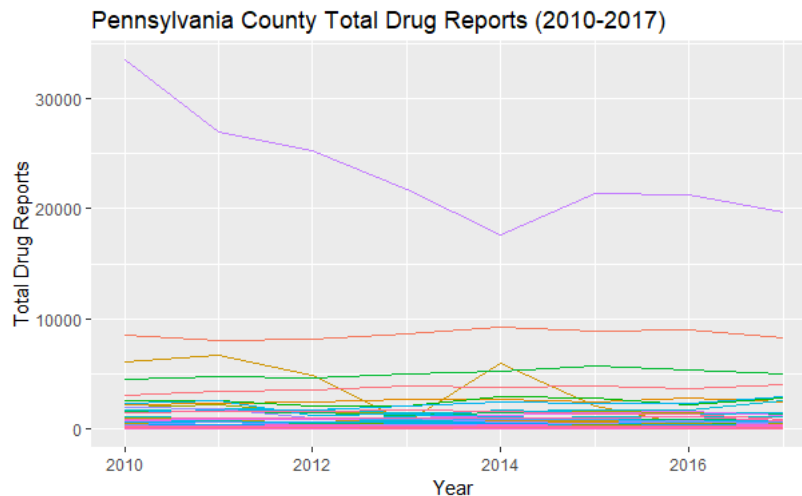


Figure 5: Total drug incidents reported for each county in Pennsylvania.

Virginia (Figure 6) has large variations within its counties. Most of the counties begin with high numbers of drug incidents in 2010 – with Richmond (pink) being the largest – that sharply decrease by 2011. There is another spike – the largest of which was again Richmond – in incidents in 2013 that again decreases. It would be interesting to see what the historical context is in relation to these spikes; was there some economic downturn or legal incident which occurred during these years? The threshold value – if we ignore the spikes in 2010 and 2013 – is approximately 1,000 incidents. The county with the greatest number of incidents is Fairfax (green) which possesses an increasing trend.

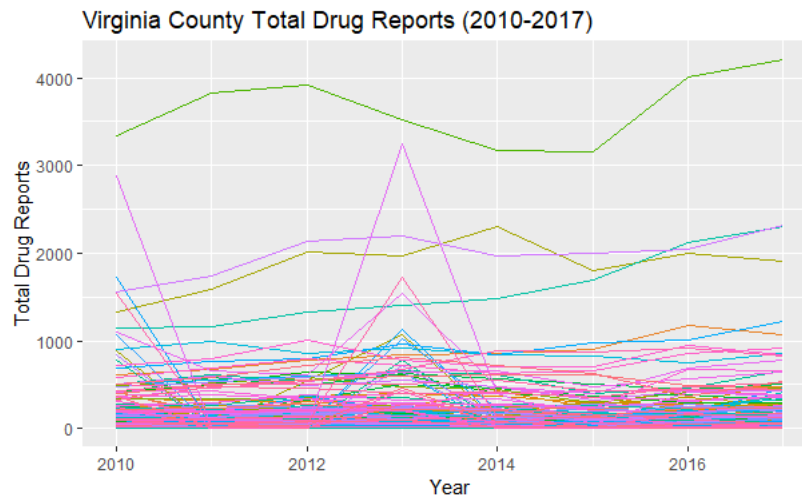


Figure 6: Total drug incidents reported for each county in Virginia.

West Virginia (Figure 7) has a fairly clear decrease in incidents over the seven year period. The threshold is very low—about 250 incidents—likely due to the smaller population. Interestingly, the county with the highest number of incidents—Kanawha (green)—appears to be the only county that has an increasing number of incidents. Furthermore, from 2012 to 2015 the number of incidents decreased dramatically.

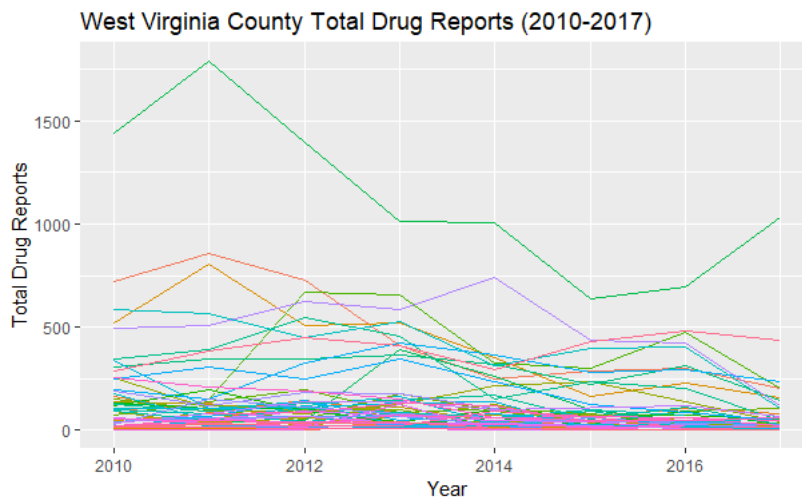


Figure 7: Total drug incidents reported for each county in West Virginia.

4.1.2 Current Trends

The current trend at the state level is a slight increase in drug incidents.

At the county level we can see that Kentucky is relatively stable besides the three counties who's number of drug incidents increases in 2017. We could hypothesize that opioid use may have started in Jefferson, since it begins and remains the most prevalent source of drug reports, and changes to such reports in Jefferson don't correspond to expected changes in other counties.

In Ohio, we see a sharp spike of reports in Cuyahoga that correspond to several other county reports in subsequent years. This could indicate that Cuyahoga incited the increase in opioid use beginning around 2012.

Pittsburgh shows a steep downward trend, so it would be logical to assume that was the origin for opioid use in Pennsylvania.

As a whole, Virginia's reports are increasing. However, it appears that there is some confounding variable that incited irregular spikes in the data. We hypothesize Richmond is the source of the opioid epidemic in Virginia as other counties mirror its spikes. Just as with Kentucky, however, these hypotheses should be taken with caution, as we are unable to tell how the growth of opioid use was influenced by Richmond.

Lastly, West Virginia has a clear downward trend in incidents, barring 2017 when Kanawha had an increase in drug incidents. It is likely that Kanawha is the source of opioid use in West Virginia, as other counties follow the trend of decreasing use.

4.1.3 Predicting Future Trends

Our model predicts that if we let the current rates of drug addiction and relapse remain the same, then from 2017 to 2022, there will be 1,496 more individuals who are drug users.

To determine a possible strategy for countering the opioid crisis, we tested our model with different values of d , the duration of drug addiction and r , the rate of relapse. Here the x-axis represents time from 2010 to 2025.

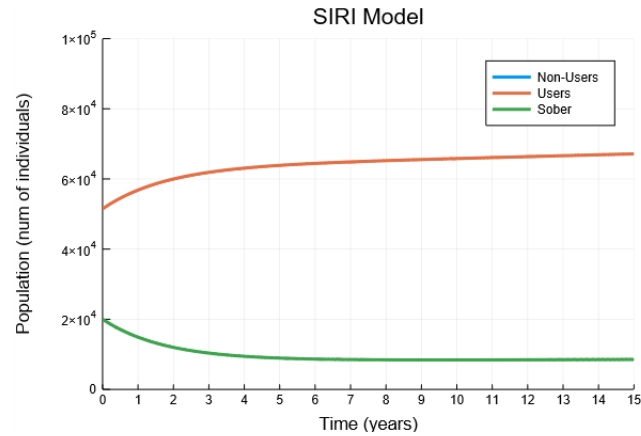
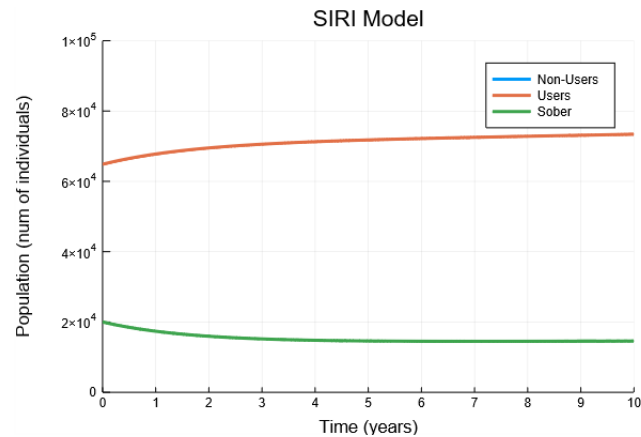
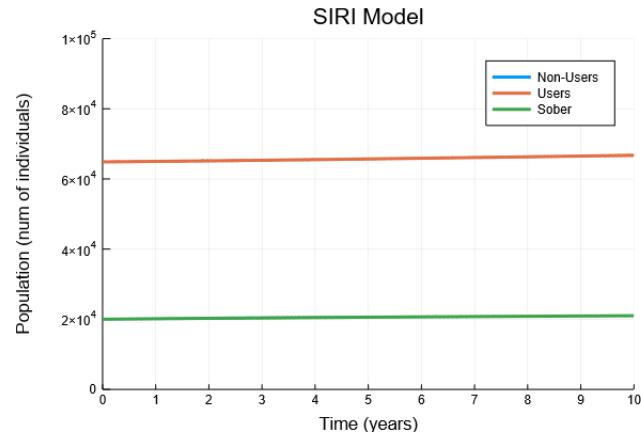


Figure 8: Original SIRI model data

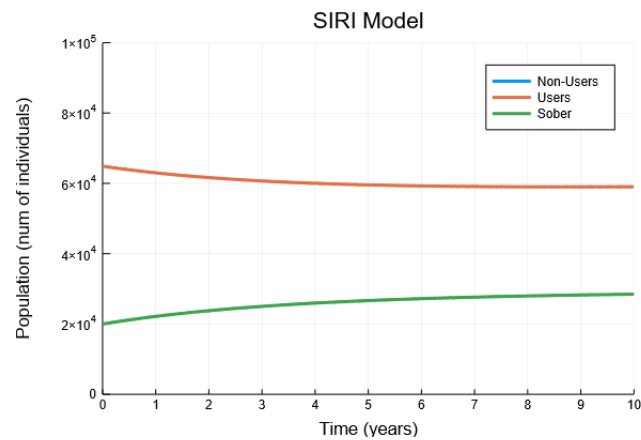
If we lower the average duration of drug addiction from 15.6 years to 10 years, our model predicts that between 2017-2022 there would be 6,919 more individuals who are drug users. This is an additional 5,423 people addicted to drugs than the original model. Here the x-axis represents time from 2017 to 2027.

Figure 9: SIRI model with $d = 10.0$

Instead, if we lower the rate of relapse from 50% to 20%, we see that between 2017-2022 there would only be 840 more individuals who are drug users, which is 656 less drug users from the original model. Again, the x-axis represents time from 2017 to 2027.

Figure 10: SIRI model with $r = 0.2$

The best results come from a combination of lowering both the duration of drug addiction and the rate of relapse. If we lower both, then between 2017-2022, our model predicts that there will be a decrease in drug users; 5,308 individuals will stop doing drugs and become sober. Yet again, the x-axis represents time from 2017 to 2027.

Figure 11: SIRI model with $d = 10.0$ and $r = 0.2$

It is important to note that there is a lack of predictive power since the parameters are constant and don't change with time.

4.2 Testing

The success of the model is determined by two factors: how well it explains the variation in the current historical data and its forecasting ability. To test this, the locally weighted smoothing technique with a span of 0.2 (loess) used to capture the trend in state total drug reports was used as a basis to test the model. Simple correlation is used to determine the effectiveness of the model in Figure 12, where predicted values obtained from the mathematical model were plotted against values obtained from locally weighted smoothing (loess). The correlation coefficient of 0.70 indicates that the model explains about 70% of the variation in the trend of the data. Additionally, Figure 13 demonstrates that the model not only overestimates, but unfortunately grows in the opposite way that the trend appears to be growing. However, we may be able to use this model to predict values in the near future.

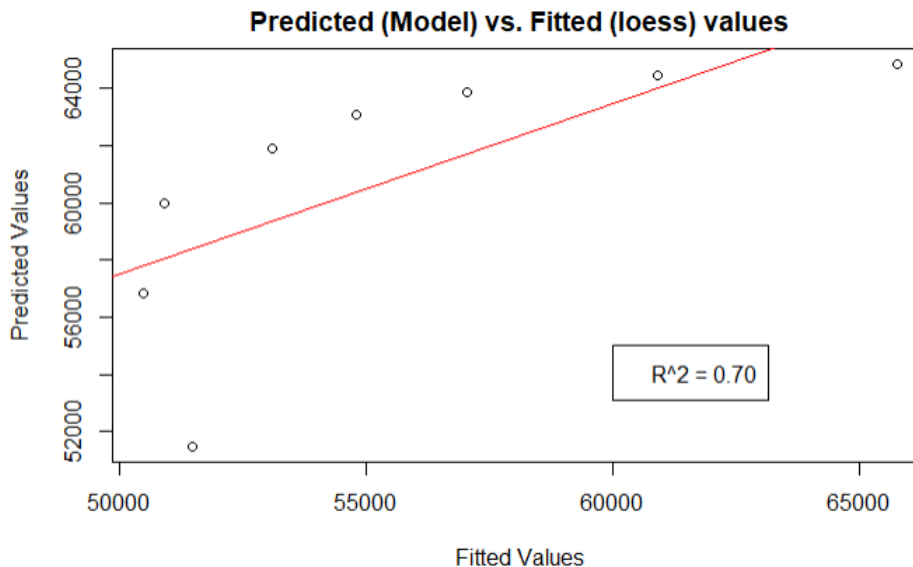


Figure 12: Model vs Data Trend

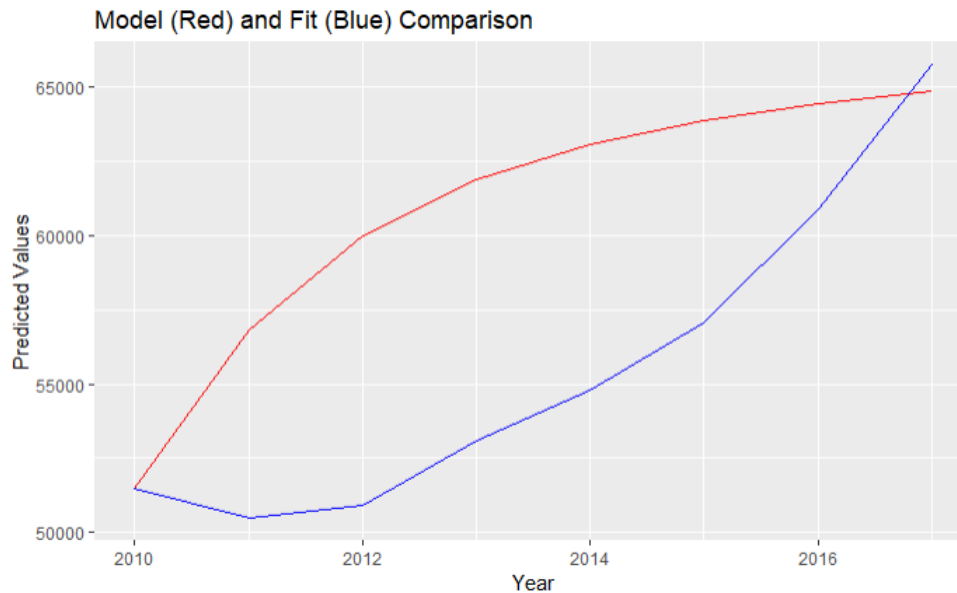


Figure 13: Model vs Data Trend Line

5 Strengths and Weaknesses

5.1 Strengths

The strengths of our model can be found in its simplicity and few parameters. This means it is easily explained and easily altered. Because we understand its shortcomings, we can add features from other population and epidemic models to make our model more effective and accurate. We can do the same by adding more parameters. Another strength of our model is that it is based off other well known models. Therefore, we are reasonably certain that our model is supported by other mathematicians.

5.2 Weaknesses

While we feel that our model fits the data within reason, there are ways we could further improve the model. Currently, our model has no way for individuals to exit the system – either by death or by moving to another state – so after an extended period of time everyone will become either a drug user or a sober individual. If we had data indicating the number of

overdoses that occur from drug users, we could adjust our model to have another compartment, the overdose or O compartment and thus, a fourth equation. Then, drug users could go from being a drug user to a sober individual or overdosing and dying.

Our model reasonably fits the given data. However, the shape of our data compared to the provided data shows that ours is concave down while the true data is concave up (Figure 13). This could be due to a variety of factors: we start with a large non-user population compared to a small user population and a small k value, which might account for the results acting strange; our values for d and r might not be correct; or the model might not be as accurate as we would like. It is a simple fix if the values for d and r are not the most accurate. However, it is more likely that our model simply doesn't fit the data like we want it to – our model initially increases faster than the data does. Furthermore, its increase slows down while the increase in the trend seen in the data speeds up.

6 Socio-Economic Trends

To explore specific socio-economic factors and their influence in total drug reports, specific counties that displayed disproportionate reports over the 2010-2017 period were selected, and variables of interest were explored in a multiple linear regression setting. The counties investigated were Hamilton County, Ohio and Fairfax County, Virginia which both demonstrate uncharacteristically high drug reports compared to their neighboring counties. The explanatory variables investigated focused on education, specifically the possession of a high school degree and/or a bachelor's degree. The model tested for both Fairfax and Hamilton was:

$$\begin{aligned} &\mu\{Total\ Drug\ Reports|Year, High\ School, Bachelor's\} \\ &= \beta_0 + \beta_1 Year + \beta_2 High\ School + \beta_3 Bachelors + \beta_4 High\ School : Bachelors. \end{aligned}$$

There was no significant evidence that any of the model parameters above were nonzero; the variation in the data was not significantly explained by these factors ($p > 0.5$). This was true despite year being included as an explanatory variable. This is echoed by the results in the plot of the autocorrelation function of overall total drug reports, where lag corresponds to one year, seen in Figure 14. We see that total drug reports alone for one year

do not possess predictive power for subsequent years ($ACF < 0.6$).

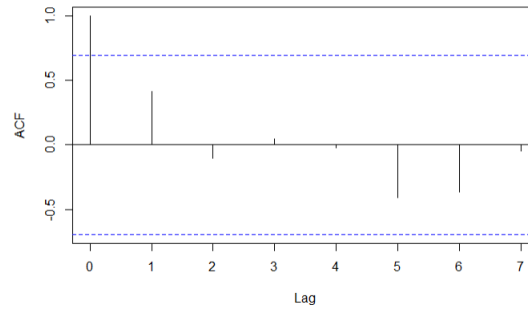


Figure 14: Auto Correlation Function. Lag correspond to Year.

7 Solutions to Problem

The simplest solution is to decrease each of the k , d , and r values. Decreasing the r values is likely the simplest; the government should invest in treatment and drug rehabilitation centers. This would prevent the sober population from falling back into the user population. Decreasing the k and d values is arguably more difficult. Educating the population on the dangers of drug addiction is certainly a step, as is providing simple and easy ways for people to get help.

Based on socio-economic factors tested, we can conclude that a model taking into account only education parameters does not account for the variation we see in opioid heavy counties. This could be evidence to focus more on regulation and recovery rather than forms of education and information about opioid abuse. We also see that time alone has barely any predictive power for future years' total drug reports.

8 Conclusion

After looking at our model and the data we conclude that with revisions our model could prove to be good at predicting future trends. Had we more time to work on this project we would have explored how altering the model would have altered its effectiveness. However, even with the time constraints,

we were able to produce a reasonably accurate model. Unfortunately, this model will likely not be very effective at predicting future trends. Though, we noted that there is very little correlation within the data in the first place, so this might not be entirely the fault of the model.

Our analysis of the data reveals that there are surprisingly few overall trends. We agree that as the populations of each state increases, so too does the number of drug incidents. On a county level, the overall trends vary; Virginia especially has unexplained spikes in the data. This is likely due to some outside factors. We note that there is no correlation between education and opioid use. Therefore, our solution to the crisis does not include increasing education. We instead recommend changing how we treat drug addiction.

If we had more time, we would test other socio-economic factors such as marital status, gender, age, and race. Furthermore, we would test the correlation of these factors between all of the counties. Unfortunately, as we were given over 600 different socio-economic variables, we did not have time to test them all.

Because we didn't have time to test the impact of the socio-economic factors on drug incidents, we were not able to alter our model to include the effects of the socio-economic factors. As it stands, because we were not able to find any correlation between education and opioid use, we did not change our model.

References

- [1] Quanyan Befekadu, Getachew K.; Zhu. Optimal control of diffusion processes pertaining to an opioid epidemic dynamical model with random perturbations. *Journal of Mathematical Biology*, 2018.
- [2] Gernot; Wallner Dagmar Caulkins, Jonathan P.; Tragler. Optimal timing of use reduction vs. harm reduction in a drug epidemic model. *International Journal of Drug Policy*, 20:480–487, 2009.
- [3] American Addiction Centers. Statistics on drug addiction, 2019. <https://americanaddictioncenters.org/rehab-guide/addiction-statistics>.
- [4] I. Greenhalgh, D.; Moneim. Sirs epidemic model and simulations using different types of seasonal contact rate. *Systems Analysis Modelling Simulation*, 43:573–600, 2003.
- [5] Addiction in Family. Length of addiction, 2019. <http://www.addictioninfamily.com/drugs/length-of-addiction/>.
- [6] A. G. Kermack, W. O.; McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115:700–721, 1927.

Appendix

We used the Julia package DifferentialEquations.jl to find numerical solutions for our model. We used the R packages ggplot2, Plyr, and tibble to analyze the data sets provided.

Julia

```
using DifferentialEquations
using Plots
function SIRI(du, u, p, t)
    du[1] = - (0.0044 / 32081000.0) * u[1] * u[2]
    du[2] = (0.0044 / 32081000.0) * u[1] * u[2] - (u[2] / 15.6) + u[3] * 0.5
    du[3] = (u[2] / 15.6) - u[3] * 0.5
end

u₀ = [32009510.0; 51490.0; 20000.0] #num of non-users, num of users, num of sober
tspan = (0.0, 15.0)
prob = ODEProblem(SIRI, u₀, tspan)
sol = solve(prob)

plot(sol, title="SIRI Model", xlabel="Time (years)", ylabel="Population (num of individuals)",
    label=["Non-Users", "Users", "Sober"], xlims=(0, 15), ylims=(0, 100000), xticks = 0:1:15)
```

Figure 15: Julia code used to solve and plot the SIRI model.

R

Here are selected pieces of our R code used to analyze

```
title: "initial_explore"
output: html_document
output: pdf
```

```
““{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(tibble)
library(ggplot2)
library(plyr)
““
```

```
## Various Plots for Initial Exploration
```

```
““{r}
# read in the data
library(readxl)
alldrugs <- read_excel("MCMNFLIS_Data.xlsx", sheet = "Data")
““

““{r}
# State Totals

ggplot(alldrugs, aes(YYYY, TotalDrugReportsState))+
  geom_line(aes(group = State, color = State))+
  geom_smooth(method = "loess", span = 0.2, size = 2)+
  ggtitle("State_Total_Drug_Reports_(2010-2017)")+
  xlab("Year")+
  ylab("Total_Drug_Reports")
““

““{r}
# store this model for use later

alldrugs.lo_fit <- loess(TotalDrugReportsState ~ YYYY, alldrugs)
““

““{r}
# County Totals

ggplot(alldrugs, aes(YYYY, TotalDrugReportsCounty))+
  geom_line(aes(group = COUNTY, color = COUNTY))+
  theme(legend.position = "none")+
  ggtitle("County_Total_Drug_Reports_(2010-2017)")+
  xlab("Year")+
  ylab("Total_Drug_Reports")
““

““{r}
```

```
# County Plots by State
```

```
# OH
```

```
alldrugs.OH <- subset(alldrugs, State == "OH")
```

```
ggplot(alldrugs.OH, aes(YYYY, TotalDrugReportsCounty))+  
  geom_line(aes(group = COUNTY, color = COUNTY))+  
  ggtitle("Ohio_County_Total_Drug_Reports_(2010-2017)") +  
  theme(legend.position = "none") +  
  xlab("Year") +  
  ylab("Total_Drug_Reports")  
““
```

```
““{r}
```

```
# KY
```

```
alldrugs.KY <- subset(alldrugs, State == "KY")
```

```
ggplot(alldrugs.KY, aes(YYYY, TotalDrugReportsCounty))+  
  geom_line(aes(group = COUNTY, color = COUNTY))+  
  ggtitle("Kentucky_County_Total_Drug_Reports_(2010-2017)") +  
  theme(legend.position = "none") +  
  xlab("Year") +  
  ylab("Total_Drug_Reports")  
““
```

```
““{r}
```

```
# PA
```

```
alldrugs.PA <- subset(alldrugs, State == "PA")
```

```
ggplot(alldrugs.PA, aes(YYYY, TotalDrugReportsCounty))+  
  geom_line(aes(group = COUNTY, color = COUNTY))+  
  ggtitle("Pennsylvania_County_Total_Drug_Reports_(2010-2017)") +  
  theme(legend.position = "none") +  
  xlab("Year") +  
  ylab("Total_Drug_Reports")  
““
```

```
““{r}
```



```
# VA
alldrugs.VA <- subset(alldrugs, State == "VA")

ggplot(alldrugs.VA, aes(YYYY, TotalDrugReportsCounty))+
  geom_line(aes(group = COUNTY, color = COUNTY))+
  ggtitle("Virginia_County_Total_Drug_Reports_(2010-2017)") +
  theme(legend.position = "none") +
  xlab("Year") +
  ylab("Total_Drug_Reports")
```

```{r}
# WV
alldrugs.WV <- subset(alldrugs, State == "WV")

ggplot(alldrugs.WV, aes(YYYY, TotalDrugReportsCounty))+
  geom_line(aes(group = COUNTY, color = COUNTY))+
  ggtitle("West_Virginia_County_Total_Drug_Reports_(2010-2017)") +
  theme(legend.position = "none") +
  xlab("Year") +
  ylab("Total_Drug_Reports")
```

Model Fitting

```{r}
alldrugs.model <- tibble(Year = 2010:2017, Data
  = c(51490, 56843, 59997, 61900, 63091, 63878, 64435, 64861))
alldrugs.rvals <- unique(alldrugs.lo_fit$fitted)

alldrugs.model$rvals <- unique(alldrugs.lo_fit$fitted)

model_rsqr <- cor(alldrugs.rvals, alldrugs.model$Data)

plot(alldrugs.rvals, alldrugs.model$Data, main
  = "Predicted_(Model)_vs._Fitted_(loess)_values", xlab
  = "Fitted_Values", ylab = "Predicted_Values")
```

```
abline(lm(alldrugs.model$Data ~ alldrugs.rvals), col = "red")
legend(60000,55000, legend = c("R^2=0.70"),col=c("red"))
““
```

```
““{r}
ggplot(alldrugs.model)+
  geom_line(aes(Year,Data), color = "red")+
  geom_line(aes(Year,rvals), color = "blue")+
  ylab("Predicted Values")+
  ggtitle("Model (Red) and Fit (Blue) Comparison")
““
```

County Based Parameter Exploration

```
““{r}
# read in the data
library(readr)
suppressMessages(censusData_2010 <- read_csv
  ("census_data/ACS_10_5YR_DP02_with_ann.csv", skip = 1))
suppressMessages(censusData_2011 <- read_csv
  ("census_data/ACS_11_5YR_DP02_with_ann.csv", skip = 1))
suppressMessages(censusData_2012 <- read_csv
  ("census_data/ACS_12_5YR_DP02_with_ann.csv", skip = 1))
suppressMessages(censusData_2013 <- read_csv
  ("census_data/ACS_13_5YR_DP02_with_ann.csv", skip = 1))
suppressMessages(censusData_2014 <- read_csv
  ("census_data/ACS_14_5YR_DP02_with_ann.csv", skip = 1))
suppressMessages(censusData_2015 <- read_csv
  ("census_data/ACS_15_5YR_DP02_with_ann.csv", skip = 1))
suppressMessages(censusData_2016 <- read_csv
  ("census_data/ACS_16_5YR_DP02_with_ann.csv", skip = 1))
““
```

```
““{r}
# Identify Counties of Interest based on model + trend
drugCounties.OH <- unique(subset
  (alldrugs.OH, TotalDrugReportsCounty > 5000)$COUNTY)
drugCounties.KY <- unique(subset
```

```

      (alldrugs.KY, TotalDrugReportsCounty > 1000)$COUNTY)
drugCounties.PA <- unique(subset
  (alldrugs.PA, TotalDrugReportsCounty > 5000)$COUNTY)
drugCounties.VA <- unique(subset
  (alldrugs.VA, TotalDrugReportsCounty > 1500)$COUNTY)
drugCounties.WV <- unique(subset
  (alldrugs.WV, TotalDrugReportsCounty > 750)$COUNTY)
'''

'''{r}
# Explore results of multiple linear regression for
parameters of interest for the model

# Hamilton, OH
hamilton_rpts <- as.numeric(
  unique(
    subset(alldrugs.OH, COUNTY == "HAMILTON",
      select = TotalDrugReportsCounty))
    $TotalDrugReportsCounty)

hamilton.dat <- tibble("Year" = 2010:2016, "Drug_Reports"
  = hamilton_rpts[1:7], "%HS" = hamilton_high_edu, "
  %College" = hamilton_college_edu)

hamilton.lm <- lm(hamilton.dat$`Drug Reports` ~
  hamilton.dat$Year + hamilton.dat$`%HS` +
  hamilton.dat$`%College` +
  hamilton.dat$`%HS`:hamilton.dat$`%College`)

summary(hamilton.lm)

'''

# Time is not a significant explanatory variable?

'''{r}

totaldrugs <- c(0,0,0,0,0,0,0,0,0)

```

```
for (i in 1:8){
  totaldrugs[i] <- sum(unique(subset(alldrugs ,
  YYYY == 2009 + i , select
  = TotalDrugReportsState))$TotalDrugReportsState)
}

totaldrugs.ts <- ts(totaldrugs , start = 2010, end = 2017)
acf(totaldrugs.ts)

'''
```