

To: Executive Council, Purdue Pharma
From: CMP Data Analysis

Esteemed Council,

As you are no doubt aware, a wave of opioid-based litigation is currently threatening companies in our sector. Purdue Pharma has weathered the storm, but every day brings a new State Attorney General who is looking to assign blame for illicit drug use. The recent actions by Attorney General Josh Stein from North Carolina should be of particular concern¹. We believe he is readying a massive lawsuit against Purdue Pharma. *In preparation, this memorandum serves to answer three questions:*

- **Is there reason to believe that major pharmaceutical companies have caused, either directly or indirectly, the opioid crisis in North Carolina?**
- **What socioeconomic factors most correlate with high overdose rates?**
- **What action can Purdue Pharma take to dampen the effects of the opioid crisis in North Carolina?**

We will answer these questions by looking at data available from the Centers for Disease Control and Prevention, The Robert Wood Johnson Foundation, the University of Wisconsin Public Health Institute, and the North Carolina Department of Health and Human Services. The data examines overdose rates and socioeconomic factors at the county level for North Carolina.

Background:

According to the CDC, there have been in excess 700K overdose deaths attributed to the opioid epidemic in the United States since 1999². Several major pharma companies are currently involved in litigation over their role in fueling this epidemic. Portions of the rural south east have been hardest hit. This memo will examine opioid overdoses across the 100 counties that comprise North Carolina. We intend to determine if the major pharmaceutical companies contributed to the opioid epidemic by encouraging the over prescription of pain medication. In order to do this, we must first make a few key assumptions.

If recent accusations are to be believed, there is a direct link between pressure from pharmaceutical and over prescription of pain medication by physicians. That is to say, public sentiment seems to indicate that pharma companies have incentivized doctors to prescribe

¹ <https://thehill.com/policy/healthcare/461779-north-carolina-sues-sackler-family-over-opioid-epidemic>

² <https://www.cdc.gov/drugoverdose/data/statedeaths.html>

opioids. While Purdue Pharma disputes this claim, for the purposes of this paper, we will assume that link exists. Additionally, we will assume that the rate at which opioids are prescribed is directly related to the amount of pressure from pharmaceutical companies³.

The majority of our data came from two different sources. The first data was provided by the Robert Wood Johnson Foundation in conjunction with the University of Wisconsin. This data primarily focused on demographic information for each county (total population, percentages by race, firearm-related crime death rates, etc.). When necessary, we normalized the data based on population. We focused on data from the 2018 calendar year, since prior years were largely incomplete.

The second data set we used for the analysis came from the North Carolina Department of Health and Human Services. This data focused mainly on opioid prescription and overdose information (emergency room visits that received an opioid overdose diagnosis, the number of opioid pills dispensed, etc.). The data set is organized by county and by fiscal year quarter from 1999-2019. We focused only on the 2017 and 2018 data sets, as they were the most complete.

Both sets contained data broken down by county, but the physical layout was largely dissimilar. We chose to join the data in excel before converting it to a JMP file. We used a combination of pivot tables and the “vlookup” functionality to create a uniform data set. We converted the information from a quarterly format into an annual format by either summing data across four quarters, or by taking the average data across the four quarters. Further, we normalized the data based on the county’s population when appropriate.

As an example, we determined the number of opioid pills as it relates to the population by taking the sum of the total number of pills prescribed in a given county for the four quarters of 2018. Next, we divided the value by the county’s population to determine the average number of pills dispensed per-person in each county. Some of the variable in the supplied data set were metrics like “percent of opioid deaths involving heroin or fentanyl/fentanyl analogues.” For these, we simply calculated the average percentage of the four quarters to determine the average value for 2018.

From there we began analyzing the data. Using the multivariate method we identified and marked for exclusion any independent variable that displayed multicollinearity. We also ran descriptive statistics in an attempt to understand the shortcomings of our data. Specifically, we looked at univariate statistics via the distribution functionality in SAS JMP. The raw data was, in general, complete and mostly free of outliers. The outliers that were present did not appear truncated, so we did not remove or modify them.

³ We realize this isn’t a valid assumption, but data for this link is extremely hard to come by because of HIPPA compliance.

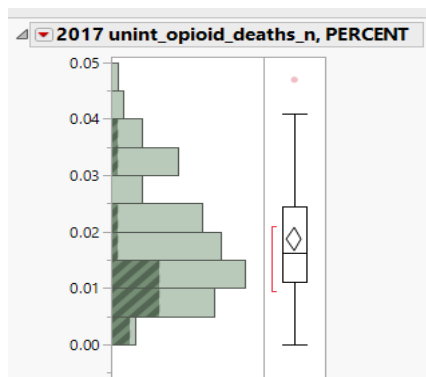


Figure 1: Example of a Non-Truncated Outlier

Expected Findings

Purdue Pharma has been convicted in the court of public opinion, but very little quantitative evidence has been presented to date. We intend to see if there is a relationship between actions taken by pharma companies / medical providers and incidences of overdoses. We will specifically look at counties in North Carolina. The popular narrative is that individuals are getting injured and doctors are overprescribing opioids for pain management. As a result, the individual develops an addiction to these pain medications.

This chain of events makes sense from a logical perspective, and it makes for sensational journalism, but it has not been proven quantitatively. As a byproduct of this research, we are hoping to be able to use cluster analysis to inform officials from North Carolina about higher risk segments within their state (based on county, income level, education level, ect). In this sense, we are creating something similar to the risk profiles developed by insurance companies.

Methods and Discussion of Results

For this analysis, we relied primarily on linear regression, logistic regression, and cluster analysis. The methodology of each approach is outlined in the following segments.

Generalized Linear Regression

As discussed previously, our data set included all of North Carolina's 100 counties. For each county, we started with approximately 93 different variables. In order to identify the most significant variables, we first created a validation column with the "formula random" functionality. 75% of the data was training, while the remaining 25% was for validation.

Next, we set up a Generalized Regression and chose “opioid ODs in 2018, Percentage” as the **dependent variable**; all the remaining variables were selected to be independent. We then added the validation column. For the estimation method, we selected “Pruned Forward Selection.” Once the analysis was completed, we noticed that only a handful of the independent variables had a non-zero estimate for their coefficients.

To determine the effect of the validation set on the results, we decided to repeat the analysis for 1,500 different randomly-generated validation sets. This was accomplished by right-clicking over the “Estimate” fields, selection “Validation” for both the Switch In/Switch out, and choosing 1,500 samples. **Appendix A** details the 20 independent variables that most frequently had a non-zero coefficient.

Armed with this information, we selected the top 15 variables with a stepwise regression approach using the full data set. The snapshots below show the list of variables with P-values less than 0.15. The top variables were: 2018’s unintended opioid deaths (percent of each county’s population), the high-school graduation rate, racial demographics (Hawaiian, a little more analysis is needed to determine whether this was an arbitrary correlation with no causation attached to it), social association rate, preventable hospitable stay rate, homicide rate, the previous year’s opioid OD percentage, overall drug overdose deaths, and the percent of uninsured adults. A detailed explanation of these independent variables is located in **Appendix B**.

Therefore, the independent variables that are most closely correlated with opioid overdoses for each county in 2018 were related to the life/community circumstances of individuals. There was no strong correlation between opioid overdoses and pharmaceutical pill prescriptions in the analysis.

Source	LogWorth	PValue
2018 unint_opioid_deaths_n, PERCENT	5.060	0.00001
High school graduation-Graduation Rate	3.214	0.00061
Demographics-% Native Hawaiian/Other Pacific Islander	3.137	0.00073
Social associations-Association Rate	3.089	0.00082
Preventable hospital stays-Preventable Hosp. Rate (by enrollees)	3.079	0.00083
Homicides-Homicide Rate	1.878	0.01326
2017 opioid, PERCENT	1.811	0.01545
Drug overdose deaths-Drug Overdose Mortality Rate	1.565	0.02725
Uninsured adults-% Uninsured	1.155	0.06996

Summary of Fit	
RSquare	0.671882
RSquare Adj	0.616164
Root Mean Square Error	0.023333
Mean of Response	0.070693
Observations (or Sum Wgts)	63

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	9	0.05908733	0.006565	12.0586
Error	53	0.02885570	0.000544	Prob > F
C. Total	62	0.08794303		<.0001*

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-0.066585	0.050825	-1.31	0.1946
2018 unint_opioid_deaths_n, PERCENT	0.8551157	0.545615	1.57	0.1217
Preventable hospital stays-Preventable Hosp. Rate (by enrollees)	0.000575	0.000285	2.02	0.0474*
Severe housing problems-% Severe Housing Problems	-0.003685	0.001663	-2.22	0.0300*
2017 opioid, PERCENT	0.2446223	0.078248	3.13	0.0026*
Mental health providers-MHP Rate	8.9889e-5	3.642e-5	2.47	0.0161*
Insufficient sleep-% Insufficient Sleep	0.0031472	0.001805	1.74	0.0858

Figure 2: Linear Regression Analysis

Logistic Regression Analysis

In the previous section, we discussed the variables that were most closely correlated with opioid overdoses in each county. In this section, we will explore which variables are correlated with counties that have relatively high rates of overdoses.

We calculated the weighted average and standard deviation of the opioid overdose (percent by population) for the state of North Carolina from the 100 counties. With this information, we determined which counties have an overdose rate that was (1) higher than the North Carolina mean, (2) 0.5 standard deviations higher than the North Carolina mean, and (3) 1 standard deviation higher than the North Carolina mean. For these scenarios, we repeated the validation simulation from the previous section using 1,500 validation samples. **Appendix C** details the variables that most frequently returned a non-zero coefficient estimate. It is worth noting that, for all three of the logistic regressions we ran, the volume of pills prescribed did not register as significant.

Detailed below are the results from counties with an overdose rate that was higher than the North Carolina mean. Using a logistic regression, we started off with the top 10 variables and removed those variables with a relatively high P-value. The most significant variables, in order, were (1) limited access to healthy foods, (2) 2018’s EMS naloxone usage (by percent of each county’s population), (3) 2018’s unintentional opioid deaths, (4) lower high school graduation rate, (5) percent of the population driving alone for long commutes, and (5) insufficient sleep. It appears that counties with higher than average rates of opioid addiction also have higher rates of high school dropout, and issues that contribute to poor mental health (long, lonely commutes and insufficient sleep).

Whole Model Test				
Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	20.833383	6	41.66677	<.0001*
Full	44.011217			
Reduced	64.844600			

RSquare (U)	0.3213
AICc	103.295
BIC	119.973
Observations (or Sum Wgts)	96

Lack Of Fit			
Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	89	44.011217	88.02243
Saturated	95	0.000000	Prob>ChiSq
Fitted	6	44.011217	0.5094

Source	LogWorth	PValue
Limited access to healthy foods-% Limited Access	4.187	0.00007
2018 ems_naloxone_n, PERCENT	2.648	0.00225
2018 unint_opioid_deaths_n, PERCENT	2.202	0.00628
High school graduation-Graduation Rate	1.133	0.07354
Long commute - driving alone-% Long Commute - Drives Alone	0.963	0.10879
Insufficient sleep-% Insufficient Sleep	0.949	0.11258

Parameter Estimates				
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	24.0234871	9.5071529	6.39	0.0115*
Limited access to healthy foods-% Limited Access	-0.3155315	0.0909011	12.05	0.0005*
2018 unint_opioid_deaths_n, PERCENT	-88.166578	34.172312	6.66	0.0099*
High school graduation-Graduation Rate	-0.1304812	0.0756423	2.98	0.0845
2018 ems_naloxone_n, PERCENT	-10.392733	3.8093569	7.44	0.0064*
Long commute - driving alone-% Long Commute - Drives Alone	-0.0487181	0.030723	2.51	0.1128
Insufficient sleep-% Insufficient Sleep	-0.1804518	0.1172668	2.37	0.1238

Figure 3: Logistic Regression Analysis

The confusion matrix is shown below. The model's hit rate was 78%.

Actual Count	Predicted	
	0	1
0	47	10
1	11	28

Figure 4: Logistic Regression Confusion Matrix

Cluster Analysis

We used hierarchical cluster analysis in an attempt to build an understanding of how the different counties relate to each other. The goal was to uncover some relationships that may be helpful in making our final recommendations. JMP indicated that 3-4 clusters were optimal; we eventually settled on 4 since in presented a clearer point of diminishing returns. Generalized details for each cluster are detailed below:

Cluster #	2018 opioid, PERCENT	Notes
Cluster 1	0.059774653	Very high rate of risky behaviors, Highest rate of pills prescribed by a doctor
Cluster 2	0.048135387	Most affluent, Most physically active,
Cluster 3	0.074073126	Very high rate of risky behaviors, Lowest number of pills prescribed by a doctor
Cluster 4	0.040324605	Highest unemployment, Highest level of income inequality, Second highest rate of pills prescribed by a doctor

Figure 5: General Trends Associated with Each Cluster

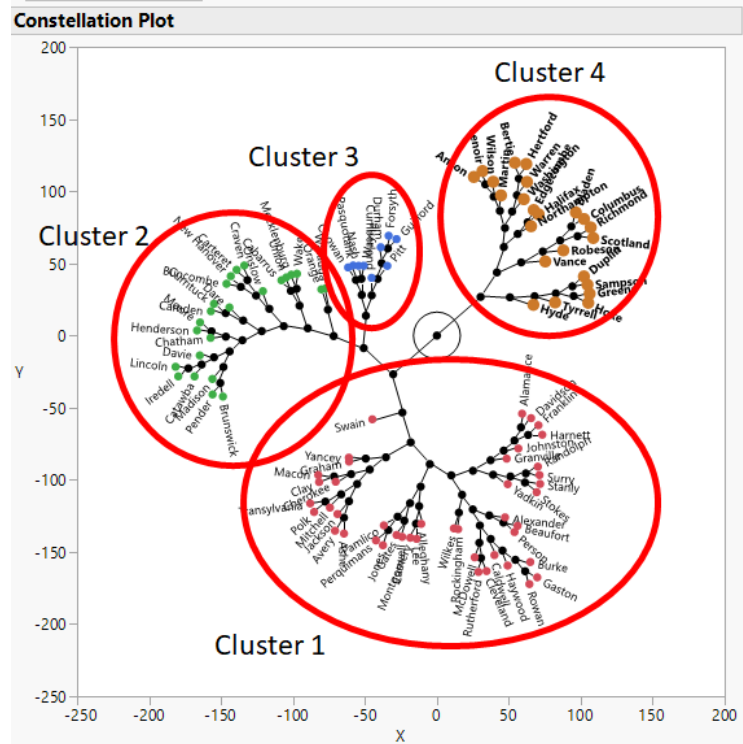


Figure 6: Constellation Plot

At this point, we have essentially agglomerated counties into different risk profiles. The goal is to understand some of the social factors that may be correlated. Noteworthy is the fact that the two clusters with the highest rates of opioid overdoses also have high incidents of risky behavior (HIV rates, violent crimes, STD rates). The rate of pills prescribed by a doctor does not appear to be a significant factor; the highest risk cluster has the lowest number of pills prescribed. Conversely, the lowest risk segment has the second highest rate of pills prescribed by a doctor. Additionally, drug dosage and incidents of drug seeking behavior did not appear to have any relationship with a given cluster’s risk profile. Finally, despite the common perceptions, we found no evidence that poverty was tied to opioid abuse.

Conclusion and Recommendations

We find no indication that opioid dosage, number of pills prescribed, injury rate, or incidence of multiple provider episodes correlate with opioid overdoses in North Carolina. This information is reinforced by the National Institute on Drug Abuse for the state of North Carolina. (Appendix D). Instead, we found that most of the predictors of opioid overdoses were tied to mental health, risky behaviors, violence, and, in limited cases, high school graduation rate. The factors that lead to high rates of opioid overdoses are wide ranging, but they do not include variables directly relating to doctors and pharmaceutical manufacturers.

If deemed necessary, the next step of this analysis is to join with an outside organization that is authorized to handle PHI. This would allow our researchers to reach beyond the county level into individual cases. Armed with such data, it would be possible to build individual risk profiles for opioid abusers. Such information could be shared with the North Carolina Department of Health and Human Services. Such an outreach would help clear Purdue Pharma's name and reinforce our commitment to safe pain management.

Appendix A: Linear Regression Independent Variables That Most Frequently Had A Non-Zero Coefficient

Rank	For Opioid OD Rates (Regular Regression) Variable	Percent of validation sets in which variable had non-zero estimate
1	2018 uint_opioid_deaths_n, PERCENT	66%
2	Firearm fatalities-Firearm Fatalities Rate	51%
3	Social associations-Association Rate	33%
4	Preventable hospital stays-Preventable Hosp. Rate (by enrollees)	31%
5	Severe housing problems-% Severe Housing Problems	26%
6	2017 opioid, PERCENT	24%
7	Drug overdose deaths-Drug Overdose Mortality Rate	22%
8	High school graduation-Graduation Rate	17%
9	Mental health providers-MHP Rate	17%
10	Demographics-% Native Hawaiian/Other Pacific Islander	13%
11	Insufficient sleep-% Insufficient Sleep	8%
12	2018 Average of fent_heroin_death_pct	6%
13	2018 ems_naloxone_n, PERCENT	5%

14	2018 opioid_dx_un, PERCENT	5%
15	Uninsured adults-% Uninsured	5%
16	Residential segregation - black/white-Segregation index	4%
17	Unemployment-% Unemployed	4%
18	2017 unint_opioid_deaths_n, PERCENT	2%
19	Driving alone to work-% Drive Alone	2%
20	Mammography screening-% Mammography	2%

Appendix B: Variable Descriptions

Variable	Description
2018 opioid, PERCENT	Number of Emergency Room visits that received an opioid overdose diagnosis (all intents) as a percentage of the county's population.
2018 unint_opioid_deaths_n, PERCENT	Number of unintentional opioid related deaths as a percentage of the county's population.
2017 unint_opioid_deaths_n, PERCENT	Number of unintentional opioid related deaths as a percentage of the county's population.
2017 pills_n PER PERSON	Number of opioid pills dispensed per person in the county.
2017 opioid, PERCENT	Number of Emergency Room visits that received an opioid overdose diagnosis (all intents) as a percentage of the county's population.
2018 opioid_dx_un, PERCENT	Number of uninsured individuals and Medicaid beneficiaries with an opioid use disorder served by treatment plans as a percentage of the county's population
2018 bupronorphine_scripts_n, PERCENT	Number of buprenorphine prescriptions dispensed as a percentage of the county population. Buprenorphine is an opioid used to treat opioid addiction.
2018 comm_naloxone_n, PERCENT	Number of community naloxone reversals as a percentage of the county population. Naloxone (also known as Narcan) is a medication called an "opioid antagonist" used to counter the effects of opioid overdose

2018 acute_hep_c_n, PERCENT	Number of acute hepatitis C cases as a percentage of the county population. Ecological studies link increases in hepatitis C virus (HCV) infection to the U.S. opioid crisis.
2018 Average of fent_heroin_death_pct	Percent of opioid deaths involving heroin or fentanyl/fentanyl analogues.
2018 Average of multiple_episodes_n	Average rate of multiple provider episodes for prescription opioids, per 100,000 residents. MPEs are intended to apply to patients who see multiple prescribers in an attempt to obtain more medication than a single provider would provide.
2018 Average of patients_90mme_pct	Percent of patients with an opioid prescription receiving more than an average daily dose of 90+ MME of opioid analgesics. The CDC advises clinicians to prescribe the lowest effective dose, and requires justification to increase dosage to > 90 MME a day.
2018 Average of rxdays_opioidandbenzo_pct	Percent of prescription days any patient had at least one opioid AND at least one benzodiazepine prescription on the same day. The CDC recommends that clinicians avoid prescribing benzodiazepines. Previous studies have also highlighted the dangers of co-prescribing opioids and benzodiazepines concurrently with opioids whenever possible.
2018 pills_n PER PERSON	Number of opioid pills dispensed per person in the county.
Poor or fair health	Percent of adults that report fair or poor health
Poor physical health days	Average number of reported physically unhealthy days per month
Poor mental health days	Average number of reported mentally unhealthy days per month
Low birthweight	Value reported but considered unreliable since based on counts of twenty or less.

Adult smoking	Percentage of adults that reported currently smoking
Adult obesity	Percentage of adults that report BMI \geq 30
Food environment index	Indicator of access to healthy foods - 0 is worst, 10 is best
Physical inactivity	Percentage of adults that report no leisure-time physical activity
Access to exercise opportunities	Percentage of the population with access to places for physical activity
Excessive drinking	Percentage of adults that report excessive drinking
Alcohol-impaired driving deaths	Number of alcohol-impaired motor vehicle deaths
Sexually transmitted infections	Number of newly diagnosed chlamydia cases per 100,000 population
Teen births	Number of births per 1,000 female population ages 15-19
Uninsured	Number of people under age 65 without insurance
Primary care physicians	Number of primary care physicians (PCP) in patient care
Dentists	Ratio of population to dentists
Mental health providers	Number of mental health providers (MHP)
Preventable hospital stays	Number of Medicare enrollees
Diabetes monitoring	Number of diabetic Medicare enrollees
Mammography screening	Number of female Medicare enrollees age 67-69

High school graduation	Percentage of ninth-grade cohort that graduates in four years.
Some college	Adults age 25-44 with some post-secondary education
Unemployment	Number of people ages 16+ unemployed and looking for work
Children in poverty	Percentage of children (under age 18) living in poverty
Income inequality	Ratio of household income at the 80th percentile to income at the 20th percentile.
Children in single-parent households	Number of children that live in single-parent households
Social associations	Number of membership associations per 10,000 population. Poor family support, minimal contact with others, and limited involvement in community life are associated with increased morbidity and early mortality.
Violent crime	Number of reported violent crime offenses per 100,000 population.
Injury deaths	Number of deaths due to injury per 100,000 population
Air pollution - particulate matter	Average daily amount of fine particulate matter in micrograms per cubic meter
Drinking water violations	County affected by a water violation: 1-Yes, 0-No
Severe housing problems	Number of households with at least 1 of 4 housing problems: overcrowding, high housing costs, or lack of kitchen or plumbing facilities
Driving alone to work	Percentage of workers who drive alone to work
Long commute - driving alone	Number of workers who commute in their car, truck or van alone

Appendix C: Logistic Regression Independent Variables That Most Frequently Had A Non-Zero Coefficient

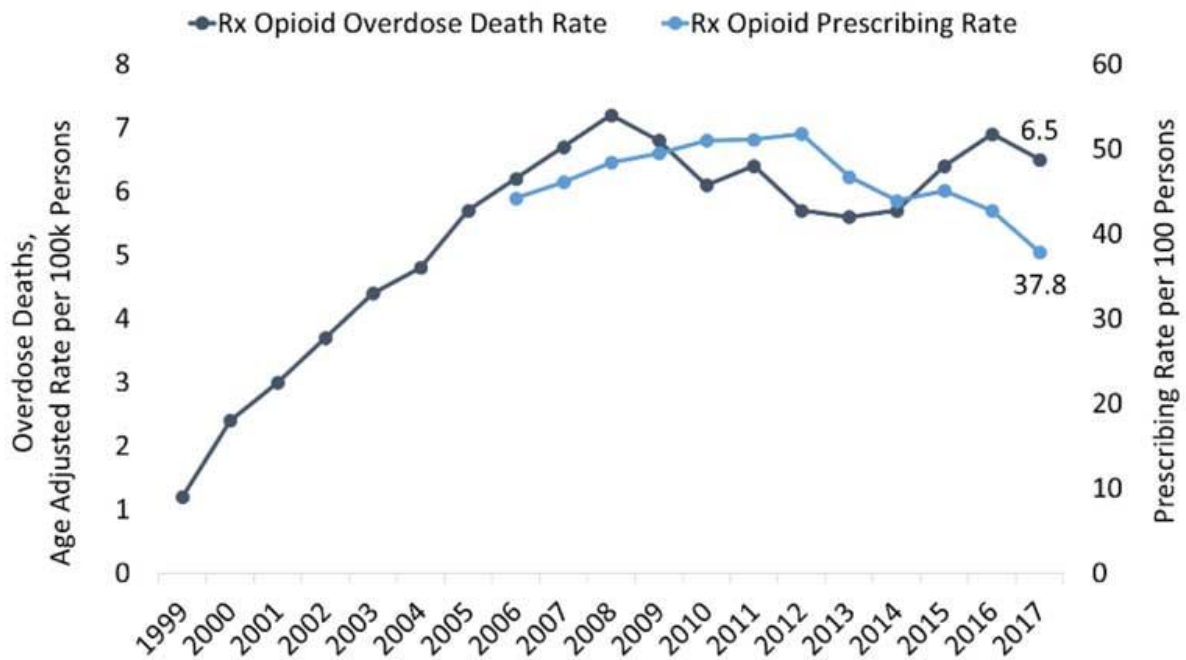
Rank	For Counties that had Opioid OD rates higher than the North Carolina mean	Percent of validation sets in which variable had non-zero estimate
1	2017 opioid, PERCENT	34%
2	Limited access to healthy foods-% Limited Access	32%
3	2018 ems_naloxone_n, PERCENT	18%
4	2018 unint_opioid_deaths_n, PERCENT	18%
5	High school graduation-Graduation Rate	12%
6	Insufficient sleep-% Insufficient Sleep	11%
7	Disconnected youth-% Disconnected Youth	9%
8	2018 Average of fent_heroin_death_pct	6%
9	Firearm fatalities-Firearm Fatalities Rate	5%
10	Long commute - driving alone-% Long Commute - Drives Alone	5%

Rank	For Counties that had Opioid OD rates more than 0.5 Standard Deviations Above the North Carolina mean	Percent of validation sets in which variable had non-zero estimate
-------------	--	---

1	2018 ems_naloxone_n, PERCENT	47%
2	Preventable hospital stays-Preventable Hosp. Rate (by enrollees)	20%
3	Severe housing problems-% Severe Housing Problems	15%
4	2017 opioid, PERCENT	14%
5	Residential segregation - black/white-Segregation index	11%

Rank	For Counties that had Opioid OD rates more than 1 Standard Deviation Above the North Carolina mean	Percent of validation sets in which variable had non-zero estimate
1	2017 opioid, PERCENT	28%
2	2018 ems_naloxone_n, PERCENT	28%
3	Firearm fatalities-Firearm Fatalities Rate	26%
4	2018 unint_opioid_deaths_n, PERCENT	9%
5	Mammography screening-% Mammography	6%

Appendix D: North Carolina Rate of Overdose Deaths Involving Prescription Opioids and The Opioid Prescribing Rate⁴



⁴ <https://www.drugabuse.gov/opioid-summaries-by-state/north-carolina-opioid-summary>

Kaiser Permanente: 2018 National Decrease in Prescription Opioids, large increase in Heroin and Fentanyl.

How do the economic dominos fall?

