Evaluating Ohio's Opioid Overdose Epidemic with AI

Zachary Elsawy, Katherine Elkins, and Jon Chun

Kenyon College, Gambier, OH

IPHS Senior Research Seminar, Fall 2021

Introduction

The epidemic surrounding unintentional drug poisoning in the United States is not showing signs of slowing down. The American Physical Therapy Association report every day, more than 1,000 people are treated in emergency departments for misusing prescription opioids. Being an Ohio resident for my entire life and losing a close friend to a drug overdose has quickly taught me the turmoil and detriment that drugs create. Over the past decade, family members and communities have struggled mightily with addiction, whether from opiates, heroin, cocaine, or other harmful substances. In 2019, Ohio had the 3rd highest drug overdose death rate, paired with the 5th highest opioid overdose death rate in the country (Kaiser Family Foundation). Today, these unfortunate rates remain alarming high (Figures 1 and 3). Prior studies have been able to identify individual-level characteristics associated with drug use, abuse, and overdose. However, Bozorgi et al., 2021 note "most of these studies have been conducted in unique populations (e.g., people injecting drugs, Medicaid recipients, veterans, and private insured populations) that may not generalize well with other US populations". Therefore, my study seeks to identify individual-level predictors of drug overdose by using machine learning. Additionally, I believe my dataset provides a more representative sample of a population compared to previous literature.

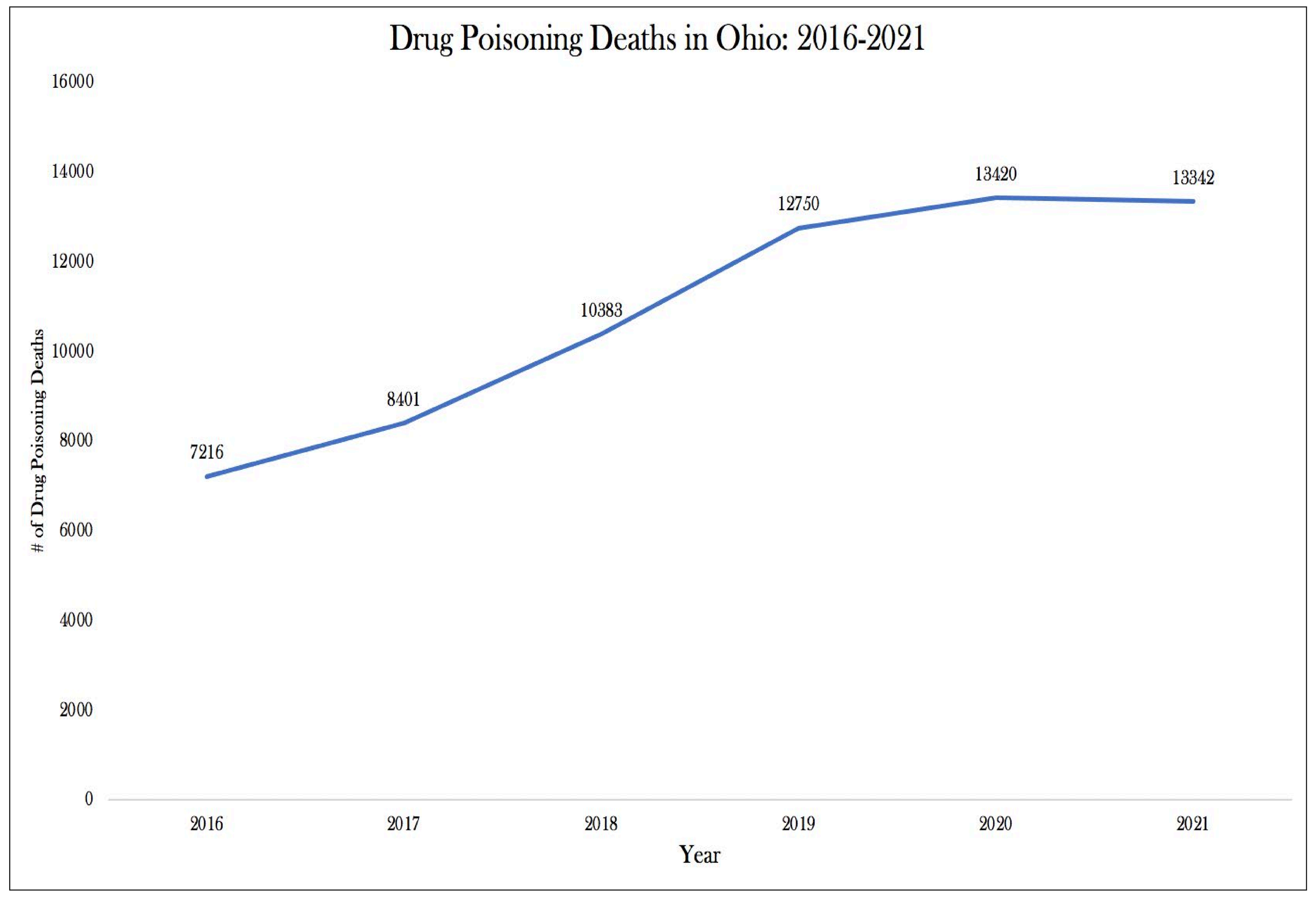


Figure 1: Drug Poisoning Deaths in Ohio: 2016-2021

Literature Review

In previous studies, many individual-level characteristics have been associated and identified with drug use, abuse, and overdose. Martins et al., (2015) find individuals with "low incomes or insecure housing, without a high school diploma, or recent release from prison are at increased risk for drug use and overdose." Additional findings help shape the selection process for variables in my machine learning model. For example, Johnson and Shreve et al., (2020) find overdose deaths tend to be higher in areas with higher concentrations of vacant and run-down buildings since unattended spaces provide good places for illicit drug use. Galea et al., (2005), among others, find population density may also affect substance abuse and the risk of overdose. Furthermore, individuals are less likely to consume illicit drugs in areas with a higher level of socialization, such as a densely populated urban community. According to King et al., (2014), "rural residents are more susceptible to opioid overdose and overdose deaths than urban residents." According to the United States Census Bureau, "rural areas comprise open country and settlements with fewer than 2,500 residents." In addition, García et al., (2019) and Keyes et al., (2014) reveal prescription rates for opioids are higher in rural areas than their urban counterparts. Depression is also linked to opioid and drug addiction (Cleland et al. 2020).

Methodology

Data Collection

The source of my dataset is www.countyhealthrankings.org, a website hosted by The University of Wisconsin Population Health Institute. Each year, the program provides health and demographic information on individuals at the state and county level. I use data from all eighty-eight Ohio counties in 2021 for my study. The dataset contains 514 variables, each categorized by specific health factors. I select nine independent variables from the Figure 2 health factors and run them against my dependent variable: drug poisoning deaths. All data cleaning and analysis is via Python 3.10.0.

Metadata

Regarding metadata, the fields consist of lists. Some of these lists contain text: whereas others have numeric values. The data types are integer (int) and string (str). I also assume the dataset derives from a normal distribution. There are no missing values for the independent variables selected. However, there are six counties, i.e., missing values, with data on opioid deaths that are still unreleased. I do not expect the missing values to hinder my regression model.

Variable Selection

Based on prior literature and the opinion of a medical professional, I chose nine variables to predict drug poisoning deaths. Figure 2 represents the County Health Rankings Model. To create a fully representative model, I choose variables from each "bucket" of health factors (clinical care, health behaviors, social and economic factors, and physical environment). These variables include high school graduation rate, the average number of mentally unhealthy days, age-adjusted death rate. Other variables, such as rural, uninsured, severe housing problems, unemployment, excessive drinking, and smoking, are expressed as a percentage.

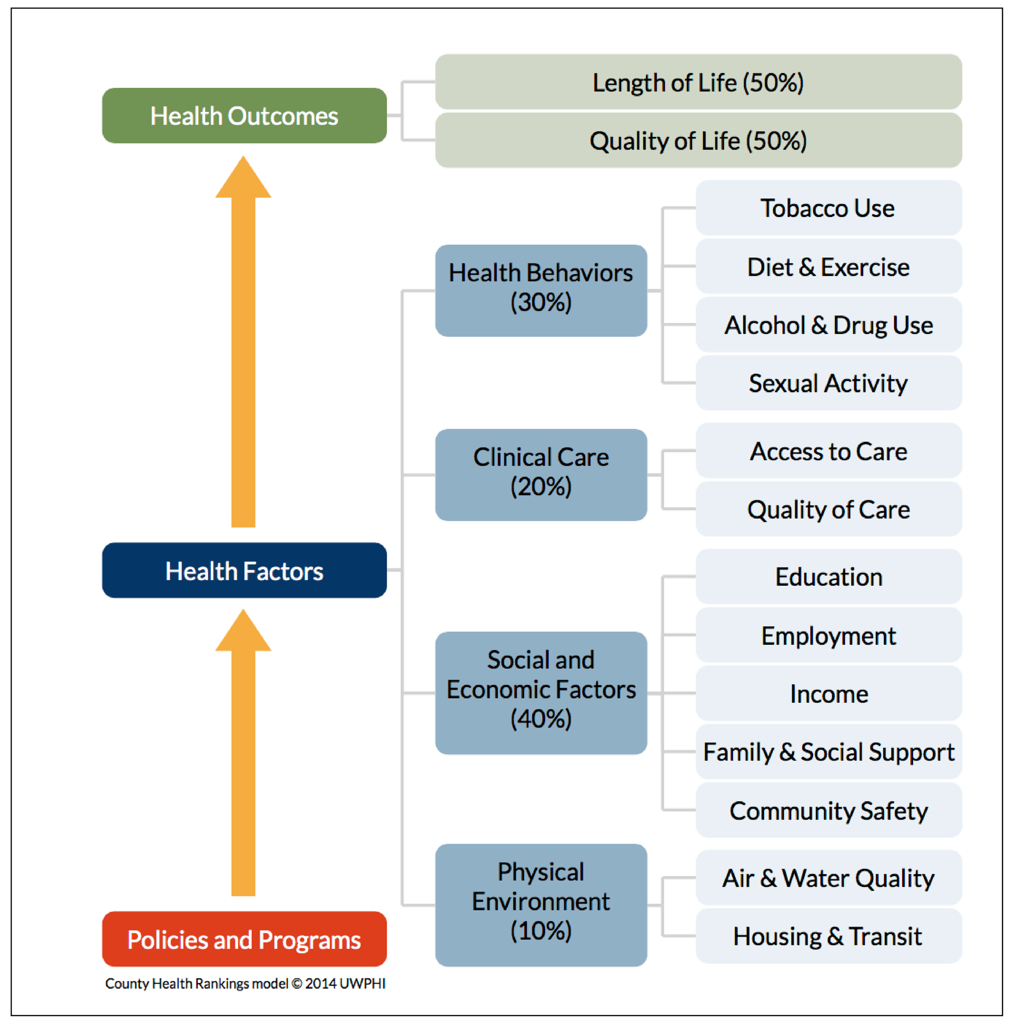


Figure 2: County Health Rankings Model

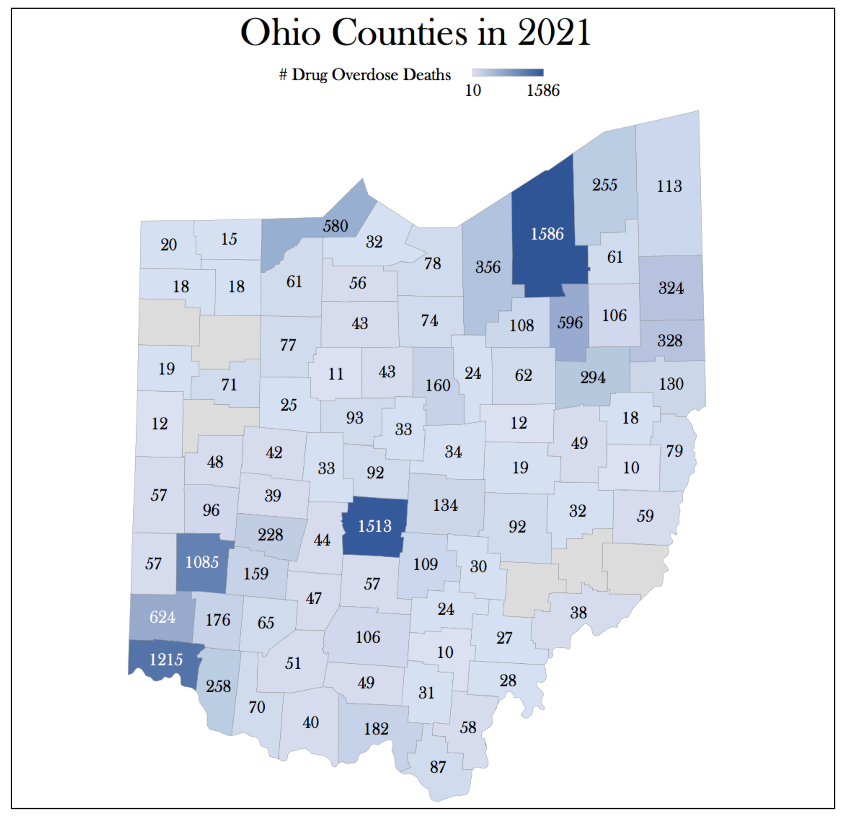


Figure 3: Drug overdose deaths per county in Ohio (2021)

Results

Tuning is the ML process of optimizing a model's performance. Machines select parameters that impact the model to enable the algorithm to perform more effectively. The top ML models for my dataset are AdaBoost Regressor, Huber Regressor, and Passive-Aggressive Regressor. I find the tuned AdaBoost regression model is the best of the three due to its relatively high R^2 (.73) and its MAE of 77.2.

In mathematics, R^2 refers to how well a regression model fits the observed data. For example, AdaBoost's R^2 is 73%, revealing 73% of the variability in drug overdose deaths can be explained by the variables in the model. Typically, higher R^2 values indicate a better fit for the model. Being said, this does not imply low R^2 values cannot still produce statistically significant results.

Mean average error (MAE) reveals how big an error one can expect from the forecast on average. The lower the MAE, the better. Of the top three tuned regressions, AdaBoost Regressor yields the lowest MAE.

Figure 4 reveals predictors of drug overdose according to the AdaBoost model. The top five predictors of drug poisoning deaths are rural (%), uninsured (%), the average number of mentally unhealthy days, severe housing problems (%), and unemployed(%). The remaining four predictors, like age-adjusted death rate, high school graduation rate, excessive drinking (%), and smokers (%), also provide significance to the model. These variables likely have a correlation, which may influence the findings.

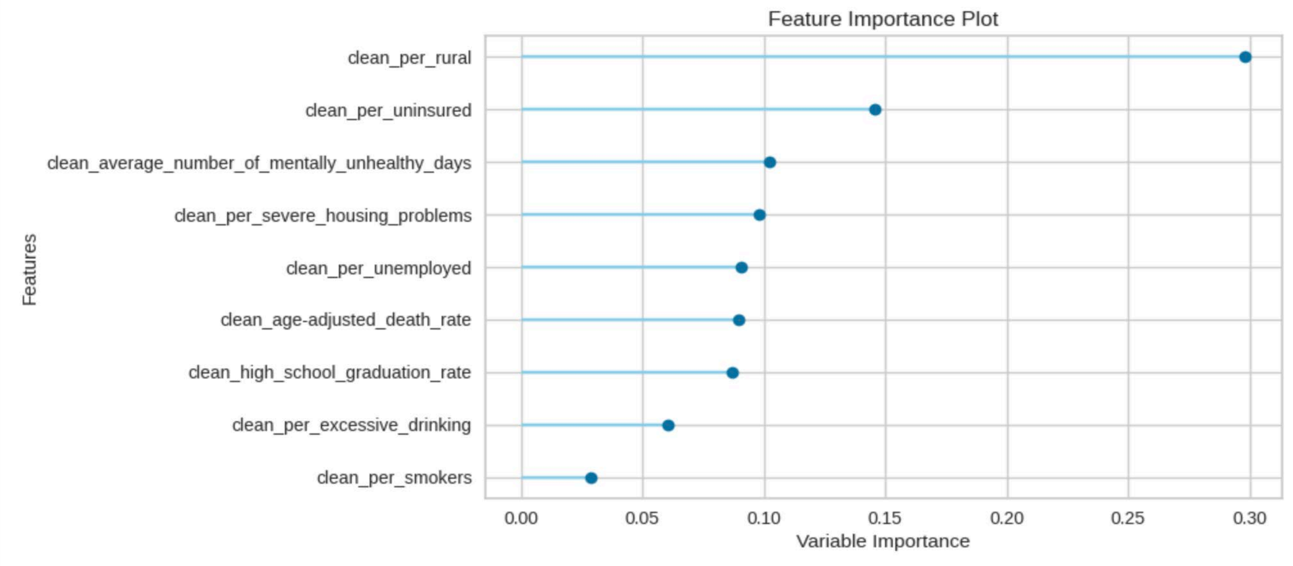


Figure 4: AdaBoost Feature Importance Plot: Predictors of Drug Poisoning Deaths

Conclusions

Since the early 1990s, machine learning models have flourished in popularity: due to their ability to create the most accurate predictions possible. The top ML models for my dataset were AdaBoost Regressor, Huber Regressor, and Passive-Aggressive Regressor. I found the AdaBoost regression model to be the best of the three.

According to the AdaBoost model, the most important predictors of drug overdose are rural (%), uninsured (%), the average number of mentally unhealthy days, severe housing problems (%), and unemployed(%). Other predictors like age-adjusted death rate, high school graduation rate, excessive drinking (%), and smokers (%) still provide significance to the model.

My results regarding rural communities and drug overdose are consistent with King et al. (2014), among others. Additional predictors in my model, such as poor mental health, appear to align with past literature as well. Since samples in prior studies derive from unique populations, my data better generalizes with the United States population. It is imperative to understand that many factors contribute to opioid addiction. Life is full of stressors that may expedite poor decision-making. Illicit drug use may not always correlate with demographics or health factors. However, I believe my study provides a data-driven argument that certain factors play a role.

Avenues for future research should consider assigning fewer variables to the model. Also, it may be beneficial to ask more professionals with domain expertise to extrapolate my findings. Given their day-to-day experience, it is possible their interpretation of my model, its variables, or findings, may differ.

I believe my study provides beneficial insight into the realm of drug abuse and overdose. One day, I wish to live in a world where public policy and treatment programs are better-suited for individuals suffering from drug addiction.

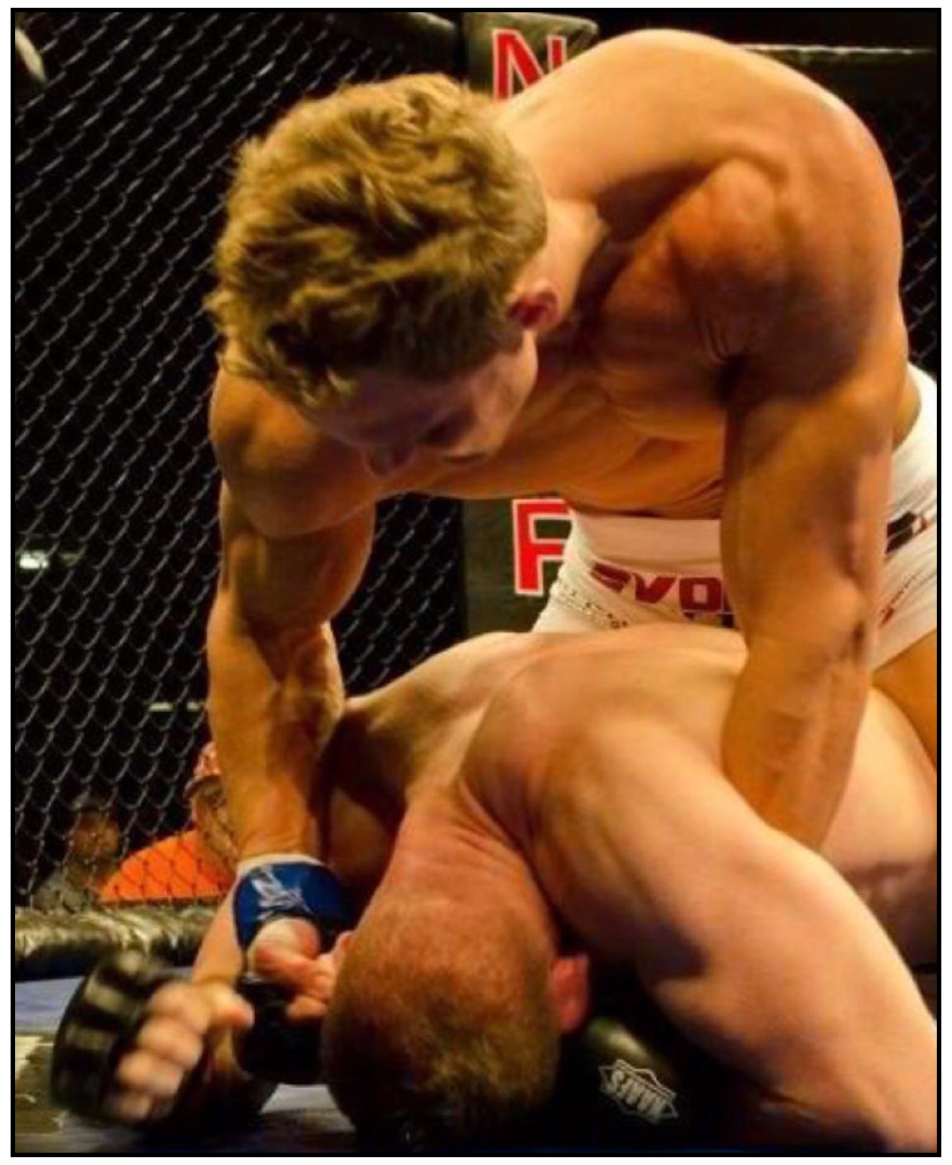
Discussion

Rose DeRoia, the mother of my late friend Tommy, lost her son to an opiate overdose in May of 2016. Recently, she interviewed on a YouTube podcast called Chronic Curiosity. She describes her personal experience regarding hardships her family faced during Tommy's six-year battle with addiction. Tommy was not only an undefeated MMA champion, college graduate, and successful businessman, but he was a beloved son, brother, and friend to all.
It started with a Xanax prescription, then led to a poor choice, and unfortunately, in a short period, his brain developed a chronic and relapsing disease.

She states, "if you start using opiates today, within three weeks, you will develop receptors in your brain that drive and direct you. If you stop using opiates, it takes an average of nine months for the brain circuitry in charge of decision-making, control, etc., to get back into place. Now, what do we see with rehab facilities? Four to six weeks if they have availability." Ohio must continue to expand access to opioid treatment centers for its constituents. I believe individuals suffering from drug addiction require more access to enhanced treatment and detoxification centers, recreational facilities, and facilities that distribute naloxone (a medicine that rapidly reverses an opioid overdose). More robust prescription drug monitoring programs seem to be beneficial as well.

Rose believes, "addiction is a 'disease' in the notion that the brokenness in which people cannot see, and do not want to see, is in the brain." Communities are too quick to stigmatize and judge those struggling with addiction. Most often, people lose sight that the disease is the one directing and controlling an addict's decisions. Society must break the stigma that addiction is a choice. Like any mental illness, it is a disease that needs consistent treatment.

In loving memory of Tommy DeRoia



Acknowledgements

Bozorgi, P. Eberth, E. Karami, I. (2021). The leading neighborhood-level predictors of drug overdose: A mixed machine learning and spatial approach, *Drug and Alcohol Dependence*, Volume 229, Part B.
Cleland, A. Bennett, S. Elliott, L. Rosenblum, A. (2020). Between-and within-person associations between opioid overdose risk and depression, suicidal ideation, pain severity, and pain interference, *Drug Alcohol Depend.*, 206, Article 107734.
Galea, S. Rudenstine, S. Vlahov, D. (2005). Drug use, misuse, and the urban environment
Drug Alcohol Rev., 24 pp. 127-136.
García, M. Heilig, C. Lee, S. Faul, M. (2019). Opioid prescribing rates in nonmetropolitan and metropolitan counties among primary care providers using an electronic health record system—United States, *Morb. Mortal. Wkly. Rep.*, 68 (2019), pp. 25-30
Johnson, L., Shreve, T. (2020). The ecology of overdose mortality in Philadelphia
Health Place, 66 ,Article 102430
Keyes, K. Cerdá, M. Brady, J. Havens, J. Galea, S. (2014). Understanding the rural–urban differences in nonmedical prescription opioid use and abuse in the United States, *Am. J. Public Health, 104* , pp. e52-e59
King, B. Fraser, V. Boikos, C. Richardson, R. Harper, S. (2014). Determinants of increased opioid-related mortality in the United States and Canada, 1990–2013: a systematic review
Am. J. Public Health, 104, pp. e32-e42
https://www.youtube.com/watch?v=_5XJX3JIkQg
https://www.kff.org/other/state-indicator/opioid-overdose-death-rates