The opioid epidemic: socioeconomic factors and opioid prescription rates

Team Opi-nope (Team 3): Steve Leung, Ningrui Li, Colleen Rhoades, Patrick Ye

Tens of thousands of Americans die every year due to the opioid addiction epidemic. Many patients become addicted when prescribed an opioid for pain management. Once addicted, people are at risk of dying from an opioid overdose or from overdosing on a substitution drug like heroin. In response, states are adopting measures to combat the opioid crisis, including limiting opioid prescriptions.

We investigated which health and socioeconomic factors were most relevant to the number of drug overdose deaths. We used data from 2015 to predict the number of deaths in the following year on a county-level basis. Furthermore, we compared statewide annual trends for both drug overdose deaths and the number of opioid prescriptions. We found that as the opioids prescription rate decreases, counterintuitively, the number of drug overdose deaths continues to increase.

Data Wrangling: We drew from several data sources to create a cleaned dataset for the years 2015 and 2016. We supplemented the county health rankings and census data with an external dataset containing opioid prescription rates by county. We verified the drug overdose deaths values with an external dataset pulled from wonder.cdc.gov. This dataset gave us ~3,000 more drug overdose data points than provided in the county health rankings. All features were normalized by county population and missing values were replaced with the median of other values for that feature.

Modeling: We investigated whether we could predict the 2016 drug overdose death rate using a model trained only on data from 2015. We chose to use a linear regression model with L1 regularization (LASSO) to obtain an interpretable model with a sparse set of features, compared to the 598 features in the cleaned dataset. We used the regularization parameter $\alpha = 0.1$ in order to balance sparsity of features with model performance and achieved a test r^2 of 0.4225. Our performance was better than a study by Clemans-Pope et al. 2018, who achieved a test r^2 of 0.4046.

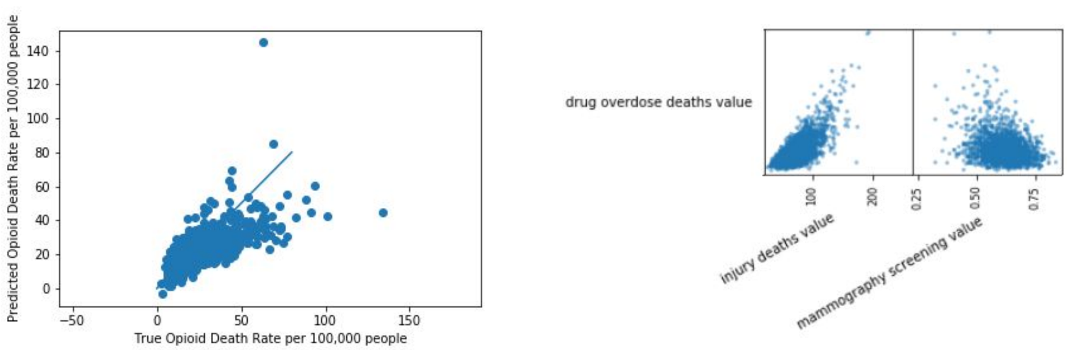

Figure 1: Model predictions in relation to the testing data (left). Most positively and negatively correlated variables with drug overdose deaths

As a comparison to understand how challenging our future prediction task was, we split the data into equal training and testing sets regardless of the year. Using $\alpha = 0.5$, the test r^2 was 0.5169. While this is an improvement over the previous model, performing a standard prediction problem did not yield impressive results. This highlights the challenge of predicting response variables in future years.

Injury deaths normalized by population had the largest coefficient in our linear model - it was also highly positively correlated with drug overdose deaths (Figure 1b). We have two possible explanations: people at high risk for injuries may also receive more surgeries and thus be prescribed more opioids. Alternatively, those who are abusing opioids are more likely to experience fatal injuries due to their altered mental state. In contrast, obtaining mammograms was negatively correlated with drug overdose deaths (Figure 1b). People who seek out preventative screenings, such as a mammogram, are less likely to make negative health decisions. Some other features that were used by the model related to: housing (age of housing complex, size of housing complex, having a mortgage, air pollution), personal health (insufficient sleep, disability status), and race (non-Hispanic white). Additional features that were strongly negatively correlated with drug overdose deaths were: employment status, years of residency in the US, nationality, college education, English-speaking proficiency, and comically, enrollment in Kindergarten.

Opioid prescription rates negatively correlated with drug overdose death rates: We found that as the opioid prescription rate decreased from 2012 to 2016, the number of drug overdose deaths actually increased in the same time period (Figure 2a). This was true for 43 states. For the remaining states, the prescription rate either increased or the drug overdose rates decreased (Figure 2b).

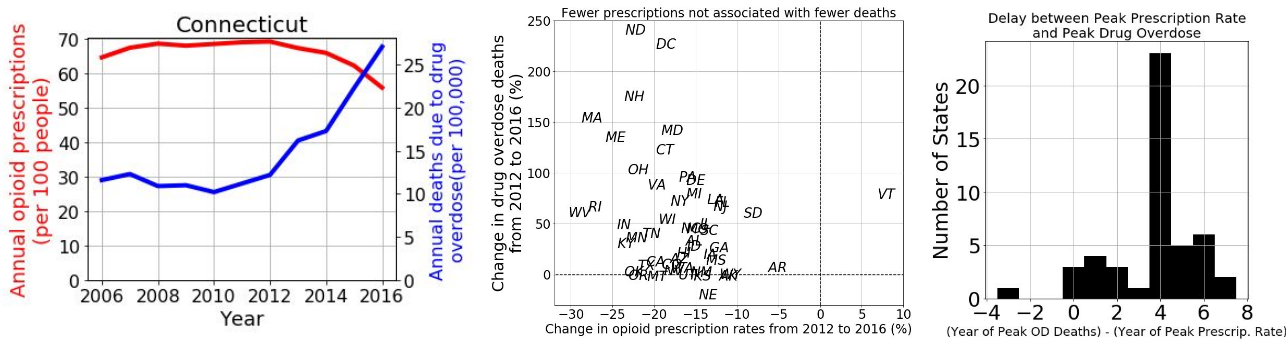

Figure 2: Comparing opioid prescription rates with drug overdose deaths by state from 2006 to 2016.

While keeping in mind that correlation does not equal causation, we offer a couple of potential explanations for this widespread phenomenon. A causative explanation: if former patients cannot continue obtaining their prescription opioids, this could lead them to take more dangerous alternatives like illicit fentanyl and heroin. Alternatively, it could take time (at least 4 years, Figure 2c) for a decrease in opioids prescriptions to have an effect on overdose deaths. Another explanation is that decreasing prescription rates truly has little effect on the crisis, which could be worsening because of more important but unaddressed factors (such as the shortage of treatment centers). These are hypotheses that require further validation, but if true, they would suggest to policy makers that a radically different approach is needed.

We were unable to evaluate many policies because most took effect outside our data timeframe. Several states have passed limits on first-time opioid prescriptions, but all were passed in 2016 and 2017. In addition, we would be interested in studying the effect of electronic Prescription Drug Monitoring Programs, but many states made this change long ago, preventing us from obtaining data both before and after digitization.

Areas for future work: In the future, we would extend the modeling effects to include different classes of models, such as tree-based methods, and ensembles of models. Additionally, we would explore autoregressive models to investigate the relationship among the features over time using a GARCH model. A model, that incorporates epidemiological methods like the spread of infectious disease, could be more accurate. Finally, more medically specific data (e.g. surgeon prescription practices by county, opioid types, addiction rates) would be more relevant and thus more predictive for our questions.

Data sources

1. County health rankings: http://www.countyhealthrankings.org/explore-health-rankings/rankings-data
2. Census data: https://www.census.gov/support/USACdataDownloads.html#VST
3. Opioid prescription rate, 2006-2016: https://www.cdc.gov/drugoverdose/maps/rxstate2016.html
4. https://www.cdc.gov/drugoverdose/maps/rxcounty2016.html
5. Drug overdose deaths, 2006-2016: https://wonder.cdc.gov/cmf-icd10.html