The Opioid and Heroin Epidemic: Predicting Fatal Opioid and Heroin Overdoses

Paper written for Class IST 718 – iSchool, Syracuse University

Authors: Farees Patel, Michael Marciano and Pooja Bathia

Abstract

Cities across the United States are targeting opioid drug manufacturers for deceptive advertising and marketing. The cities have begun actively suing these manufactures alleging that their marketing campaigns have significantly contributed to opioid related overdose deaths. In West Virginia communities, which boast the nation's highest incidence of opioid induced overdoses, legal firms have begun targeting drug manufacturers for their role in the opioid epidemic. The lawsuit alleges that the drug manufacturers lack substantive oversight and control over the surplus of opioid painkillers imported into the state, and that this is directly correlated to opioid overdoses. A further example was cited by Dr. Art Van Zee in 2009 wherein he highlighted the growth of OxyContin sales driven by divisive marketing campaigns. Sales of OxyContin increased 30-fold from 1996 to 2000. This growth, which also corresponded with increased availability, was positively correlated with the increased abuse of this drug. Purdue Pharma attempted to influence physicians' prescribing patterns through deceptive marketing, specifically by citing that OxyContin was less addictive than other opioids.

The blame does not stop with drug manufacturers, however. Physicians and pharmacies can be blamed for opioid related addiction and deaths. The West Virginia Supreme Court recently ruled that physicians can be held liable for the addiction of their patients, opening the door to patient lawsuits against their physyscians[3]. Recently, Dr. Hsiu-Ying Tseng was convicted of second-degree murder for overprescribing opioids that led to the deaths of three patients[4]. Coupon programs and the pressure of a large sales force are also responsible for physicians overprescribing opioids.[1] The ripple effects of the opioid epidemic have morphed into significant domestic health issues that correspond to potentially significant monetary losses for drug manufacturers, insurance companies, pharmacies and individual physicians.

Introduction

The opioid and heroin epidemic have been increasingly present in the last decade. The number of deaths due to opioid and heroin overdoses have increased tremendously. Establishing a link between opioid prescriptions and overdose deaths would allow insurance companies to establish risk factors for each specialty and adjust the premiums accordingly. Additionally, this would establish a type of causality for the opioid epidemic. Hence the objective of this project is to develop models to identify types of prescribers that are a high-risk for opioid related fatalities and predict most influential opioids leading to deaths due to overdose.

The increase in heroin use and overdose deaths has been theorized to be the result of the increase in opioid prescriptions. Therefore, identifying a predictive model would permit a type of root cause analysis and allow more effective monitoring systems.

Materials and Methods

A. *Data Set:*

Table 1: Overview of data files used to complete the assignment including column headers and descriptions

Data Set	Columns	Details
Prescriberinfo.csv	NPI	Prescriber identifier
	State	State abbreviation
	Credentials	Degrees or certifications
	Specialty	110 medical specialties
	Drug prescription counts	Per drug
	Opioid Prescriber	1-Yes 2-No
Overdoses.csv	State	
	State Population	
	Deaths	Prescribed opioids
Opoids.csv	Drug name	
drug_poisoning_deaths_by_state-_us_2013_2014-v7.csv	State	
	2014 Number	Deaths - heroin overdose

The data was obtained from kaggle (www.kaggle.com/apryor6/us-opiate- prescriptions and the cdc (www.cdc.gov/drugoverdose/data/statedeaths.html.)

Data Munging:

a. Presciberinfo.csv
   1. Removed all prescribers who have not prescribed opiates (0) using the last column where value 1 indicates Yes while value 2 indicates No
   2. Removed all columns (parsing column 6 through 256) that were not opioids. Used opioids.csv to identify and remove all non-opioids columns
   3. Standardized naming conventions of credentials and specialty columns. The original dataset contained plenty of anomalies in enlisting the specialty of the prescribers
   4. Created column indicating the total opioids prescribed per row for each prescriber

b. Overdoses.csv
   1. Joined overdose.csv columns to Prescriberinfo.csv. using state abbreviation column
   2. Joined the drug_poisoning_deaths_by_state-_us_2013_2014-v7.csv with prescriberinfo.csv using state

We started with four csv files with unclean data which we transformed into 25000 rows of clean and transformed data which captures the number of opioid deaths grouped by states. Thus, our cleaned data set contained metrics of the population, total prescribers, total number of deaths opioid prescription, specialties of the prescribers by state and total number of deaths due to heroin by state in United States of America.

B. Modeling:

Linear regression forms the crux of our analysis as we try to weigh different features that influence the deaths due to opioid overdose. Regression analysis is a statistical technique for investigating and modeling the relationship between variables[5]. We used linear regression to develop predictive models using pyspark – Spark via Jupyter ipython notebook (which was provided as a separate document). Using combination of simple linear regression model and multiple linear regression models we identified the key features which are more influential in identifying the high-risk prescribers and opioid drugs leading to death due to overdose. Eight models were evaluated, four related to opioid deaths (Table 2) and four to heroin deaths (Table 3).

Table 2: Features used in the linear regression models developed to predict opioid overdose deaths.

Model	Features		
1-O	total opioid prescribed		
2-O	total opioid prescribed	specialties	
3-O	total opioid prescribed	number prescriptions per opioid	
4-O	total opioid prescribed	specialties	number prescriptions per opioid

Table 3 : Features used in the linear regression models developed to predict heroin overdose deaths.

Model	Features			
1-H	total opioid prescribed	opioid deaths		
2-H	total opioid prescribed	opioid deaths	specialties	
3-H	total opioid prescribed	opioid deaths	number prescriptions per opioid	
4-H	total opioid prescribed	opioid deaths	specialties	number prescriptions per opioid

The opioid feature and the heroin features were normalized to the state population and are in units per 1 million.

C. Evaluation:

The data set was split into training, validation and testing in the ratio of 7:2:1. We developed the models using the training data set and then calculated the root mean square error (rmse) for each of the model listed above using validation data set. Models were compared using the rmse of the validation performance. The model with the lowest rmse will be chosen as the highest performing model. The generalizability of the model was evaluated using the rmse of the model when used to predict based on the data in the testing set.

Results

A. Opioid Overdose prediction

Model 1-O outperformed all other models, exhibiting a rmse of 45.06 (intercept = 136.05;coefficient = 0.0033) (Figure 1). This model has a generalizability (rmse of testing set) of 64.77. This model was the simplest model evaluated in this project, with one feature, the total opioids prescribed. Model 4-O performed similarly, with an rmse of 52.35 obtained from the validation set evaluation. While model 4-O (validation rmse = 52.36) did not perform as well as 1-O, it was comparable. Model 4-O has three features, and does not include specialty.

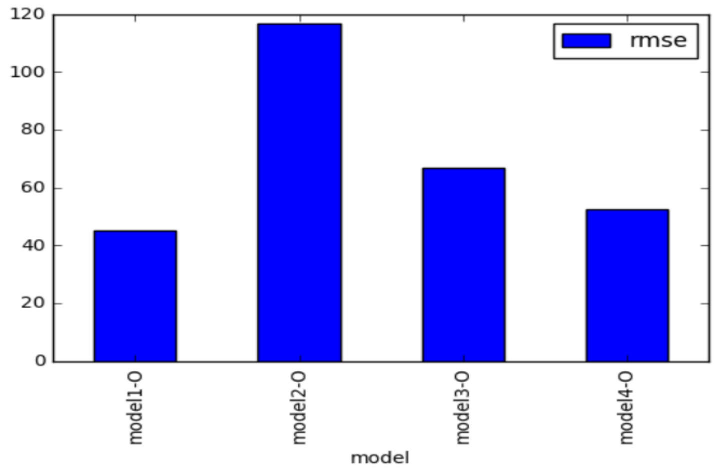

Figure 1:RMSE: opioid models(validation data)

A goal of this study was to identify medical specialties that have a high risk for patients experiencing fatal overdoses from opioids. Although Model 2-O outperformed by other models in predicting opioid deaths, it does identify trends in the medical profession (Figure 2). The higher the weights are indicative of the specialties that are more likely to prescribe opioids and lead to deaths due to an overdose on opioids. Thus, we can see that doctors or prescribers in the field of psychiatry, internal medicine, cardiac electrophysiology are more likely to prescribe opioids. While Neuropsychiatry, Nuclear_Medicine, Hospice_and_Palliative_Care, Plastic_Surgery and Cardiac_Surgery had weights of 0 which means that they are least likely to influence deaths due to opioid overdose.

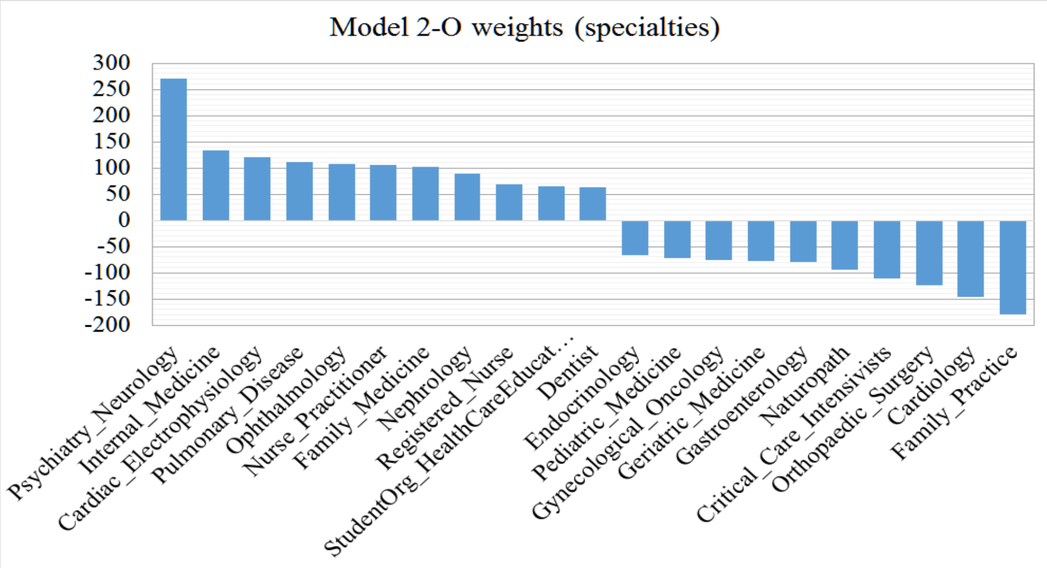

Figure 2: The weights of the medical specialties relating to the linear regression model 2-O.

B. Heroin Overdose prediction:

Model 4-H (rmse = 20.81) predicts heroin related overdose deaths with the lowest error among the models evaluated (using the validation data set) (figure 3). The generalizability of this model assessed through rmse is 23.68. This model has four classes of features and we believe that the number prescriptions per opioid is the feature that is likely due to the improved responses when compared to the other models. Using weights (coefficients), we are able to identify the individual opioid drugs that influence fatalities due to heroin overdoses (Figure 4).

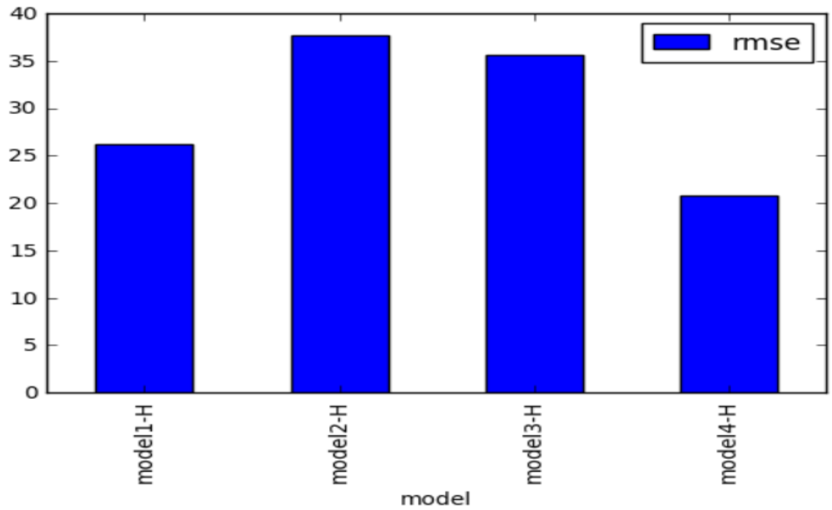

Figure 3: RMSE: heroin models (validation data)

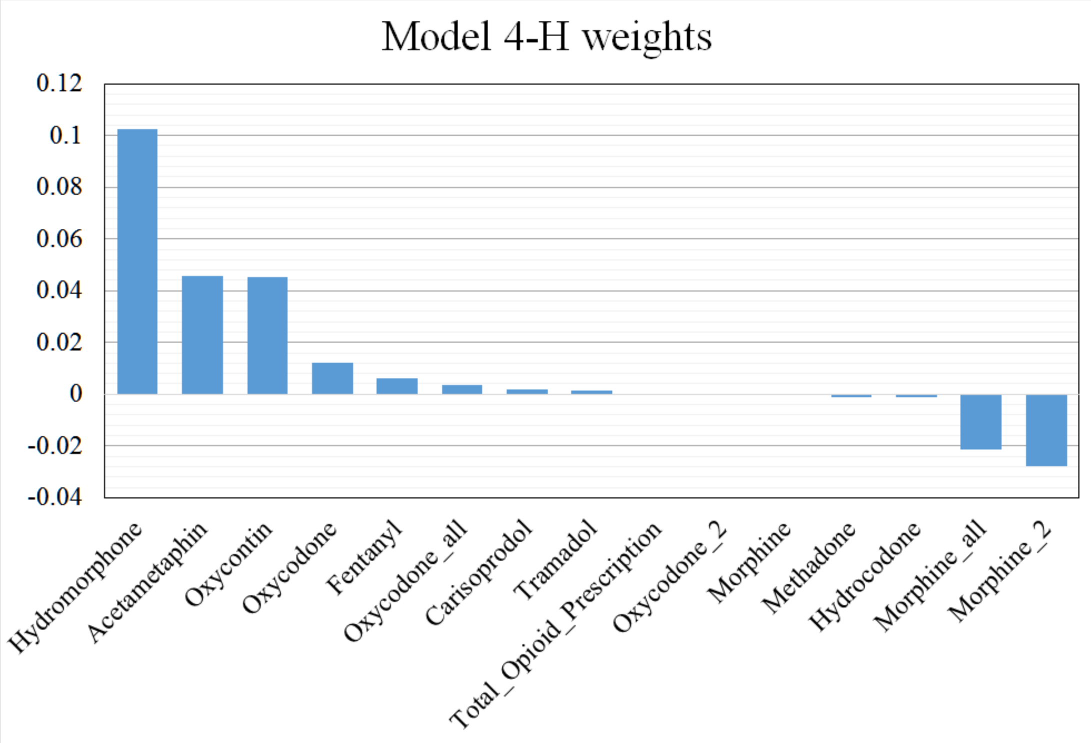

Figure 4: Results of linear model 4 for Heroin

Discussion and Conclusion

While using modeling for analysis it is important to understand that there is not a single correct method to arrive at a conclusion. Linear regression is likely not the optimal method model this problem set however it was appropriate based on the level of knowledge the group had at the time.. A random forest for example may have better predictive power than the linear regression model. Despite this the results indicate that there is a link between opioid prescriptions and the amount prescribed in predicting opioid overdose deaths and a link between the total opioids prescribed, opioid overdoses deaths, the type of opioid and specialty. This link provides enough evidence to warrant further analyses. Hence, even though linear regression analysis gives us the list of features that might be most influential in predicting the high-risk prescribers it is possible to get deeper insights by using mode modelling techniques and methodologies.

References:

[1] \https://baronandbudd.com/mydrugjustice/opioid-lawsuit/

[2] Van Zee A. The Promotion and Marketing of OxyContin: Commercial Triumph, Public Health Tragedy. American Journal of Public Health. 2009;99(2):221-227. doi:10.2105/AJPH.2007.131714.

[3] http://epmonthly.com/article/you-re-suing-me-for-what/

[4] http://www.latimes.com/local/lanow/la-me-ln-doctor-prescription-drugs-murder-overdose-verdict-20151030-story.html

[5] Montgomery Douglas, Peck Elizabeth and Vining Geofrrey. Introduction to Linear Regression Analysis Fifth Edition. April 2012. ISBN: 978-0-470-54281-1