



The Opioid and Heroin Epidemic : Predicting Overdose Deaths

Pooja Bathia, Farees Md, Michael Marciano
iSchool, Syracuse University, New York, US

Purpose

The purpose of this project is to develop models to:

- (1) Identify types of prescribers that are a high-risk for opioid related fatalities
- (2) Predict opioid related deaths
- (3) Predict heroin related deaths

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. Cities across the United States are targeting opioid drug manufacturers for deceptive advertising and marketing, alleging the a significant contribution to opioid related overdose deaths. This has been extended to physicians and pharmacies.

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.

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

Data Overview

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	
Heroin_deaths_2014.csv	State	
	2014 Number	Deaths - heroin overdose

Materials and Methods cont.

Data Munging using Jupyter Notebook/python (pandas) to generate cleaned data frame (25000 rows) grouped by state (50 rows) containing:

- Population
- Total Prescribers
- Total number of deaths opioid - prescription
- Specialties of the prescribers
- Total number of deaths - heroin

Linear regression was used to develop predictive models in Spark (pyspark).

Opioid Models (4) and features:

- Model 1: total opioid prescribed
- Model 2: Model 1 + specialties
- Model 3: Model 1 + number prescriptions per opioid
- Model 4: Model 2 + number prescriptions per opioid

Heroin Models (4) and features:

- Model 1: total opioid prescribed + opioid deaths
- Model 2: Model 1 + specialties
- Model 3: Model 1, + number prescriptions per opioid
- Model 4: Model 2 + number prescriptions per opioid

All features were normalized to the state population and are in units per 1 million.

Evaluation:

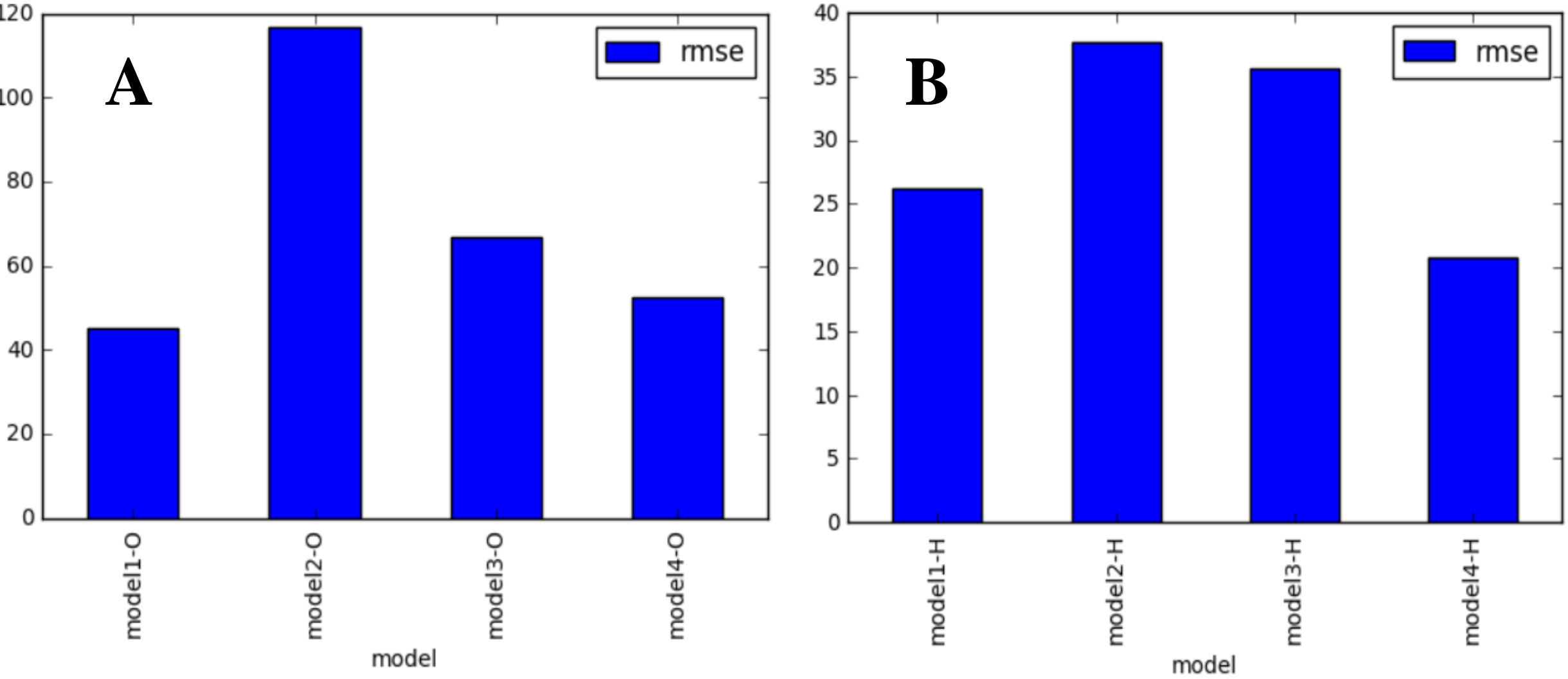
- Training, validation, testing split: 0.7,0.2,0.1
- Models were compared using root mean square error (rmse) of the validation performance
- Generalizability is evaluated using rmse of the model when used to predict based on data in the testing set.

Data Source:

- www.kaggle.com/apryor6/us-opiate-prescriptions
- www.cdc.gov/drugoverdose/data/statedeaths.html

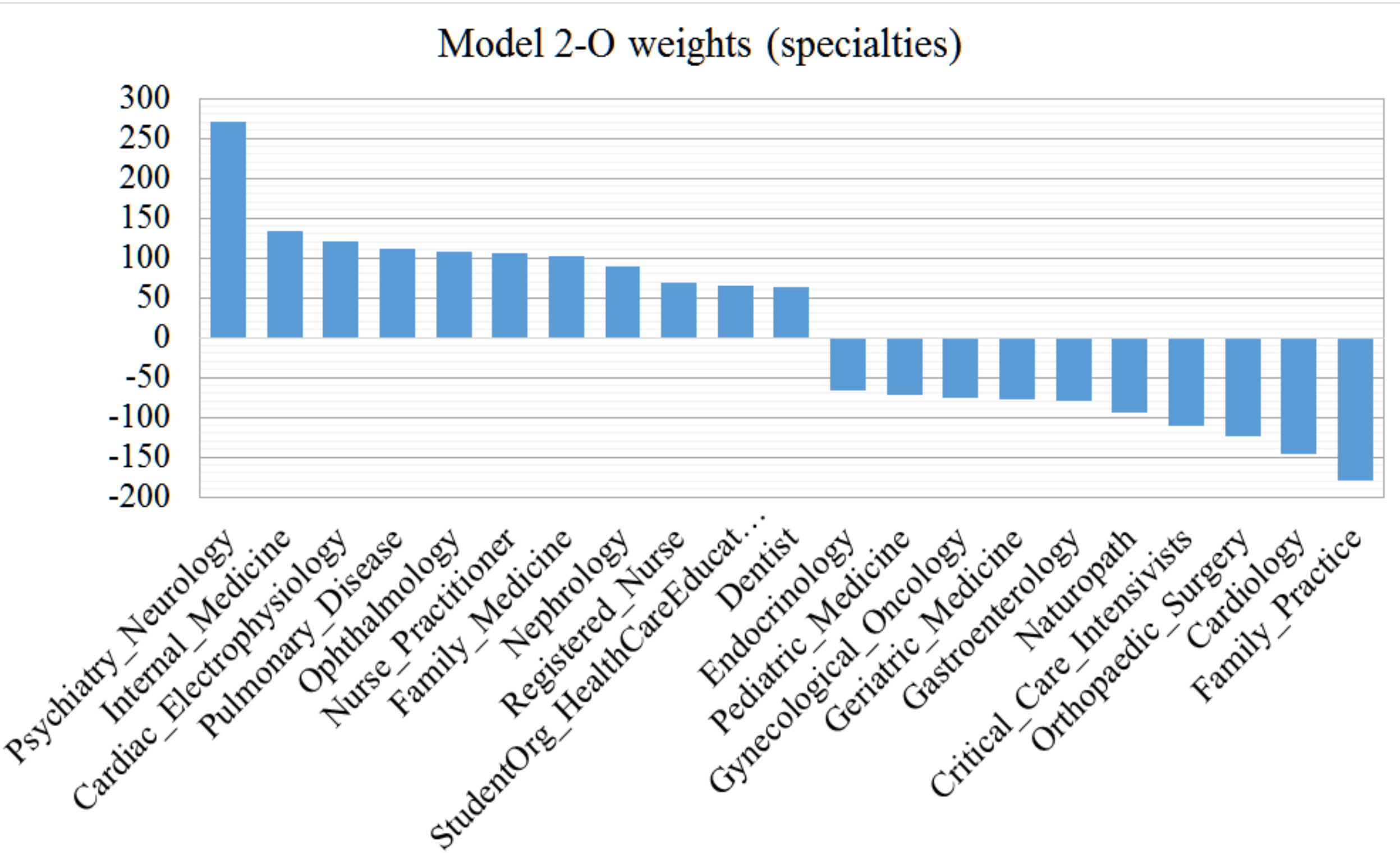
Results

RMSE :opioid (A) and heroin (B) models (validation data)



Opioid overdose prediction

- Model 1-O
 - Lowest RMSE = 45.06.
 - Generalizability (testing): RMSE= 64.77
 - Intercept: 136.05; coefficient: 0.0033
- Model 2-O
 - Included prescriptions/specialty, although not optimal, the higher risk specialties may be tentatively identified.



- ✓ Note: Neuropsychiatry, Nuclear_Medicine, Hospice_and_Palliative_Care, Plastic_Surgery and Cardiac_Surgery had weights of 0.

Results cont.

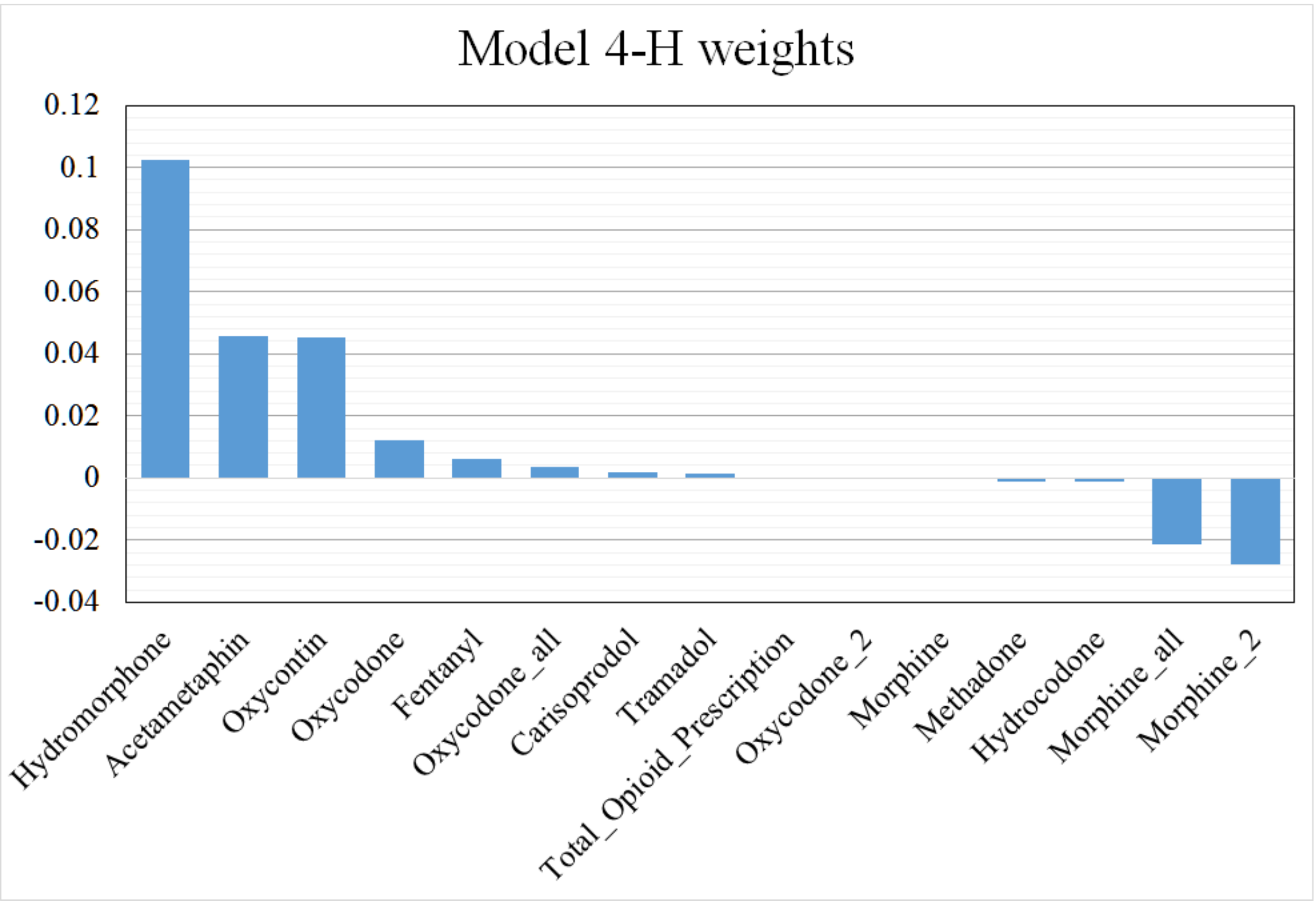
- Model 4-O
 - Performance similar to model 1.
 - This model can be used to summarize the effects of individual drugs on the opioid overdose deaths

Conclusion: Opioid Models

Generally, the performance of all opioid death models were suboptimal. Next steps would be to use random forests and cross validation to combat the small training set.

Heroin Overdose Prediction

- Model 4H:
 - Lowest validation set RMSE, 20.81
 - Generalizability (testing) : RMSE = 23.68
 - Number of opioid deaths/million most influential feature (weight 1.1188)
 - Intercept: -13.75
 - Additional feature weights included below.



Conclusion: Heroin Models

Model 4H appears to be a satisfactory predictor of heroin overdose deaths. However, increasing the size of the training, validation and testing set or using cross validation would provide significant benefits. Additional, non-linear models should also be examined.