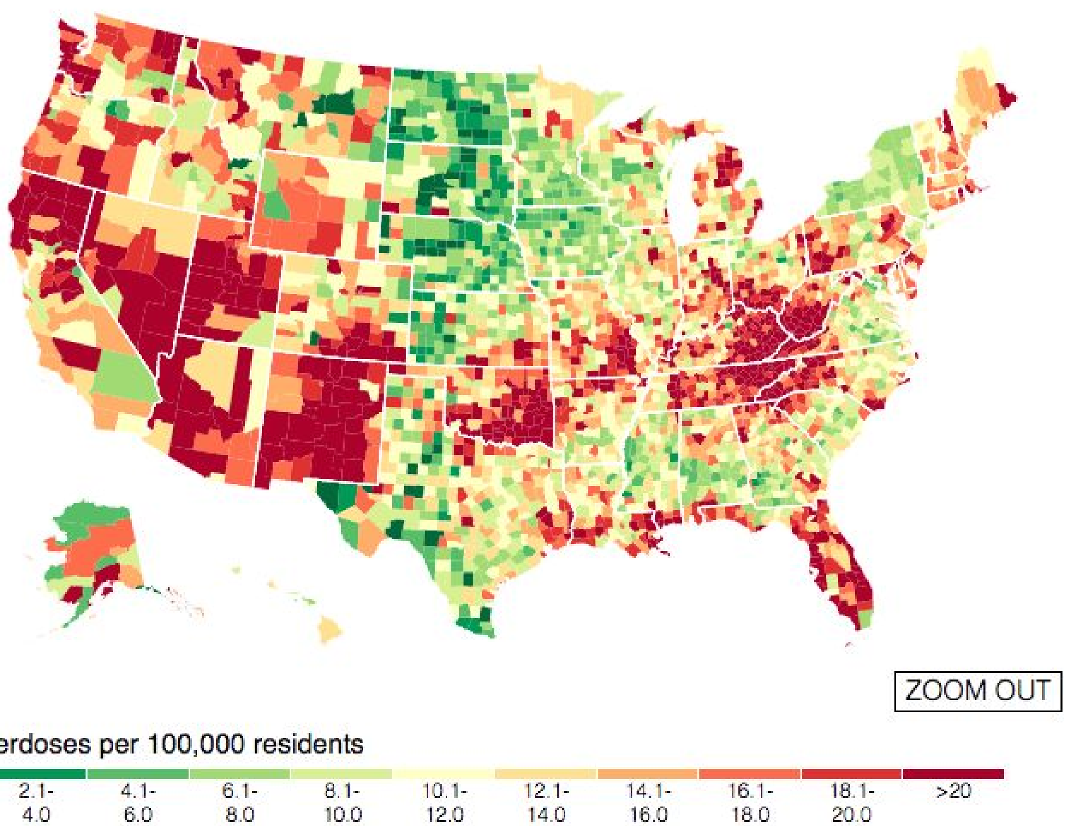
Data & Visual Analytics Project Proposal

Nopioid: Targeting High-Rate Opioid Prescription Medical Providers

Matt Liedtke, Prasanta Lenka, Supriya, Pallavi Joshi, Zach Olivier



[0]

Objective

Problem Statement: Determine which US medical providers are writing above expected levels of opioid prescriptions. Provide insights into which factors are driving provider prescription rates.

The National Institute of Drug Abuse estimates roughly 21%-29% of patients prescribed opioids for chronic pain misuse the opioids, and around 80% of people who use heroin first misused prescription opioids. Identifying practices that write higher than normal levels of opioid prescriptions relative to their attributes will help efforts in education, awareness, and audits to help solve the opioid crisis at its core. [1]

Current State

Authorities need to identify where resources should be targeted to reduce opioid prescribing. The Department of Health and Human Services conducts audits on medical practices to analyze opioid prescribing and dispensing practices. Audits are typically very costly and inefficient. Existing visualizations are sufficient for viewing trends but fail to identify over prescribing providers. Currently, most analytics and modeling efforts have focused on predicting patient level risk to opioid addiction [2], which ignores the role the provider takes in the crisis.

Innovation - How are we different?

Our solutions will target providers whose rate of opioid prescription is above an expected rate based on provider attributes. Provider targeting allows for interventions across all practice types.

Attributes include: location, medical school rank, distance to peers, provider type and more. [3] A study focused on opioid prescribing for pediatric dentistry in Nova Scotia included provider graduation year as an attribute. [4] While statistically significant relationship was not found, this study population was limited compared to our broad national, multi-specialty data, and there is interest if significance changes.

Project Plan:

- Develop a standardized training dataset connecting together all medical practice attributes based on provider data from the CMS API. [5]

- Encode categorical data in a format that allows for predictive modeling by leveraging feature extraction methods like vTreat [6]. vTreat encodes high cardinal features in a statistically sound manner. These encodings are not substitutes for domain knowledge.

Additional variables like smoking rate, age, gender,and prescription history are additional domain-focused variables to consider including in our analysis [7]

- Exploratory analysis to understand relationships between provider attributes and prescription rate. Includes expanding on recent analysis performed on only the top 10% of opioid prescription writers [8] with a more detailed Hierarchical Latent Class Clustering Analysis [9]. HLC provides meta-cluster groups based on categorical variables. Many options are available to solve categorical variable clustering including a variants of K-Means, K-Modes [10] and K-Prototypes [11]. Further, if our dataset includes a mix of categorical and numerical variables, clustering becomes more difficult to extract insights. Probabilistic based clustering algorithms like t-SNE [12] and DBSCAN [13] are options to counteract traditional clustering issues with our dataset.

- Model to estimate expected prescription rate based on provider attributes. Comparing the expected to the actual rate allows us to determine which providers are over-scripting. High cardinal categorical features makes modeling difficult. We will employ feature selection methods that eliminate redundant levels through *artifact contrast variables*, which utilize traditional variable importance output to eliminate features from our model. [14]

- Visualize outputs of the predictive model in order to determine the “where” and “why” of our results. Employing model interpretation methods like SHAP [15] and LIME [16] will allow us to pin-point which attributes are contributing to our model’s prediction. SHAP and LIME values both aim to quantify the non-linear relationship between variables in a fitted machine learning model on the *observation level*. Allowing us to show *by provider* which features influenced our model’s prediction.

- Developing a simple interface showing residual values of our model will help us identify over-scripting geographically. Insights into the “why” and “where” of prescription rates together in an engaging interface will help users target over-scripting providers and trust our recommendations. Interactivity with our insights is extremely important. Our team plans to take advantage of visualization toolkits such as D3.js, Vega and ECharts. ECharts framework allows for multi-thread rendering of our results which provides efficient visualization of complex visuals sourced from large datasets. Further, the up front coding of these visualizations is handled with only a few lines of code, a huge advantage for D3.js and other web-based visualization frameworks. [17]

Figure: Classification example of SHAP machine learning interpretation. Clear explanation of input variables to prediction drives understanding and trust. [18]

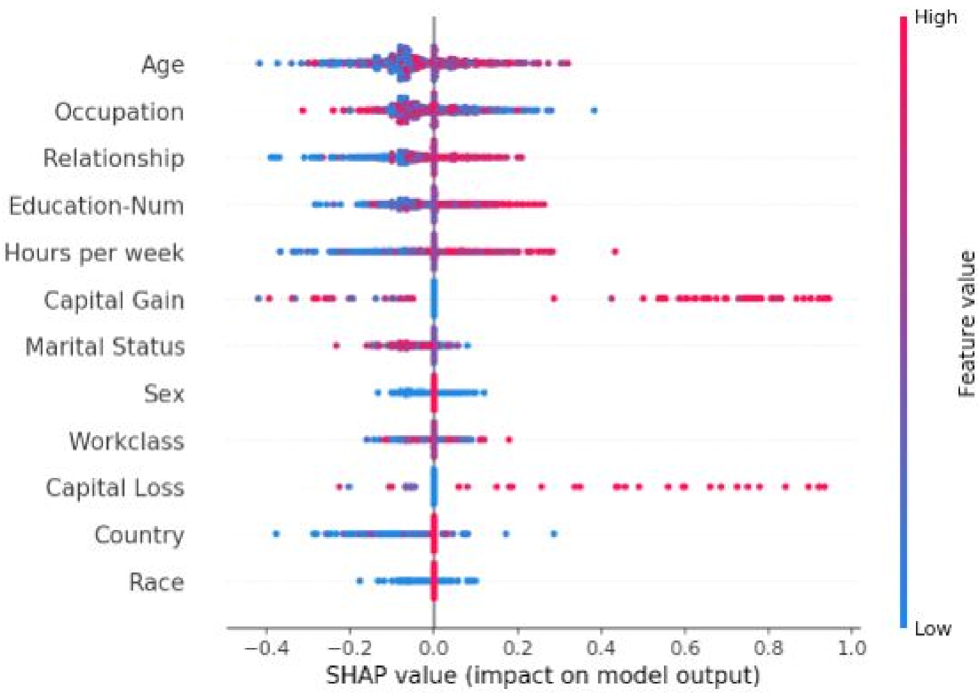
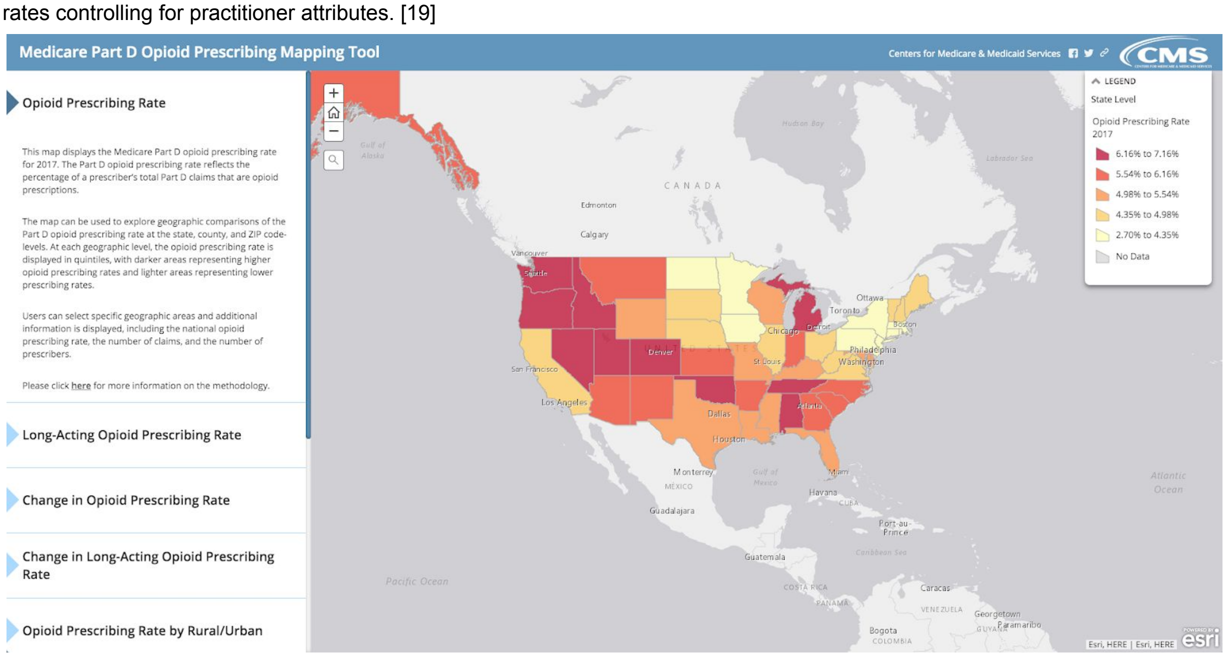


Figure: Current state analysis is done at patient level, not at the provider level. Our analysis will aim to provide opioid prescription rates controlling for practitioner attributes. [19]



Why does this matter?

The Center for Disease Control and Prevention (CDC) estimates that around 68% of the more than 70,200 drug overdose deaths in 2017 involved an opioid. According to their research, the number of overdose deaths involving opioids was 6 times higher in 2017 than 1999. [20]

Directly targeting over-scripting practices will positively affect one of the largest factors in the ignition of the opioid crisis – opioid prescriptions.

Measuring Success

We will be successful if we enable a public health official or concerned community member to identify overprescribing providers in their community. Second, the user will be able to tell whether the provider is an outlier, or part of a larger cluster. We enable the user to take appropriate action, such as specific provider consultation, or broader policy.

Risks, Payoffs, Cost Implications

The data we will be using is limited to the US population aged 65 and older. We may fail to capture a provider who sees few or no patients in this population. Additionally, the majority of deaths related to opioids are people under the age of 50. Still, the 65 and older demographic is an important demographic to focus on, because the literature suggests that the elderly are susceptible to more chronic conditions. If a patient is taking different drugs for multiple conditions, they are at higher risk for drug interactions. [21]

The second risk is evidence that shows providers over-scripting opioids is becoming a slightly less significant part of solving drug overdose deaths. At the state level opioid prescribing rates are decreasing. Opioid prescription overdose deaths, while at a 15 year high, have dropped. This is positive news, and our analysis can help ensure this trend continues. [22]

Developing our model and analysis is not costly - there is plenty of data available to determine which medical providers to target.

Milestones

- Develop a relevant and efficiently stored dataset that allows us to perform our analysis.
- Perform data exploration, identify the features to predict, draft models
- Develop a tested model that generalizes well, with relevant supporting analysis.
- Model explanation and targeting landscape in a easy-to-use interactive interface.
    - Estimated Time: 9 Weeks Total

Team Roles

- Data Team: Responsible for data storage, cleaning, and feature development.
  - Matt Liedtke, Zach Olivier, Prasanta Lenka
- Analytics Team: Responsible for developing clustering analysis and prediction model.
  - Prasanta Lenka, Supria, Zach Olivier, Pallavik
- Visualization Team: Responsible for the interactive output of the model and analysis.
  - Pallavik, Prasanta Lenka, Matt Liedtke

Reference List

[0] Wilson, C. (n.d.). Drugs in America: Watch the Epidemic Spread. Retrieved from https://time.com/4260798/drug-epidemic-america/

[1] National Institute on Drug Abuse. (2019, January 22). Opioid Overdose Crisis. Retrieved from https://www.drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis

[2] Ellis1, R. J., Genes2, N., & Ma'ayan, A. (2019, January 29). Predicting opioid dependence from electronic health records with machine learning. Retrieved from https://biodatamining.biomedcentral.com/articles/10.1186/s13040-019-0193-0

[3] Schnell, M., & Currie, J. (2018). Addressing the Opioid Epidemic: Is There a Role for Physician Education? *American Journal of Health Economics*, *4*(3), 383–410. doi: 10.1162/ajhe_a_00113

[4] Matthews, D. C., Brillant, M., Jimoh, K. O., Singleton, W., McLean-Veysey, P., & Sketris, I. (2019). Patterns of opioid prescribing by dentists in a pediatric population: a retrospective observational study. CMAJ open, 7(3), E497–E503. doi:10.9778/cmajo.20190021

[5] Socrata Developer Portal. (n.d.). Retrieved from https://dev.socrata.com/foundry/data.cms.gov/6wg9-kwip

[6] Zumel, Nina, & John. (2019, September 22). vtreat: a data.frame Processor for Predictive Modeling. Retrieved from https://arxiv.org/abs/1611.09477

[7] Skala, K., Reichl, L., Ilias, W., Likar, R., Grogl-Aringer, G., Wallner, C., … Walter, H. (2013). Can we predict addiction to opioid analgesics? A possible tool to estimate the risk of opioid addiction in patients with pain. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/24284844

[8] P.GuyJr., G. (2019, February 16). Identifying opioid prescribing patterns for high-volume prescribers via cluster analysis. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0376871618306926

[9] (PDF) Hierarchical Latent Class Models for Cluster Analysis. (n.d.). Retrieved from https://www.researchgate.net/publication/228679557_Hierarchical_Latent_Class_Models_for_Cluster_Analysis

[10] A Fast Clustering Algorithm to Cluster Very Large ... (n.d.). Retrieved from https://grid.cs.gsu.edu/~wkim/index_files/papers/fastclusteringHuang.pdf

[11] Huang, Z. (1997, January 1). [PDF] Clustering Large Data Sets with Mixed Numeric and Categorical Values - Semantic Scholar. Retrieved from https://www.semanticscholar.org/paper/Clustering-Large-Data-Sets-with-Mixed-Numeric-and-Huang/d42bb5ad2d03be6d8fefa63d25d02c0711d19728

[12] Visualizing Data using t-SNE - cs.toronto.edu. (n.d.). Retrieved from http://www.cs.toronto.edu/~hinton/absps/tsne.pdf

[13] A Density-Based Algorithm for Discovering Clusters in ... (n.d.). Retrieved from https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf

[14] Feature Selection with Ensembles, Artificial Variables ... (n.d.). Retrieved from https://www.researchgate.net/publication/220320233_Feature_Selection_with_Ensembles_Artificial_Variables_and_Redundancy_Elimination

[15] Lundberg, Scott, & Lee. (2017, November 25). A Unified Approach to Interpreting Model Predictions. Retrieved from https://arxiv.org/abs/1705.07874

[16] Ribeiro, Tulio, M., Singh, Sameer, & Carlos. (2016, August 9). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Retrieved from https://arxiv.org/abs/1602.04938

[17] Li, D., Mei, H., Shen, Y., Su, S., Zhang, W., Wang, J., … Chen, W. (2018). ECharts: A declarative framework for rapid construction of web-based visualization. Visual Informatics, 2(2), 136–146. doi: 10.1016/j.visinf.2018.04.011

[18] Slundberg. (n.d.). slundberg/shap. Retrieved from https://github.com/slundberg/shap/blob/master/README.md

[19] Story Map Series. (n.d.). Retrieved from https://cms-oeda.maps.arcgis.com/apps/MapSeries/index.html?appid=735f83ac6e984d6fade11b241d295585

[20] Understanding the Epidemic | Drug Overdose | CDC Injury Center. (n.d.). Retrieved from https://www.cdc.gov/drugoverdose/epidemic/index.html

[21] Bonnie, R. J., Ford, M. A., & Phillips, J. K. (2017). *Pain management and the opioid epidemic: balancing societal and individual benefits and risks of prescription opioid use: Committee on Pain Management and Regulatory Strategies to Address Prescription Opioid Abuse*. Washington, D.C.: The National Academies Press.

[22] Goodnough, A., Katz, J., & Sanger-katz, M. (2019, July 17). Drug Overdose Deaths Drop in U.S. for First Time Since 1990. Retrieved October 8, 2019, from https://www.nytimes.com/interactive/2019/07/17/upshot/drug-overdose-deaths-fall.html.