Spatial and Temporal Analysis of Trends in Opioid Abuse within the US

Mario Iurlaro, Christopher Nelson, Thomas Swarbrick, Daniel Vela, Patrick A. Wall

1. INTRODUCTION

Prescription opioid abuse in the United States is an epidemic that has affected millions, with an estimation of drug abusers in recent years exceeding 15 million [Basak et al, 2019]. Determining the factors that create a greater risk of opioid overdose hospitalisation or death (overdose harm) is key to creating effective public health interventions to reduce the effects of the epidemic. Our objective for this project is to build an interactive visual tool to identify emerging trends in the opioid epidemic, utilising available open-source data to analyse opioid abuse trends regionally, demographically, and over time. If successful, this could provide a new way for local public health departments and organisations to identify the unique circumstances that contribute to opioid abuse in their region, or determine which demographics in their jurisdictions are the most at-risk for prescription opioid abuse, to focus their interventions where they are most needed.

2. LITERATURE SURVEY

In recent years, an increasing amount of studies have used analytical approaches to potentially identify contributing factors, causal relationships and geographical variations of the opioid epidemic. However, the current practice is often restricted to small regional approaches or limited in the scope of potential factors. Recent studies have used machine learning approaches to predict the risk of opioid use disorder on a population level [Hasan et al. 2021] or applied probabilistic models to find variation in prescription trends [Hu et al., 2015], but were restricted to Massachusetts or areas in Queensland (Australia), respectively. Areas of higher prescription rates have also been a central focus of research, in an attempt to identify areas of urgent action and causality factors. One approach used spatial cluster detection to monitor local patterns of prescription opioid abuse in New Mexico [Brownstein et al.]. A similar approach identified smaller hot-spot areas where the opioid prescription rate was significantly higher than the baseline measured as the rate of the whole state [Basak et al. 2019]. All these studies highlight the potential of using quantitative approaches to identify areas and predictive factors associated with the opioid abuse emergency, but also reveal limitations in the data.

Understanding the spatial-temporal effect of opioid abuse in the United States is a crucial aspect of this effort, as it might help highlight systemic differences that are at the basis of the increase in opioid usage. An argument is to be made that the misuse of opioids reflects the most substantial public health crisis that the US has faced [Maclean et al., 2020]. A multi-level analysis has shown as much as 10% of the variation in opioid intake can be explained by geographical data itself [Webster et al. 2009], even if using state-level data only. More recent studies have shown how there is a large geographical variation in opioid prescription rate and a general lack of consensus regarding opioid use to treat pain [McDonald et al. 2012; McDonald and Carlson, 2014].

Opioid abuse data can be hard to track as many health problems, including causes of death, may not seem opioid-related. Polysubstance abuse often occurs in those who are addicted to opioids. Studies have shown that the less widely used, the more likely a drug is to be used with other drugs [Compton et al., 2021]. Because of this, we could potentially use other types of drug overdoses as potential predictors in our analysis [Unick & Ciccarone, 2017; Unick et al., 2013].

There seem to be clear differences in opioid intake between urban and rural areas. An analysis of prescription opioid poisoning across counties in California indicated factors, such as higher pharmacy density in urban areas and low income with more manual labour industries in rural areas as potential causes for this variation [Cerda et al., 2017]. In the case of Purdue Pharma, they would target low-income miners in rural communities to target Oxycontin which led to a major crisis [Whelan & Asbridge, 2013]. Availability of healthcare services, differences in demographics and unemployment rates have also been indicated as potential causes of this difference [Sun et al. 2022; Duan & Hand, 2021]. All these indicate the importance of including the rural/urban aspect in our analysis, while also clearly highlighting the limitation associated with single-state data or lack of accountability of confounding variables.

Public health departments and governments need to know where to best allocate resources to combat the ongoing prescription opioid epidemic in the US. Current research focuses on a single specific state or on a larger country scale. Our analysis will investigate potential factors that contribute to the opioid epidemic at a state level using county-level data and build a visually interactive tool to visualise trends annually.

3. PROPOSED METHOD

A county-level dataset [Griffith et al. 2021] with U.S. opioid prescription pill distributions, demographics and several other county-level variables from 2006-2013 will be used in the analysis to determine the factors that contribute to specific geographic hot spots in the opioid epidemic. The dataset leverages several U.S. federal data resources such as the Automation of Reports and Consolidated Orders System (ARCOS), the Center for Disease Control (CDC) Wide-ranging Online Data for Epidemiologic Research (WONDER) database and the Health Resources & Services Administration's (HRSA) Area Health Resource File (AHRF). It combines extracted pill shipments for oxycodone and hydrocodone to retail pharmacies in the U.S. from the database ARCOS, data on opioid-related deaths and cancer deaths from WONDER, and county-level characteristics such as annual data on demographics, healthcare workforce, rurality, unemployment rate, etc, from AHRF. The ARCOS data was aggregated to calculate county-level annual per capita pill volume (PCPV). Since WONDER suppresses data for counties with less than 10 deaths, the authors used imputation methods to estimate Opioid-Related Deaths (ORD).

Based on our literature review, analysis to date has been completed on either individual states such as New Mexico, West Virginia or North Carolina [Brownstein et al. 2010, Basak et al. 2017] or on the entire country [Griffith et al. 2021]. To determine the features that create a greater risk of opioid abuse, both geographically and through time, we will be performing multiple linear regressions on county-level demographic data for each state individually, and over time. The regression analysis will also include regularization techniques to determine feature relevance and overcome issues such as multicollinearity. PCPV and ORD have been selected as features to predict as we believe they will provide the most insights into opioid abuse.

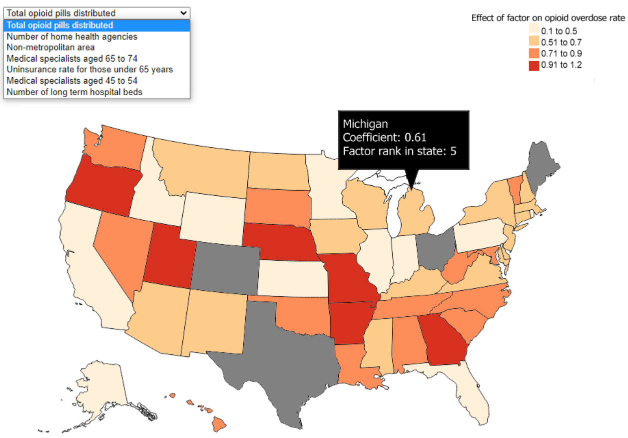

Figure 1. Proof of concept demonstration of Interactive tool. Values do not represent any actual data.

An interactive visual tool identifying opioid overdose trends over time will be developed with use of the results from our regression analysis. This tool will allow public health departments to design intervention techniques and policies focused on specific geographic areas and demographics. A proof of concept interactive tool is shown in Figure 1. The interactive visual tool will be a map of the U.S. that can demonstrate the impact on opioid harm of any given factor from a dropdown list of factors. Once a factor is selected, the legend and colour-coding of each state will change to demonstrate which states have an opioid harm rate that is more impacted by that factor, based on the coefficient of the corresponding factor produced by the analysis. States with a greater coefficient will be colored with a more saturated colour, and states which exclude the factor as part of their feature selection process will be coloured grey. Hovering mouse cursor over any given state will produce a tooltip that will name the state, specify the coefficient corresponding to the selected factor, and specify the rank of the factor in the equation to estimate the opioid harm rate, meaning the placement of that the factor's coefficient in an ordered descending list of every coefficient in the equation. Time permitting, we may include an option in a separate dropdown menu to colour-code the states according to the rank of the factor instead of the value of its corresponding coefficient, to visually identify groups of states that all hold a given factor in the same place of importance in determining the opioid overdose rate; note that this feature is not shown in the proof of concept illustration of the interactive visual tool. The use of county-level data will also allow us to look at counties that show higher risk compared to the average of the state they are in.

If successful, our study will help to identify factors that contribute to the increasing rate of opioid overdose hospitalisations and deaths and to indicate specific areas where these issues are increasing at a more urgent rate. Our goal is to identify the potential factors contributing to these geographic/demographic hot spots to allow public health to implement strategies efficiently.

A risk of the study is to incorrectly identify a factor contributing to an area's increased opioid use. If resources are allocated to this factor, it would be a waste of time and money. The opioid epidemic has been greatly studied with lots of research identifying common factors. Therefore, the factors we identify for an area are most likely to correspond to factors identified in other studies. The potential payout of contributing to the decline of opioid use and overdose would be far more rewarding than the drawbacks.

An additional concern is that given the large number of demographic features provided in the county-level dataset (156), it will be difficult to determine feature relevance due to factors such as multicollinearity. To address this problem, we will be using regularisation techniques, such as LASSO regression, to determine which features provide the most predictive power and explain the geographic and temporal trends observed. Based on our literature review, this type of feature selection has not been done before on opioid prescription and mortality data at this scale.

4. PLAN OF ACTIVITIES

The data used in this project will be open-source that will not present any financial cost. The software used, such as Python, HTML, CSS, and D3.js, will also be free of charge.

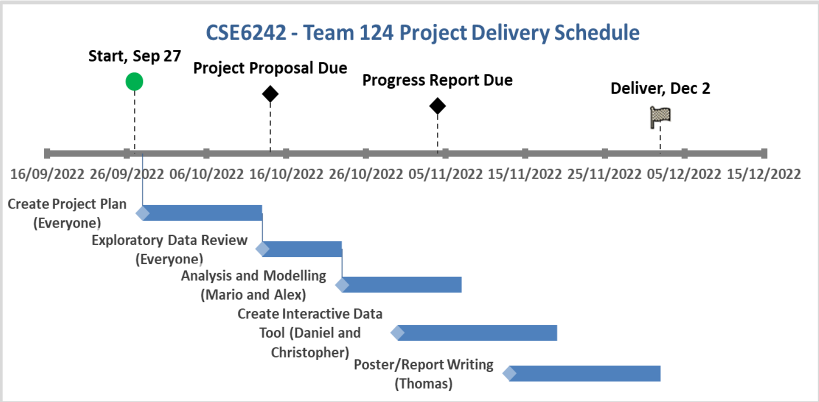

Figure 2. Project plan and timeline as a Gantt chart

Each team member has performed a literature review for 3 papers, proposed at least one open-source dataset for the project and completed written components for the documents. The dataset has been selected and proof of concept regression analysis has been implemented. The team will now be split into groups responsible to finish statistical analysis, creating the visual data tool and documentation. This leaves us with a month to finalise the visual components and complete the documentation as our "final". All team members have contributed a similar amount of effort thus far. The project plan outlining our previous milestones and future activities is shown in Figure 2.

5. EXPERIMENTS/ EVALUATION

A subset of the dataset from Griffith et al. was selected to validate the workflow and determine the demographic features that are most impactful for the counties within a state. To achieve this, an initial evaluation was performed only on counties in West Virginia and during the year 2013. West Virginia was selected due to its disproportionately high rates of opioid abuse compared to the rest of the country [Brownstein et al. 2010].

The dataset has 156 features recorded for each of the 55 counties in West Virginia. The dataset contained 20 counties in 2013 with suppressed values for Cancer Deaths, Opioid Deaths and Crude Opioid Death Rate, which occurs when there are less than 10. The regression analysis so far was done using values from imputation methods for these counties. Given the significant difference in range between the features, standardisation was performed such that the transformed feature had a mean of 0 and a standard deviation of 1.

LASSO regression was used initially as a regularisation technique given the number of features far exceeded the number of observations. LASSO regularisation is a technique used for variable selection and regularisation to enhance the predictive ability of a model. The analysis was performed to predict PCPV and ORD within the state of West Virginia in 2013. The LASSO regression included 5-fold cross-validation in order to determine the optimum value of regularisation which minimises squared error. The LASSO regression analysis determined 16 of the 156 were relevant for predicting PCPV and 14 for predicting ORD. Both models achieved an R2 of over 0.8 on both the test and training set, as shown in the figures below, indicating good predictive power. Shown is PCPV model for reference.

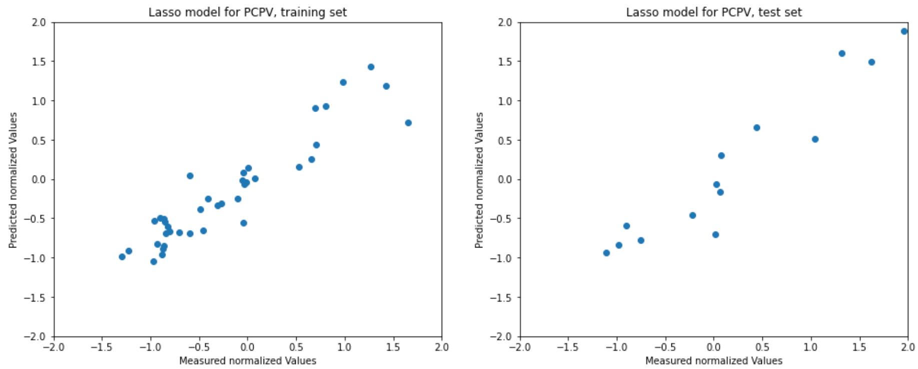

Figure 3. Predictive performance of the linear model with LASSO regularisation for predicting PCPV in West Virginia, 2013

The selected features with a coefficient magnitude greater than 0.1 are shown below for predicting PCPV:

Feature	Coef	Description
DOSAGE_UNIT	0.579750	Total number of opioid pills distributed
ORD_CDR	0.462571	Crude Opioid-Related Death Rate
F13214	0.143770	Number of home health agencies
NONMETRO	0.128336	Non-metropolitan indicator
SPEC_65T74_PC	0.106408	Medical specialists aged 65 to 74
F15474	-0.121680	Uninsurance rate for those under age 65 years
SPEC_45T54_PC	-0.176921	Medical specialists aged 45 to 54
F13911	-0.357173	Total black female population

Figure 4. Significant features for predicting PCPV in West Virginia, 2013

The workflow described above provides accurate predictions, however further analysis could be done using different regularisation techniques, such as Ridge Regression or Elastic Net.

6. CONCLUSION/ DISCUSSION

Work in Progress

7. REFERENCES

1. Basak, A., Cadena, J., Marathe, A., & Vullikanti, A. (2019). Detection of Spatiotemporal Prescription Opioid Hot Spots With Network Scan Statistics: Multistate Analysis. *JMIR Public Health Surveill*, *5*(2). 10.2196/12110
2. Brownstein, J., Green, T., Cassidy, T., & Butler, S. (n.d.). Geographic information systems and pharmacoepidemiology: using spatial cluster detection to monitor local patterns of prescription opioid abuse. *Pharmacoepidemiol Drug Saf.*, *19*(6), 627-37. 10.1002/pds.1939
3. Cerda, M., Gaidus, A., Keys, K. M., Ponicki, W., Martins, S., Galea, S., & Gruenewald, P. (2017). Prescription opioid poisoning across urban and rural areas: identifying vulnerable groups and geographic areas. *Addiction*, *112*(1), 103-112. https://doi.org/10.1111/add.13543
4. Cifuentes, M., Barbara S Webster, Verma, S., & Pransky, G. (2009). Geographic variation in opioid prescribing for acute, work-related, low back pain and associated factors: A multilevel analysis. *American Journal of Industrial Medicine*, *52*(2), 162-171. https://doi.org/10.1002/ajim.20655
5. Compton, W. M., Valentino, R. J., & DuPont, R. L. (2021). Polysubstance use in the U.S. opioid crisis. *Molecular Psychiatry*, *26*(1), 41-50. https://doi.org/10.1038/s41380-020-00949-
6. Duan, W. R., & Hand, D. J. (2021). Association between opioid overdose death rates and educational attainment – United States, 2010-2019. *Preventive Medicine*, *153*. https://doi.org/10.1016/j.ypmed.2021.106785.
7. Griffith, K.N., Y Feyman, SG Auty, EL Crable, TW Levengood. (2021). County-level data on U.S. opioid distributions, demographics, healthcare supply, and healthcare access. Data in Brief 35: e106779. https://doi.org/10.1016/j.dib.2021.106779
8. Hasan, M. M., Young, G. J., Patel, M. R., Modestino, A., Sanchez, L. D., & Noor-E-Alam, M. (2021). A machine learning framework to predict the risk of opioid use disorder. *Machine Learning with Applications*, *6*. 100144
9. Hu, X., Gallagher, M., Loveday, W., Connor, J.P., & Wiles, J. (2015). Detecting anomalies in controlled drug prescription data using probabilistic models. *Australasian Conference on Artificial Life and Computational Intelligence*, 337-349. https://doi.org/10.1007/978-3-319-14803-8_26
10. Maclean, J., Mallatt, J., Ruhm, C. J., & Simon, K. (2020). Economic Studies on the Opioid Crisis: A Review. *NBER Working Paper Series*. https://doi.org/10.3386/w28067
11. McDonald, D. C., Carlson, K., & Izrael, D. (2012). Geographic Variation in Opioid Prescribing in the U.S. *Journal of Pain Official Journal of the American Pain Society*, *13*(10), 988–996. https://doi.org/10.1016/j.jpain.2012.07.007
12. McDonald, D.C., & Carlson, K.E. (2014). The ecology of prescription opioid abuse in the USA: geographic variation in patients' use of multiple prescribers ("doctor shopping"). *Pharmacoepidemiology and Drug Safety*, *23*(12), 1258–1267. https://doi.org/10.1002/pds.3690
13. *Overdose Death Rates*. (2022, January 20). National Institute on Drug Abuse. Retrieved October 12, 2022, from https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates

14. Sun, F. (2022). Rurality and opioid prescribing rates in U.S. counties from 2006 to 2018: A spatiotemporal investigation. *Social Science & Medicine*, *296*(1), 114788–114788. https://doi.org/10.1016/j.socscimed.2022.114788
15. Unick, G. J., & Ciccarone, D. (2017). US regional and demographic differences in prescription opioid and heroin-related overdose hospitalizations. *International Journal of Drug Policy*, *46*, 112-119. https://doi.org/10.1016/j.drugpo.2017.06.003
16. Unick, G. J., Rosenblum, D., Mars, S., & Ciccarone, D. (2013). Intertwined epidemics: national demographic trends in hospitalizations for heroin- and opioid-related overdoses, 1993-2009. *PloS one*, *8*(2). https://doi.org/10.1371/journal.pone.0054496
17. *U.S. Overdose Deaths In 2021 Increased Half as Much as in 2020 - But Are Still Up 15%*. (2022, May 11). Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2022/202205.htm
18. Whelan, E., & Asbridge, M. (2013). The OxyContin crisis: Problematisation and responsibilisation strategies in addiction, pain, and general medicine journals. *International Journal of Drug Policy*, *24*(5), 402-411. https://doi.org/10.1016/j.drugpo.2013.01.007