Demographic Cofactor Analysis of Opioid Use in Tempe, Arizona and Cincinnati Ohio

Drew Gobbi
George Washington University
Washington, D.C., USA
drewgobbi@gwu.edu

Maddie Warndorf
George Washington University
Washington, D.C., USA
mwarndorf65@gwu.edu

Peiyu Wang
George Washington University
Washington, D.C., USA
oliviawpy@gwu.edu

***Abstract*— Misuse of prescribed and illicit opioids has become an epidemic in the United States that public health professionals are still struggling to understand the root causes for. This paper introduces a relational study between two proxies for opioid usage, overdose-related emergency service calls and the wastewater estimated number of users per/1000 to test for potential associations with medical facilities data and demographic cofactors in Tempe, Arizona, and Cincinnati Ohio. We find that opioid use tends to increase with the poverty rate, population density, and racially integrated neighborhoods in Cincinnati, and increases with the concentration of Hispanic and male populations in Tempe. We also identify key associations between measures of opioid use and certain medical facilities such as pain management centers and nursing homes in both cities. Both classification and regression results in this study are limited to low shares of explainable variance due to significant levels of noise in the datasets. To overcome this issue, we recommend that future models studying the relationships between opioid use and potential cofactors distinguish between opioid type in their analyses, leverage localized spatial units such as block groups, and focus on improving construct validity.**

Keywords—Opioids, Public Health, Machine Learning

I. Introduction

Misuse of prescribed and illicit opioids has become an epidemic in the United States that claims more than an 130 lives per day [1]. Understanding the root causes of this public health issue is an important step for public health professionals to help identify vulnerable populations and develop more effective treatment interventions. In this paper, we focus on two proxy variables of interest in Cincinnati, Ohio and Tempe, Arizona, opioid-related emergency service (EMS) calls and wastewater estimated number of opioid users per 1,000, to test for potential associations with demographic cofactors and medical facilities data.

We rely on common health narratives to form the following hypotheses on how opioid use may occur in our analyses:

- Indicators of low economic status, such as household income and the percentage of population below poverty could be positively associated with high heroin use.
- Wealthy, young professionals or middle-aged, advantaged population who were previously prescribed oxycodone could move to misuse the drug after becoming addicted.
- Indicators of gentrification, like neighborhood change, and economic inequality could be associated with opioid use.
- Certain health facilities (like substance abuse facilities) may have a positive association with opioid use while others (like pharmacies where people can drop off drugs) can have a negative association with opioid use.

II. Unit of Analysis and Dependent Variables

Our study uses two spatial-temporal measures, opioid-related EMS calls and wastewater estimated users per 1,000, to represent opioid usage.

EMS calls are measured as monthly counts per census block group in Cincinnati, OH and Tempe, AZ to represent overdoses and public health interventions to treat users. EMS calls represent only the extreme instances of opioid misuse where a medical professional would need to get involved. Because of this, we expect the EMS calls to miss instances of opioid misuse where a user exhibits less severe symptoms or does not receive direct treatment from an EMS provider. This makes EMS calls a conservative estimate of only the most extreme instances of illicit opioid use in a given area.

Figure 1: Heat map of EMS Calls in Cincinnati by block group, 2017-2019

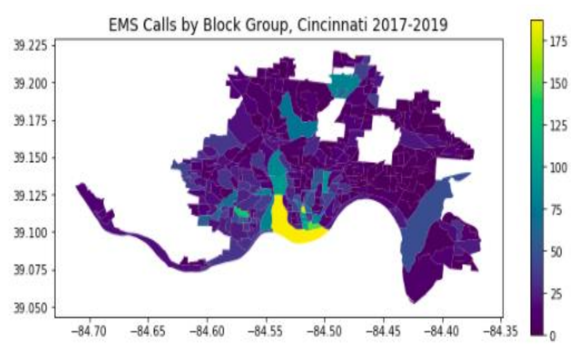
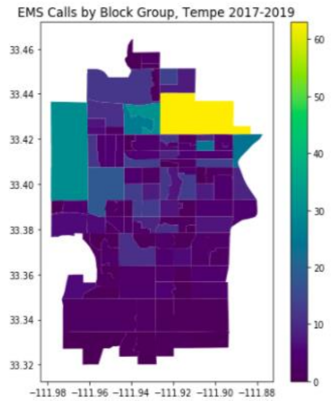
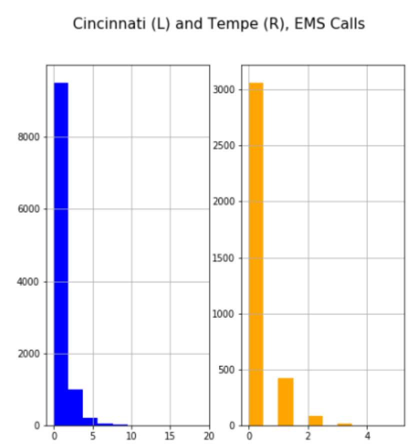


Figure 2: Heat map of EMS calls by block group in Tempe, 2017-2019



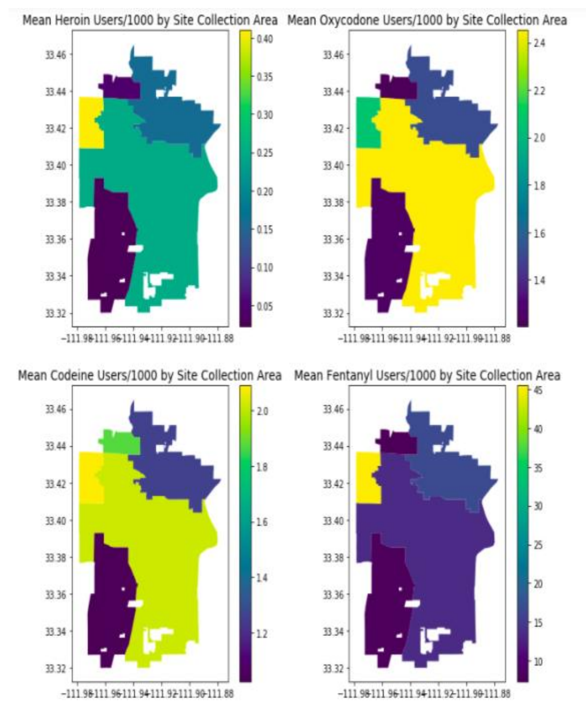
EMS calls in Cincinnati represent heroin related incidents, while calls in Tempe are representative of all opioid related incidents in the cities. Both datasets contain basic description information about the EMS call along with coordinate points and a timestamp. The number of points that occur within each block group are totaled up and segmented into the unique year-months that occur in the dataset. For example, a typical unit we could observe of opioid usage in the data would be three EMS calls that occurred in block group 13 in Cincinnati during the month of September 2017. There are 5,912 opioid-related EMS calls in Cincinnati across 10,792 possible observations (38 months multiplied by 284 block groups). In Tempe there are 675 EMS calls across 3,597 observations (33 months multiplied by 109 block groups). Histograms of both datasets that are shown below exhibit an exponential right skew because so many block group/year-month observations are zero. This indicates that the data may follow a Poisson distribution.

Figure 3: EMS Call Counts Distribution Cincinnati (L) and Tempe (R)



The second opioid usage proxy we examine is the wastewater estimated number of opioid users per 1,000. Wastewater estimated numbers of users are available for heroin, oxycodone, codeine, and fentanyl in Tempe only. The wastewater estimate is calculated by dividing the population normalized mass load (PNML) of an opioid compound by the average daily dose of that compound in mg/day. For example, if the heroin-related PNML reading is 1,000 mg/day/1,000, and the average heroin user consumes 50 mg/day, then we would estimate that there were 20 users per 1,000 in the area where the PNML reading was taken [2]. Wastewater estimates represent the total amount of an opioid “passed” by the population living in a given area. This means that the wastewater estimates represent total opioid use, both legal and illegal, regardless of whether a user receives treatment. As such, we consider the wastewater estimates to be an upper bound of potential misuse. The heat maps presented below show the mean wastewater estimated users in each of the five collection sites for each opioid.

Figure 4: Heat maps of mean wastewater estimated opioid users/1000 for Heroin (Top L), Oxycodone (Top R), Codeine (Bottom L) and Fentanyl (Bottom R)



Wastewater data are also available monthly, but are sampled at a different spatial unit than the EMS calls. Wastewater samples are taken from five site collection areas in the city of Tempe across 18 months. This accounts for a sample size around 80 observations, depending on how many values are dropped due to measurement error in the data. Histograms of the wastewater estimates vary depending on the type of drug, but generally exhibit a right skew. We take the natural logarithm of each wastewater estimate and assume the data to be log-normal. Even though this data transformation does not lead to perfectly gaussian data in some instances, we assume by the law of large numbers that the data will still tend towards normality as the low sample size increases, as we do not have any intuitive reason to believe that the data follow a different generating process.

Figure 5: Raw Distribution of Wastewater Estimated Users/1000 for Heroin (Top L), Oxycodone (Top R), Codeine (Bottom L) and Fentanyl (Bottom R)

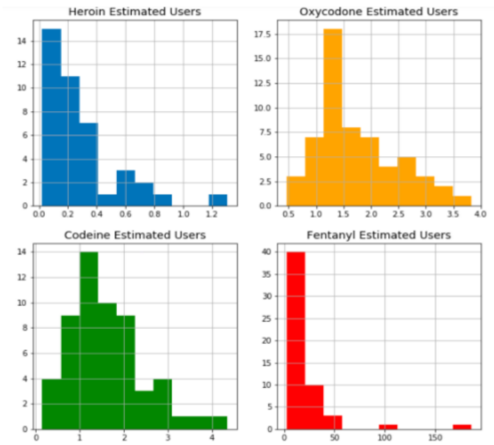
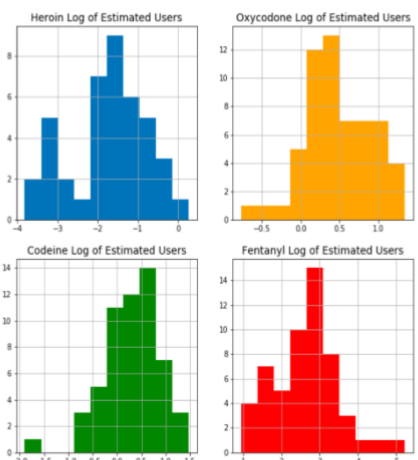


Figure 6: Log Distribution of Wastewater Estimated Users/1000 for Heroin (Top L), Oxycodone (Top R), Codeine (Bottom L) and Fentanyl (Bottom R)



Both the EMS call data and wastewater estimates exhibit some temporal variation that needs to be dealt with in our analysis to prevent from serially correlated error terms and potential spurious regression. We use augmented dickey fuller (ADF) tests and plots of the autocorrelation functions (ACF) and partial-autocorrelation functions (PACF) to examine the underlying time structure of the observations in each spatial unit in our study (e.g. either block group or wastewater site collection area). For the EMS calls, 14% of block groups in Cincinnati and 30% of block groups in Tempe are non-stationary in levels. Plots of the PACF and ACF for these series imply that they are integrated only to the first order. Accordingly, we examine non-stationary block groups in a separate sub-analysis for each city that focus on the monthly differences in opioid-related EMS calls per block group. In the wastewater data, two to three of the five site collection areas (depending on the drug) are non-stationary and levels and first differences. However, a further examination of the ACF and PACF plots tend to show only a single extremely high autocorrelation and partial-autocorrelation value at seemingly random lag points for each drug and site collection area (e.g. month five or twelve of the eighteen months). We attribute these random shocks to measurement error and opt to ignore them in our analysis. Only evaluating the time series levels is a potential source of bias in our model.

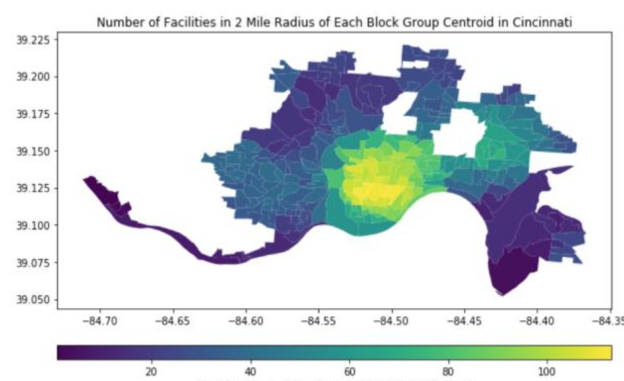
III. Cofactors Studied

We extract demographic cofactors to be used as independent variables in our analysis from the United States Census bureau's planning database file (PDB) at the block group level [3]. The PDB data contains measures from both the decennial census and the American Community Survey (ACS) five-year estimates. Because these variables are not available annually, the five-year ACS estimates or decennial Census estimates are held constant across our study. This means that our demographic predictors only vary over spatial units, not over time. To get demographic measures for the wastewater collection areas, we average the PDB measures of the block groups contained in each wastewater collection site.

We use cofactors representing different age groups, races, ethnicities, population density, and different indicators of economic status (such as median household income and home value) in our study. These covariates are similar to those used by Li et al. in a related analysis to ours [4].

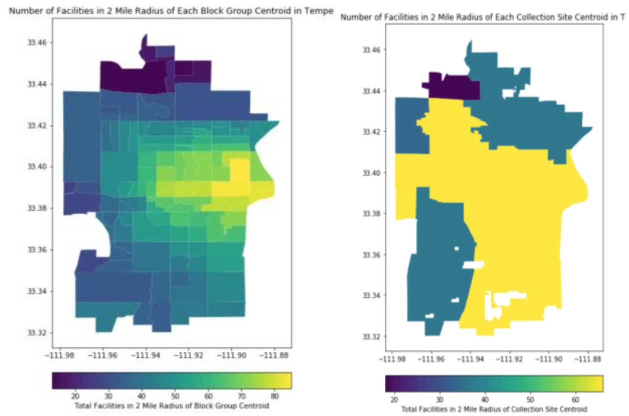
We extracted the facilities to represent medical facilities as well as drug-related facilities where opioids would either be distributed or collected legally. The facility data for Tempe was provided primarily through Arizona Department of Health Services' ArcGIS Service Directory [5, 6, 7]. The facility data for Cincinnati was provided through Ohio Public Health Information Warehouse along with several other leading opioid-based ArcGIS feature layers [8, 9, 10, 11, 12]. The facilities were counted within a 2 mile radius of the centroid for each spatial unit. The heat maps below in Appendix A5 shows the facility counts for Cincinnati and Tempe.

Figure 7: Heat map of medical facilities by block group in Cincinnati



Identify applicable funding agency here. If none, delete this text box.

Figure 8: Heat map of medical facilities by block group (L) and wastewater collection area in Tempe



The medical facilities represent hospitals, long-term care facilities, substance abuse facilities, mental facilities, physical therapy facilities, and urgent care facilities. The community facilities include pharmacies, naloxone distribution places, and drug drop-off locations. A full list of cofactors included is in the appendix.

IV. Experimental Design

We leverage a different framework to test for significance amongst our cofactors for each of our proxy variables. For EMS calls, we employ three different experiments in each city. First, we use a Logistic Regression to classify all zero and non-zero observations to determine which cofactors may increase the likelihood of an overdose or other opioid-related event occurring in a given block group and point in time. Then, we utilize a Poisson regression amongst the non-zero observations to test whether increases in these cofactors lead to an increase or decrease in the number of EMS calls in each block group and point in time. We assume that the natural log of the intensity parameter of the Poisson variable of interest (either monthly EMS calls or monthly EMS call differences in a block group) is linear for the covariates we include in the model. This Poisson regression is tested separately on the stationary and non-stationary subsets in levels and first differences respectively.

$$LOGIT(P_{EMSit}) = X_i^T\beta \qquad (1)$$

$$y_{it} \sim Poisson(\lambda_{it}) \qquad (2)$$

$$Log(\lambda_{it}) = X_i^T\beta \qquad (3)$$

For the wastewater data, we use a simple ordinary least squares (OLS) regression of the cofactors on the estimates of opioid users in each spatial-temporal unit. We test this model once for each type of opioid, for a total of four specifications.

$$y_{it} = X_i^T\beta \qquad (4)$$

V. Dicussion of Results

The logistic regression model performs poorly in terms of its ability to classify observations with and without EMS calls in Cincinnati and Tempe. The model AUC in Cincinnati is 0.60, while it is only 0.52 in Tempe. This suggests that using just demographic and medical facilities data alone to classify the probability of an EMS related event has only a 50-60% success rate in our sample. Despite poor performance, there are still some interesting patterns that emerge amongst our cofactors in each city.

Figure 9: Receiver Operator Characteristic - Cincinnati

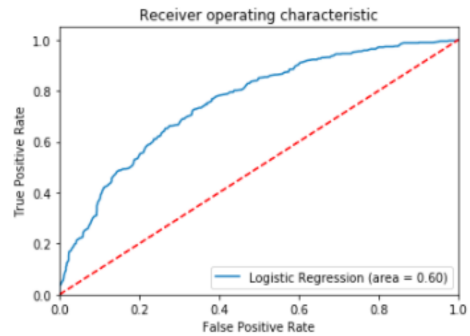
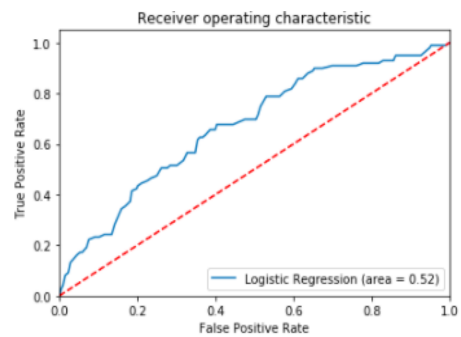


Figure 10: Receiver Operator Characteristic - Tempe



In Cincinnati, heroin use appears to increase with the rate of poverty, population density, and the number of pain or substance management facilities within each block group. The logistic regression model also finds that heroin use is less likely to occur in block groups with a large percentage of individuals with a college degree or higher, where there are larger numbers of households with all black or all white members, or around physical therapy centers, nursing homes, and children's hospitals. From a demographic standpoint, the results seem to imply that poor urban areas with diverse, less educated populations are more likely to see opioid-related EMS calls. Likewise, EMS calls appear to be more likely near pain or substance management centers, and less likely near nursing homes and children's hospitals. Both of these findings make sense intuitively. Individuals in or around pain or management facilities who overdose are among the most likely to have an EMS call placed to help them due to interventions by the staff. Similarly, we would not expect opioid use - and particularly Heroin use - to take place in areas where there is a high concentration of children's hospitals or nursing homes. These results largely extend to the Poisson regression analysis of all stationary non-zero observations, but do not replicate well in the analysis of the non-stationary block groups even after differencing has taken place. We are unsure if this divergence is model related, though it is not immediately apparent that the differences we observe are a functional difference in the characteristics of these block groups.

Table 1: Opioid EMS Call Results, 5% Statistical Significance

	Cincinnati EMS Logit	Cincinnati EMS Stationary Block Groups	Cincinnati EMS in Differences Non-Stationary Block Groups	Tempe EMS Logit	Tempe EMS Block Groups	Tempe EMS Non-Stationary Block Groups
Positive Cofactors	%25-44, %Asian Alone, %Below Poverty, Log Pop, Women's Clinics, Pain Management, Substance Abuse, Other Medical Facilities, Pharmacies	%Below Poverty, Log Pop, Pain Management, Substance Abuse, Naloxone Distribution	%Male, Pain Management, Other Medical Facilities	%Males, %Hispanic, %College, Psych Hospitals, Child Facilities, Drug Drops	N/A	%Male, %Hispanic, %Asian, %Below Poverty, %College, Hospitals, Urgent Care, Physical Therapy, Nursing Homes, Drug Drops, Naloxone
Negative Cofactors	%18-24, %White Alone, %Black Alone, %College, Urgent Care, Physical Therapy, Nursing Homes, Children's Hospitals	%White Alone, %Black Alone, %College, Physical Therapy, Nursing Homes, Children's Hospitals, Pharmacies	%25-44, %45-64, %Asian, Med HHI, %Below Poverty, Pharmacies	%18-24, %25-44, %65+, Median House Value, Surgical Centers, Urgent Care, Pain Management, Mental Facilities	N/A	%25-44, %Black Alone, Med HHI, Log Pop, Substance Abuse, Pharmacies

Results in Tempe are less conclusive than in Cincinnati. EMS calls reflect the city's different demographics compared to Cincinnati and tend to increase with the number of males, Hispanic individuals, and individuals with a college degree or higher in each block group. These results replicate in the sub-analysis of the differenced data in the non-stationary block group, but no cofactors appear to be statistically significant in our analysis of stationary block groups. A potential reason for these null results could be the inherent noise built into the EMS call data, as all opioid related responses are grouped together in Tempe. Different types of populations may be positively or negatively associated with different drugs, and these relationships might offset when examined together in the same model. For instance, if higher income individuals were negatively associated with heroin use and positively associated with oxycodone use and both types of drugs had the same number of overdoses in the EMS call data, we may see a null effect.

Wastewater data in Tempe helps to account for the noise created by mixing different opioid EMS calls together because it isolates the drugs by parent compound. Here, we see a positive association between Hispanic individuals and heroin and fentanyl use, as well as an additional positive relationship between men and fentanyl. We also observe some similarities between oxycodone and codeine users, which both tend to increase with the number of individuals 45-64 and the concentration of nursing homes and pharmacies. These appear to represent legal opioid use, where high pain individuals may be more or less likely to be prescribed these opioids. There is also a considerable amount of noise and the data, including what appears to be glaring false positive associations between things like the number of nursing homes in heroin use in the site collection areas. We attribute the detection of these relationships to noise introduced by the large spatial units in the dataset. The land area of several of the water site collection areas - particularly site collection area 1, which is about half the city -- encompasses several block groups and distinct neighborhoods. This may be grouping together distinct neighborhoods and their characteristics in our analysis that may not actually be related to one another, creating the potential for confounds.

Table 2: Tempe Wastewater results by Drug, 10% Statistical Significance

	Heroin	Oxycodone	Codeine	Fentanyl
Positive Cofactor	%Hispanic, Psych Hospitals, Mental Facilities, Other Medical Facilities, Nursing Homes, Drug Drop	%45-64, %65+, %White Alone, Other Medical Facilities, Nursing Homes, Pharmacies	%45-64, Mental Facilities, Nursing Homes, Pharmacies	%Male, %25-44, %Hispanic, Mental Facilities, Other Medical Facilities, Pharmacies, Drug Drop
Negative Cofactor	%18-24, Median House Value, Hospitals, Urgent Care, Substance Abuse, Hospice, Child Facilities, Pharmacies	%18-24, %Below Poverty, EMS calls, Drug Drop	%18-24, Hospitals, Hospice, Child Facilities, Physical Therapy, Naloxone Distribution	%18-24, %Black Alone, Med HHI, Urgent Care, Substance Abuse, Nursing Homes, Hospice

VI. Limitations, Future Considerations, and Recommendations to Policy Makers and Data Analysts

As previously discussed, there are several key limitations in our analysis that could be impacting our results and should be considered in future studies of opioid misuse. First, we believe that there could be an ecological fallacy in the wastewater site collection areas that may be conflating neighborhood level characteristics in the dataset. Because site collection areas encompass multiple block groups, they may average out or misrepresent demographic effects (such as a high population of males for example) in smaller spatial subunits. This has implications for our ability to detect useful findings in the data, and may also introduce spatial noise into the analysis. Similarly, we believe that the wastewater data indicates that user characteristics may vary substantially depending on the type of opioid under study. EMS data, like what is available in Tempe, that groups all opioid related incidents together may encompass noise generated by offsetting relationships in the data. Finally, research by Bates et al. has indicated that overdoses and EMS calls -- particularly for highly addictive substances like heroin -- are more likely to occur near where a user were to buy the substance than in the area where they live. This may mean that household demographic characteristics may be less predictive of opioid use than other neighborhood environmental features, such as the concentration of vacant lots or parking lots, where misuse may be more likely to occur.

These limitations imply a few distinct takeaways for city administrators and public health professionals as they continue to collect and analyze opioid data available to them. First, opioids and opioid users are not interchangeable from one another and should be considered in separate analyses. Different substances appear to have different user profiles and root causes, and will likely need different intervention strategies to deliver effective treatment to the population. Second, because opioid use varies spatially, data collection should continue to focus on smaller spatial units of analysis to effectively localize the opioid issue. The majority of opioid related measures are not available in units smaller than a city, and most do not extend to the census tract, block group, or geographic point levels. Providing these data can help combat ecological fallacies that may occur in larger spatial units and tie misuse to specific household or environmental characteristics within a neighborhood. Finally, because opioid use itself is such a difficult construct to measure, public health administrators and data analysts should continue in their exploration and standardization of novel datasets to represent this construct. EMS calls and wastewater derived estimates of opioid use are both limited snapshots of opioid use, and should be compared to more measures to help improve the construct validity of findings.

References

[1] "Opioid Basics", *Opioid Overdose*, 2019. [Online]. Available: https://www.cdc.gov/drugoverdose/epidemic/index.html. [Accessed: 10- Dec- 2019].

[2] Arcgis.com. 2020. Fighting Opioid Misuse By Monitoring Community Health. [online] Available at: https://www.arcgis.com/apps/Cascade/index.html?appid=92073d7f6a6a498b987f2afdab1b9471. [Accessed 23 April 2020].

[3] "2019 Planning Database Tract Data", Planning Database with 2010 Census and 2013 – 2017 American Community SurveyData, 2019. [Online]. Available: https://www.census.gov/topics/research/guidance/planning-databases.2019.html. [Accessed: 10- Oct- 2019].

[4] Z. Li et al., "Suspected heroin-related overdoses incidents in Cincinnati, Ohio: A spatiotemporal analysis", *PLOS Medicine*, vol. 16, no. 11, 2019. Available: https://journals.plos.org/plosmedicine/article/file?id=10.1371/journal.pmed.1002956&type=printable. [Accessed 10 November 2019].

[5] "AZLicensedFacilities (FeatureServer)", *ArcGIS REST Services Directory*, 2019. [Online]. Available: https://services1.arcgis.com/mpVYz37anSdrK4d8/ArcGIS/rest/services/AZLicensedFacilities/FeatureServer. [Accessed: 10- Oct- 2019].

[6] "DrugDropOffLocations (FeatureServer)", *ArcGIS REST Services Directory*, 2019. [Online]. Available: https://services1.arcgis.com/mpVYz37anSdrK4d8/ArcGIS/rest/services/DrugDropOffLocations/FeatureServer. [Accessed: 10- Oct- 2019].

[7] "Pharmacies (FeatureServer)", *ArcGIS REST Services Directory*, 2019. [Online]. Available: https://services1.arcgis.com/mpVYz37anSdrK4d8/ArcGIS/rest/services/Pharmacies/FeatureServer. [Accessed: 10- Oct- 2019].

[8] "Ohio OneSource", *Ohio Public Health Information Warehouse*, 2019. [Online]. Available: http://publicapps.odh.ohio.gov/EDW/DataBrowser/Browse/OhioOneSource. [Accessed: 10- Oct- 2019].

[9] “Health Care Facility Directories", *Health Care Facilities and Regulations*, 2019. [Online]. Available: https://chfs.ky.gov/agencies/os/oig/dhc/Pages/hcf.aspx. [Accessed: 10- Oct- 2019].

[10] J. Lindemann, "National Prescription Drop Box Layer", *ArcGIS*, 2018. [Online]. Available: https://www.arcgis.com/home/item.html?id=7726ad6095c543b2a75ffae24b27e419#overview. [Accessed: 10- Oct- 2019].

[11] T. Peacock, "Walgreens Drug Kiosk Locations", *ArcGIS*, 2016. [Online]. Available: https://opioidepidemic.maps.arcgis.com/home/item.html?id=835b14782ffb4c259c3c2913e68ca0ef#overview. [Accessed: 10- Oct- 2019].

[12] "Ohio Pharmacies Dispensing Naloxone Without a Prescription", *Naloxone Pharmacy*, 2019. [Online]. Available: https://www.pharmacy.ohio.gov/Licensing/NaloxonePharmacy.aspx. [Accessed: 10- Oct- 2019].

[13] Savannah Bates, Vasiliy Leonenko, James Rineer, Georgiy Bobashev, “Using synthetic populations to understand geospatial patterns in opioid related overdose and predicted misuse” *Computational and Mathematical Organization Theory* (2019): 25: 36-47