Geospatial Clustering Analysis on Drug Abuse Emergencies

Jinha Lee
Bowling Green State University
jinhal@bgsu.edu

Jung Im Choi
Bowling Green State University
choij@bgsu.edu

Bai-Yau Yeh
Bowling Green State University
byeh@bgsu.edu

Qizhen Lan
Bowling Green State University
qlan@bgsu.edu

Hyojung Kang
University of Illinois at Urbana-Champaign
hyokang@illinois.edu

Abstract

The epidemic of drug abuse is a serious public health issue in the U.S. The number of overdose deaths involving prescription opioids and illicit drugs has continuously increased over the last few years. The objective of this study is to develop a geospatial model that identifies geospatial clusters in terms of socioeconomic, and demographic characteristics with an unsupervised machine learning algorithm. Then, we suggest the most important features affecting heroin overdose both negatively and positively. The findings of this study may inform policymakers about strategies to mitigate the drug overdose crisis.

1. Introduction

Although slightly decreasing from 2017 to 2018, opioid-related overdose remains a leading cause of injury-related mortality in the US, with nearly 70% of drug overdoses involving opioids [1]. In general, opioids are a class of drugs used in reducing pain. The categories of opioids include natural opioid analgesics (morphine and codeine), semi-synthetic opioid analgesics (oxycodone, hydrocodone, hydromorphone, and oxymorphone), methadone, synthetic opioid analgesics (other than methadone, includes drugs, such as tramadol and fentanyl). Lastly, heroin is also an illegal opioid processed from morphine and extracted from certain poppy plants. Its use has also increased across the US among men and women, most age groups, and all income levels. In 2017 alone, there were 70,000 fatalities in the US which is three times more than the number reported in 2000 [2].

In particular, Ohio is one of the most seriously affected states regarding opioid abuse and death. A rate of 39.2 deaths per 100,000 persons is the second-highest rate in the US and 63.5 opioid prescriptions for every 100 persons is also much higher than the national average [3]. Figure 1. shows that Ohio's drug overdose rate is also higher than the US average and rapidly increasing [4]. Addiction and overdose-related to opiates have reached an epidemic level, creating an unprecedented crisis. In addition, the costs of this problem extend beyond just healthcare, including those tied to lost productivity, addiction treatment, and criminal justice involvement, as well as the many social costs that are challenging to quantify. The epidemic's effects are being felt by commercial healthcare, pharmacies, government agencies and programs, and every industry which employs its victims. Therefore, it is imperative to identify individuals most likely to develop opioid abuse or dependence to inform large-scale, targeted prevention efforts [5].

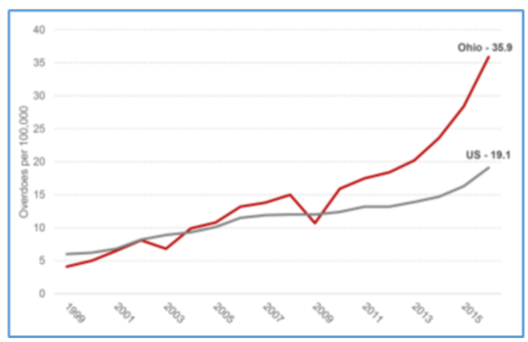

Figure 1. Drug Overdose Rates 1999-2016

The Centers for Disease Control and Prevention (CDC) has been implementing various efforts for preventing opioid overdoses, such as Prescription Drug Monitoring Programs (PDMPs), Enhanced State Opioid Overdose Surveillance (ESOOS), Overdose Data to Action (OD2A), and Data-Driven Prevention Initiative (DDPI) [6]. However, Ohio is currently funded for only PDMPs at the statewide level and only three counties in Ohio are funded for OD2A. In addition, because the prevalence of opioid addiction and resources to address the crisis vary across Ohio, there is no standard prevention and monitoring model, and limited resources

URI: https://hdl.handle.net/10125/80036
978-0-9981331-5-7

for opioid addiction prevention services are often not allocated optimally, based on the areas of highest need.

To date, research has shown that opiate addiction is associated with various socioeconomic factors. Several types of research have shown the importance of the measure of racial/ethnicity [7], opioid treatment [8], and trend analysis [9] with the modifiable areal unit problem (MAUP) to analyze the geospatial patterns of the opioid problem in terms of community-level as well as personal level. However, the geospatial analysis of opioid deaths by epidemiologists and healthcare researchers had been limited to higher geographical aggregates such as cities or, more often, provinces and states [10]. The primary reason for this is that, historically, deaths due to opioid overdose were significantly fewer than for other drugs. Although the rising number of annual opioid overdose deaths indicates that the opioid epidemic has not yet peaked, the relative contribution of different drug types to the epidemic is changing [11]. The dynamic nature of the opioid overdose epidemic poses continuous challenges to prevention efforts [12]. Lack of knowledge about vulnerabilities in a specific community, such as "hotspots" and "red-flagged times" causes challenges in responding to opioid-related incidents at the local level. Therefore, there is a critical need for local communities to understand accurate risk "patterns" in opioid-related incidents, to develop and deliver a more nuanced prevention strategy, based on local needs. To effectively deploy policies and strategies for drug abuse in local communities, it is important to understand the spatial and temporal distributions of abuse risk promptly [13].

In this paper, we present a geospatial analysis of the locations of reported heroin-related incidents associated with EMS dispatches in the city of Cincinnati, Ohio. We investigated the geospatial profile variability as a function of socioeconomic and demographic covariates, accessibility of medical facilities, and characteristics of the community environment. We applied an unsupervised machine learning algorithm to stratify the city of Cincinnati into subgroup clusters with similar covariates in terms of geospatial socioeconomic features.

2. Materials and methods

2.1. Cincinnati EMS data processing

Emergency medical services (EMS) and first responders are critical parts of the emergency care system in the US and the first phase of emergency care [14]. There are more than 20 million EMS transports each year, and emergency 9-1-1 services offer immediate access to an operator who can provide basic life support coaching until help arrives on the scene [15]. In most cases, states and EMS have time limits within which patient care records must be submitted (24-72 hours), offering more timely information about suspected overdoses [13]. EMS dispatch datasets usually also have a high spatial resolution, with global positioning system (GPS) locations or addresses in the call records, making them a valuable resource for understanding characteristics of each overdose incident that happens [16] and for developing opioid use prevention programs [17]. However, EMS calls labeled by the dispatcher as related to overdose or opioids may not represent all such incidents and calls to EMS may be incorrectly labeled by dispatchers as heroin-related based on information obtained from the caller [13]. Although EMS records may contain glitches, information from EMS records can be considered the most timely and readily available data to local authorities for appropriate response [18], [19].

In this paper, we obtained EMS response data related to heroin overdose from the City of Cincinnati's computer-aided dispatch (CAD) database [20]. Figure 2. depicts a data processing process with the EMS data set retrieved from the City of Cincinnati. The EMS data is publicly available and captures all responses by the Cincinnati Fire Department to reported heroin overdose incidents. The CAD's EMS data contains up to 6.3K heroin-related overdoses (OD) and 1.8K of other overdoses incidents in Cincinnati from 08/01/2015 to 01/30/2019. Each incident recorded location information of an incident with GPS locations (i.e., latitude and longitude), address, neighborhood (e.g., Downtown, West End, Queensgate), start and end date/time of the incident, and disposition of the incident response (e.g., medic transport, investigation, cancel) [21]. We excluded incidents outside of the study area, without geospatial information, and with disposition codes not associated with medical events (e.g., not a disposition, fire disregard, reassigned), canceled, duplicated, or false alarms (e.g., false medical situation, medical response false) [21].

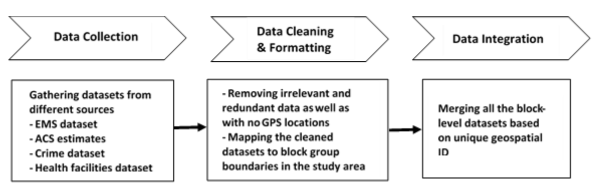

Figure 2. Data collection and processing process

To count heron incidences by census block groups during the study period, we linked the latitude and longitude of each incident to the regions of block groups in the study area that can be represented on a map. Then, the average number of heroin-related incidents was estimated throughout 2015-2019 in each of these regions, and it was used to estimate a block-level heroin-related incident rate per 1000 adult population [21].

2.2. Covariates

We collected covariates information from variable sources. First of all, we gather information on the population in each block group from the 2013 to 2017 estimates from US Census Bureau's American Community Survey [22]. In addition, healthcare facilities datasets were from Health Resources and Services Administration (HRSA) [23] Data Warehouse and SAMHSA OTP Directory [24]. Demographic information included the adult population size, and the percentage of the population by age, gender, and race/ethnicity. Table 1 shows the geospatial covariates to conduct our data analytics approach.

Table 1. Complete list of covariates

Variable Name	Description	Variable Name	Description
population	Population size	pc_nonhispanic_white	The proportion of nonhispanic white
pc_bachelor	The proportion of bachelor's degree or higher	pc_nonhispanic_black	The proportion of nonhispanic black
pc_poverty	The proportion in poverty	pc_hispanic	The proportion of Hispanic
pc_bus_half	The proportion of half-mile bus coverage	EF Theil.s.entropy	Ethnicity Fractionization Theil's entropy score (Diversity score)
fire	Distance to fire departments	pc_male	The proportion of male
pc_park	The proportion of parks	pc_age18_24	The proportion of aged 18-24
pharm	Distance to pharmacies	pc_age25_34	The proportion of aged 25-34
hospital	Distance to hospitals	pc_age35_49	The proportion of aged 35-49
fqhc	Distance to federally qualified health centers	pc_age50_64	The proportion of aged 50-64
otp	Distance to opioid treatment programs	pc_age65up	The proportion of aged 65 up
bup	Distance to Buprenorphine practitioners	crime_rate	Crime rate per population
pc_commercial	The proportion of commercial zoning	Appalachian	Appalachian score
pc_downtown	The proportion of downtown development zoning	popdens	Density of population
pc_manufacturing	The proportion of manufacturing zoning	per_cap_income	Per capita income
pc_office	The proportion of office zoning	housing_units	Number of housing units
pc_residential_other	The proportion of other residential zoning	pc_vet_XXX	The proportion of veteran/non-veteran
pc_development	The proportion of planned development zoning	pc_pop_age25+_XXX	The proportion of educational levels
pc_riverfront	The proportion of riverfront zoning	pc_pop_age3+_XXX	The proportion of public/private school enrollment
pc_singlefamily	The proportion of single-family zoning	pc_urban_mixed	The proportion of urban mixed zoning

Each covariate was calculated based on the census block level. In addition, additional socioeconomic covariates such as Theil's entropy score, Appalachian score developed for an analysis of social needs in the City of Cincinnati (See Appendix) [25], veteran status, educational levels of adults, and public/private school enrollments were retrieved from US Census Bureau [18]. Lastly, the crime incidents data between 2015 and 2019 were obtained from the Cincinnati Police Department [26]. Likewise, the same procedure was performed to compute the average crime rate per 1,000 adult population in each census block.

2.3. Methods

To effectively identify geospatial similarity in terms of the covariate, we applied a geospatial clustering model armed with an unsupervised machine learning algorithm.

Machine learning methods are commonly classified into supervised and unsupervised methods. Supervised methods, such as Support Vector Machines [21] and Random Forests [22], [23] have been extensively used in various fields. These methods classify new objects to a determinate set of discrete class labels while minimizing an empirical loss function (e.g., mean square error) [24]. However, supervised methods require the use of a training set that contains a priori information of several objects' class labels. In contrast, unsupervised methods do not require a training set that contains a priori information of objects' class labels as input. Unsupervised methods can detect potentially interesting and new cluster structures in a dataset. Moreover, they can be implemented when class label data is unavailable [27]. Therefore, unsupervised machine learning is well appropriate for our research since the objective of our study is to discover the class labels that best describe a set of data. Clustering has been one of the most popular unsupervised machine learning algorithms. Clustering refers to techniques for grouping similar objects in clusters [28]. Since the objective of the study is to discover the class labels that are determined by similarity as stated above, we applied an unsupervised machine learning clustering algorithm, especially the K-Means algorithm to define clusters in the city of Cincinnati based on EMS data.

K-Means algorithm partitions the data set into several cluster K that have been set up in the beginning. Partition data sets are performed to determine the characteristics of each cluster, so clusters that have similar characteristics are grouped into one cluster and that have different characteristics grouped into other clusters [29]. The advantages of the K-Means algorithm are that the required execution time is relatively fast and easy to implement. However, it is very tricky to determine the centroid of the cluster or the initial centroid randomly selected. Therefore, we evaluated the centroid determination process by the K-Means algorithm using the Davies-Bouldin index (DBI). DBI is a metric for evaluating clustering algorithms which are widely used for measuring the goodness of split by a K-Means clustering algorithm for a given number of clusters [30]. Cluster evaluation using the DBI uses an internal evaluation scheme in which the cluster results can be seen whether the quantity and proximity of the cluster data result. DBI's criteria are based on the ratio in clusters and the distance between clusters. In the K-Mean's formulation, the cohesiveness of the corresponding clusters and the separation between them is the main parameter that distinguishes one cluster from another. Thus, k is the number of clusters, the smaller the DBI value obtained, the better the clusters obtained from clustering using the K-Means clustering algorithm. As a result, we could produce a proper number of clusters that have a good level of similarity with given EMS data and covariates.

All analyses were conducted with Python 3.9, including the packages "scikit-learn 0.24.2" for determining the number of clusters and cluster exploration, "Matplotlib 3.4.2" and "seaborn 0.11.1" for visualization

3. Results

3.1. Number of the clusters

Before identifying clusters based on geospatial covariates, we evaluated the goodness of split by a K-Means clustering algorithm to determine a proper number of the clusters. To avoid preselecting input parameters a priori (e.g. the number of clusters), previous researches have implemented cluster validation metrics [29]-[35]. Hence, we applied the DBI score to the corresponding k randomly selected and determine a proper number of the clusters based on the minimum DBI score. Figure 3 shows the result of the DBI score analysis.

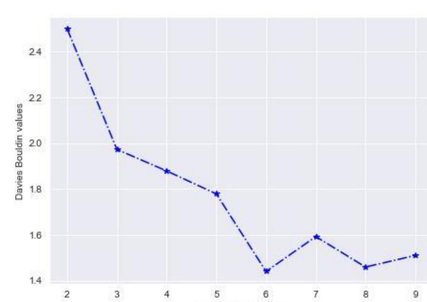

Figure 3. Davies-Bouldin Index Analysis

The best DBI score was 1.439826 at six clusters when we applied the K-Mean clustering algorithm with the data set.

3.2. Clustering results

Since the proper number of the clusters was identified, the clustering procedure with the unsupervised machine learning techniques was used for the City of Cincinnati's EMS data with covariates. In particular, our analyses were conducted based on K-Means clustering algorithms with six clusters on 280 US Census blocks in terms of geospatial and socioeconomic covariates presented in Table 1. The clusters identified by the K-Means algorithm are shown in Figure 3, express as geographical mapping. Since two independent cities, Norwood and St. Bernard in Hamilton County, OH, are not governed by the city of Cincinnati, two white blocks in Figure 4 are excluded in the analysis.

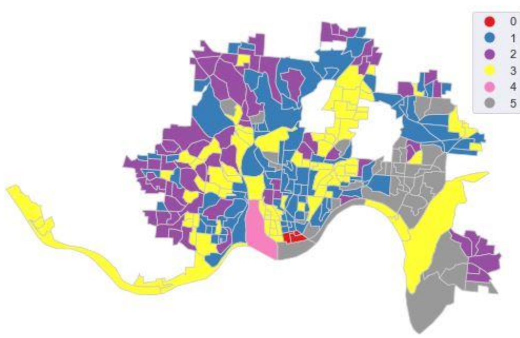

Figure 4. Clustering result in City of Cincinnati with K-Means clustering algorithm

Table 2 shows how many blocks each cluster has.

Table 2. Number of blocks in each cluster

Cluster	Number of blocks	Average heroin overdose incident rate
0	3	85.6314
1	92	15.1278
2	66	9.1742
3	80	18.2073
4	1	334.6696
5	38	5.0973

Cluster 4 was identified as the highest heroin overdose incident group with a single block, cluster 0 was also identified as a relatively higher heroin overdose incident group with 3 blocks. Meanwhile, clusters 1, 2, 3, and 5 were identified as relatively lower heroin overdose incident groups.

Collapsing the results across features within each cluster can provide further insight into cluster-level characteristics. Table 3 shows the clustering result with some selected features. Cluster 4 shows the characteristics of the highest crime rate, higher proportion of the male population, the lowest educational level, mostly manufacturing zone, low housing units, very young populations (age 18-24), and the close distance to the Buprenorphine practitioners. In general, cluster 4 is matched to the research result of a heroin overdose in a micropolitan area [36]. However, one thing particularly interesting in cluster 4 is that this area shows the highest income level among 240 blocks in the City of Cincinnati. This result is not well matched the characteristics of a micropolitan area. Cluster 4 area in the City of Cincinnati is "Queensgate" which sits in the valley of Downtown Cincinnati and has been dominated by industrial and commercial warehouses. The population of Queensgate has drastically decreased since 2010 and it caused the highest variation in the per-capita income. In other words, the very little number of highest income group dominates the income effect on the analysis. Maloney and Auffrey reported that the social needs should be addressed in the Queensgate area to reduce various problems including opioid addiction [25]. Another problem in the cluster 4 Queensgate area is that this area has been the hot-spot in illicit drug trading [21], [26].

Cluster 0 shows different characteristics compare to cluster 4 despite both clusters record higher overdose incident rates. The cluster 0's characteristics can be summarized as the highest education level, completely downtown area, white-collar working population (age 25-49), less racial diversity, higher income level, and the closest distance to the Buprenorphine practitioners. Cluster 0 is relatively similar to the characteristics of the metropolitan area [36]. Cluster 0 is the downtown area in the City of Cincinnati that shows the features of the built environment, including the proportion of parks, commercial, manufacturing, and downtown districts and the number of fast-food restaurants, exhibit strong positive associations with the number of heroin-related calls.

Clusters 1, 2, and 3 show that they are suburban areas with low-and middle-income matched to the small-town characteristics [36]. Among the cluster 1, 2, and 3, clusters 1 and 2 show less economic disparities such as poverty level and income level. Meanwhile, cluster 3 shows a relatively lower education level, lower income level, and higher poverty level than clusters 1 and 2. In other words, economic stressors could be one of the contributors to a heroin overdose in small-town suburban areas.

However, cluster 5 shows that it is a wealthy suburban area [37] with very low racial diversity, very high income, and very high educational level. The cluster 5 area is identified as a non-Hispanic white residence area well equipped with support programs such as community-based opioid overdose recognition and response training programs, and a quick response team to revisit overdose victims within 2 weeks [13].

3.3. Feature selection

To develop further geospatial profiling and community-based overdose prevention strategy, a feature selection procedure based on random forest regression was conducted with a 10-fold cross-validation random search. With the complete list of covariates, we ranked the most important features to contribute to the incident rate. Figure 5 shows the top 15 important features based on the random forest regression algorithm.

The most important covariate is the crime rate in the block. The crime rate has a positive relationship with the overdose incident rate.

Table 3. Clustering results with covariates

Cluster	incident_rate	crime_rate	popdens	pc_male	pc_bachelor
0	85.6314	0.0424	5862.7817	0.5680	0.5225
1	15.1278	0.0124	6738.4864	0.5063	0.2171
2	9.1742	0.0075	6510.6867	0.4625	0.1776
3	18.2073	0.0138	4947.7543	0.4785	0.1321
4	334.6696	0.1276	131.7923	0.8265	0.0255
5	5.0973	0.0037	5074.8135	0.4904	0.5339
Cluster	**pc_downtown**	**pc_manufacturing**	**Theil.s.entropy**	**pc_nonhispanic_white**	**pc_nonhispanic_black**
0	0.9928	0.0000	0.8115	0.7697	0.1012
1	0.0107	0.0937	0.9972	0.4859	0.3583
2	0.0000	0.0240	0.7194	0.4602	0.4736
3	0.0071	0.0914	0.4735	0.3390	0.6307
4	0.0000	0.8081	1.2764	0.2092	0.4694
5	0.0217	0.0165	0.5349	0.8389	0.0766
Cluster	**pc_age18_24**	**pc_age25_34**	**pc_age35_49**	**pc_age50_64**	**pc_age65up**
0	0.1563	0.2943	0.2242	0.1461	0.1537
1	0.1700	0.1953	0.1691	0.1719	0.0983
2	0.1100	0.1763	0.1724	0.1781	0.1231
3	0.1042	0.1462	0.1645	0.2225	0.1426
4	0.5918	0.0459	0.0663	0.0306	0.0255
5	0.0806	0.2174	0.1992	0.1705	0.1529
Cluster	**pc_poverty**	**household_income**	**housing_units**	**bup**	**hospital**
0	0.2267	71835.0000	443.3333	0.1706	1.1332
1	0.2101	36576.1975	520.6196	1.0367	1.5749
2	0.2216	37807.9419	838.0303	2.0681	1.5129
3	0.3178	30569.1129	430.5750	1.4061	1.7458
4	0.1000	129167.0000	5.0000	0.3802	1.7354
5	0.0415	84819.6053	606.1579	0.9110	2.7898

In addition, population density, the proportion of downtown zoning, the proportion of the male population, the proportion of manufacturing zoning, and diversity score are positively associated with the overdose incident rate. Meanwhile, educational level, age group, and diversity are negatively associated with the overdose incident rate. Interestingly, economic burden measures such as income and overdose support programs such as the distance to the Buprenorphine practitioners are a relatively low impact on the overall incident rate in the given data set. Another interesting point is that diversity is negatively associated with the overdose incident rate. It is known that the annual age-adjusted death rates for drug overdose deaths that involved an opioid significantly increased for all racial/ethnic groups in metropolitan and non-metropolitan areas from 1999 to 2017. The largest average annual increases in rates occurred among non-Hispanic whites in non-metropolitan areas and medium-small metropolitan areas, followed by non-Hispanic blacks in medium-small metropolitan areas [38]. However, the city of Cincinnati shows that higher diversity of racial/ethnicity could lead to a lower heroin incidents rate.

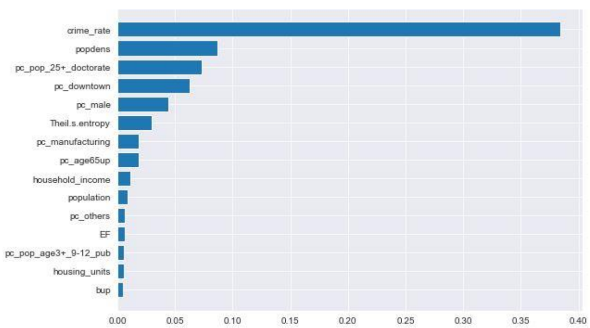

Figure 5. Top 15 important features

4. Discussion

The primary objective of this study was to identify geographic profiling in terms of socioeconomic and demographic features in the place of residence of patients who abused heroin reported to the EMS. In particular, we identified areas as heroin-related incidents by EMS dispatchers in the city of Cincinnati, along with sociodemographic variables and features of the built environment associated with overdose counts. We used K-Means unsupervised machine learning algorithms to identify clusters that consist of blocks with similarities in terms of given covariates. In addition, we determined the appropriate number of the clusters using the Davies-Bouldin index to avoid preselecting input parameters a priori. As a result, we could split the City of Cincinnati into six distinct clusters stemming from the similarity of each level of census block groups.

Originally fueled by prescription opioids, recent rises in overdoses are now driven by heroin and fentanyl, which is causing serious overdose mortality. Applying unsupervised machine learning models to geospatial overdose incident data, as demonstrated in this analysis can help communities struggling with overdoses, forecast overdose trends, and develop a targeted approach to early intervention and prevention efforts corresponding to the clustering results. This analysis provides inferences based on the current state, scope, and availability of data on heroin-related EMS calls in Cincinnati. EMS data, as well as data from other first responders, and additional demographic, social, and economic covariates coming from local settings may help develop a strategy to respond to the overdose crisis. To apply this analysis to other locations will require localized covariates because the pattern of opioids and socioeconomic backgrounds may differ from each place [39], [40].

This study has several unique implications comparing to other research. First, this study used EMS data including geospatial information that can tell us the dynamic trends of heroin-overdose incidents. Therefore, this study demonstrates the usefulness of open source EMS data on how to rapidly detect changes in overdose problems in a local community. Secondly, this study used a refined common K-Means unsupervised clustering algorithm to detect a proper number of clusters given data sets. It is common knowledge that policymakers are experiencing troubles on how to define target areas in the local community regarding the opioid epidemic crisis [36]. This study could be a guide to find steps to implement monitoring and surveillance strategy to respond to opioid problems with publicly available information about opiate overdoses, combined with data on geospatial/socioeconomic risk factors, which may help municipalities plan, implement, and target harm-reduction measures. Finally, this study presents that racial/ethnic diversity could be an important factor to reduce heroin-related incidents in the City of Cincinnati. Other cities and communities similar to the City of Cincinnati could consider a geospatial social mix diversity strategy to overcome the opioid crises in the region.

This study has a few limitations. First, we studied overdose incidents classified as heroin-related at the time of dispatch. The conclusion on the scene may be different but recoding does not apply on-site. Therefore, the classifications for the dispatches could be inaccurate. Second, we used census block levels to generate sociodemographic covariates. In addition, we generated data set without temporal considerations, rather retrospective analysis. Thus, we had to calculate the mean value of the covariates in each census block level as input values. Finally, even though Cincinnati is a large city in the United States, it may not be representative of a major metropolitan area in the US that is experiencing an influx of opioid-related overdoses, nor do our results apply to rural areas which may have different demographics and risk factors. Our analysis shows that Cincinnati is located between metropolitan and micropolitan settings in terms of sociodemographic characteristics.

Despite some limitations, our geospatial analysis of the most current data on suspected overdose calls can inform community programs on trending of overdose as well as help target specific populations that are experiencing increased overdose events by including certain demographic characteristics in the analysis. We are expanding our analyses on 1) applying unsupervised clustering machine learning algorithms to the other EMS data retrieved from various cities, 2) developing spatial clustering models which can observe the characteristics of cluster and assess the relationship between the heroin-related OD incidents and healthcare accessibility, 3) developing local spatial clustering with local indices of spatial association (LISA) statistics to identify hot-zone of substance abuse in the community, and 4) implementing various supervised machine learning predictive models with several classifiers such as support vector machine (SVM), neural network, random forest and gradient boosting machine (GBM) to the larger EMS data sets.

5. Acknowledgement

The work for this article was funded by the Ohio Department of Higher Education, BOR01-QD00006943. The authors have no conflicts of interest to report.

6. Appendix

The Appalachian score is the binary variable defined by:

1. Greater than 23% of the families are below the poverty level,
2. Less than 41.0% of families are African American
3. Less than 80% of the persons 25 years or older are high school graduates
4. More than 7% of the persons 16-19 years old who are not in school are not high school graduates
5. More than 62% of the persons 16-19 years old are jobless (includes those unemployed and those not in the civilian labor force)
6. More than 3 persons per average family

If at least four criteria were met, the neighborhood was identified as having a significant Appalachian population, but not as long as the African American population was more than 41.0 (the city-wide) percentage.

7. References

[1] N. Wilson *et al*, "Drug and Opioid-Involved Overdose Deaths — the United States, 2017–2018," *MMWR Morb Mortal Wkly Rep.69(11):290-297,* vol. 69, *(11)*, 2020. Available: https://stacks.cdc.gov/view/cdc/99464.
[2] National Institutes of Health. *Overdose Death Rates.* Available: https://www.drugabuse.gov/drug-topics/trends-statistics/overdose-death-rates.
[3] National Institutes of Health. *Ohio Opioid Summary.* Available: https://www.drugabuse.gov/opioid-summaries-by-state/ohio-opioid-summary.
[4] M. Rembert, C. William Swank Program in Urban-Rural Policy and Ohio State University, *Taking Measure of Ohio's Opioid Crisis.* 2017. Available: https://aede.osu.edu/sites/aede/files/publication_files/Swank%20-%20Taking%20Measure%20of%20Ohios%20Opioid%20Crisis.pdf
[5] B. N. Cochran *et al*, "Factors predicting development of opioid use disorders among individuals who receive an initial opioid prescription: Mathematical modeling using a database of commercially-insured individuals," *Drug and Alcohol Dependence,* vol. 138, pp. 202-208, 2014. Available: https://www.clinicalkey.es/playcontent/1-s2.0-S0376871614007716.
DOI: 10.1016/j.drugalcdep.2014.02.701.
[6] US Centers for Disease Control and Prevention "CDC'S EVIDENCE-BASED STRATEGIES FOR PREVENTING OPIOID OVERDOSE," *US Official News,* 2018.
[7] W. C. Goedel *et al*, "Association of Racial/Ethnic Segregation With Treatment Capacity for Opioid Use Disorder in Counties in the United States," *JAMA Netw Open,* vol. 3, *(4),* pp. e203711, 2020. Available: https://doi.org/10.1001/jamanetworkopen.2020.3711. DOI: 10.1001/jamanetworkopen.2020.3711.
[8] R. M. DeWolf, "Predicting Opioid Treatment Admissions." Doctoral Dissertation, Utica College, United States -- New York, 2019.
[9] B. Bearnot, J. F. Pearson and J. A. Rodriguez, "Using Publicly Available Data to Understand the Opioid Overdose Epidemic: Geospatial Distribution of Discarded Needles in Boston, Massachusetts," *Am. J. Public Health,* vol. 108, *(10),* pp. 1355-1357, 2018. Available: https://doi.org/10.2105/AJPH.2018.304583. DOI: 10.2105/AJPH.2018.304583.
[10] J. Cordes, "Spatial Trends in Opioid Overdose Mortality in North Carolina: 1999–2015," *Southeastern Geographer,*

vol. 58, *(2),* pp. 193-211, 2018. Available: https://muse.jhu.edu/article/698836. DOI: 10.1353/sgo.2018.0022.
[11] M. K. Bohm *et al*, "Heroin and healthcare: patient characteristics and healthcare prior to overdose," *The American Journal of Managed Care,* vol. 25, *(7),* pp. 341, 2019. Available: https://www.ncbi.nlm.nih.gov/pubmed/31318507.
[12] T. J. Cicero, M. S. Ellis, and J. Harney, "Shifting Patterns of Prescription Opioid and Heroin Abuse in the United States," *N. Engl. J. Med.,* vol. 373, *(18),* pp. 1789-1790, 2015. DOI: 10.1056/NEJMc1505541 [doi].
[13] Z. R. Li *et al*, "Suspected heroin-related overdoses incidents in Cincinnati, Ohio: A spatiotemporal analysis," *PLOS Medicine,* vol. 16, *(11),* pp. e1002956, 2019. Available: https://doi.org/10.1371/journal.pmed.1002956.
[14] C. Bauer *et al*, "Patterns and risk factors of opioid-suspected EMS overdose in Houston metropolitan area, 2015-2019: A Bayesian spatiotemporal analysis," *Plos One,* vol. 16, *(3),* pp. e0247050, 2021. Available: https://doi.org/10.1371/journal.pone.0247050.
[15] S. Kashani, S. Sanko, and M. Eckstein, "The Critical Role of Dispatch," *Cardiol. Clin.,* vol. 36, *(3),* pp. 343-350, 2018. DOI: S0733-8651(18)30820-8 [pii].
[16] A. Garza and S. Dyer, "EMS data can help stop the opioid epidemic," *Journal of Emergency Medical Services,* vol. 41, *(11),* 2016.
[17] B. Bearnot, J. F. Pearson and J. A. Rodriguez, "Using publicly available data to understand the opioid overdose epidemic: geospatial distribution of discarded needles in Boston, Massachusetts," *Am. J. Public Health,* vol. 108, *(10),* pp. 1355-1357, 2018.
[18] D. P. Watson *et al*, "Lay responder naloxone access and Good Samaritan law compliance: postcard survey results from 20 Indiana counties," *Harm Reduction Journal,* vol. 15, *(1),* pp. 1-8, 2018.
[19] M. Tracy *et al*, "Circumstances of witnessed drug overdose in New York City: implications for intervention," *Drug Alcohol Depend.,* vol. 79, *(2),* pp. 181-190, 2005.
[20] The City of Cincinnati. *Cincinnati Fire Incidents (CAD) (including EMS: ALS/BLS)*. Available: https://data.cincinnati-oh.gov/Safety/Cincinnati-Fire-Incidents-CAD-including-EMS-ALS-BL/vnsz-a3wp.
[21] J. I. Choi *et al*, "Spatial Clustering of Heroin-Related Overdose Incidents and Accessibility of Healthcare Facilities," working paper, September, 2021.
[22] United States Census Bureau. *2017 American Community Survey (ACS)*. Available: https://data.census.gov/cedsci/.
[23] The Health Resources and Services Administration. *HRSA Data Warehouse*. Available: https://findahealthcenter.hrsa.gov/.
[24] The Substance Abuse and Mental Health Services Administration. *OTP Directory*. Available: https://dpt2.samhsa.gov/treatment/directory.aspx.
[25] M. Maloney and C. Auffrey, "The social areas of Cincinnati: The analysis of social needs," Cincinnati, Ohio. Available: http://socialareasofcincinnati.org/report5.html
[26] The City of Cincinnati. *PDI (Police Data Initiative) Crime Incidents; 2019*. Available: https://data.cincinnati-oh.gov/Safer-Streets/PDI-Police-Data-Initiative-Crime-Incidents/k59e-2pvf.
[27] C. Lopez *et al*, "An unsupervised machine learning method for discovering patient clusters based on genetic signatures," *J. Biomed. Inform.,* vol. 85, pp. 30-39, 2018.
[28] T. S. Alsulimani, "Improving Prevention of drug addiction based on social media via clustering approach," .
[29] S. Nawrin, M. R. Rahman and S. Akhter, "Exploreing k-means with internal validity indexes for data clustering in traffic management system," *International Journal of Advanced Computer Science and Applications,* vol. 8, *(3),* pp. 264-272, 2017.
[30] B. J. D. Sitompul, O. S. Sitompul and P. Sihombing, "Enhancement clustering evaluation result of davies-bouldin index with determining initial centroid of K-means algorithm," in *Journal of Physics: Conference Series,* 2019, .
[31] S. Khakabimamaghani and M. Ester, "Bayesian biclustering for patient stratification," in *Biocomputing 2016: Proceedings of the Pacific Symposium,* 2016, .
[32] R. Shen, A. B. Olshen and M. Ladanyi, "Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis," *Bioinformatics,* vol. 25, *(22),* pp. 2906-2912, 2009.
[33] A. Sakellariou, D. Sanoudou and G. Spyrou, "Combining multiple hypothesis testing and affinity propagation clustering leads to accurate, robust and sample size independent classification on gene expression data," *BMC Bioinformatics,* vol. 13, *(1),* pp. 1-19, 2012.
[34] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.,* vol. 20, pp. 53-65, 1987.
[35] H. Clifford *et al*, "Comparison of clustering methods for investigation of genome-wide methylation array data," *Frontiers in Genetics,* vol. 2, pp. 88, 2011.
[36] V. A. Pear *et al*, "Urban-rural variation in the socioeconomic determinants of opioid overdose," *Drug Alcohol Depend.,* vol. 195, pp. 66-73, 2019. Available: https://www.sciencedirect.com/science/article/pii/S0376871618308408. DOI: https://doi.org/10.1016/j.drugalcdep.2018.11.024.
[37] C. D. Copemann and P. L. Shaw, "Selected Demographic Characteristics and Patterns of Drug Abuse among Treated Addicts from a Suburban Community," *British Journal of Addiction to Alcohol & Other Drugs,* vol. 70, *(2),* pp. 205-212, 1975. Available: https://doi.org/10.1111/j.1360-0443.1975.tb00028.x. DOI: https://doi.org/10.1111/j.1360-0443.1975.tb00028.x.
[38] K. Lippold and B. Ali, "Racial/ethnic differences in opioid-involved overdose deaths across metropolitan and non-metropolitan areas in the United States, 1999−2017," *Drug Alcohol Depend.,* vol. 212, pp. 108059, 2020. Available: https://www.sciencedirect.com/science/article/pii/S0376871620302246. DOI: https://doi.org/10.1016/j.drugalcdep.2020.108059.
[39] G. Helmerhorst *et al*, "An epidemic of the use, misuse and overdose of opioids and deaths due to overdose, in the United States and Canada: is Europe next?" *The Bone & Joint Journal,* vol. 99, *(7),* pp. 856-864, 2017.

[40] V. A. Pear *et al*, "Urban-rural variation in the socioeconomic determinants of opioid overdose," *Drug Alcohol Depend.*, vol. 195, pp. 66-73, 2019.