ABSTRACT

Title of Dissertation:	TRACKING THE DYNAMICS OF THE OPIOID CRISIS IN THE UNITED STATES OVER SPACE AND TIME
	Zhiyue Xia, Doctor of Philosophy, 2022
Dissertation directed by:	Professor Kathleen Stewart Department of Geographical Sciences

Millions of adolescents and adults in the United States suffer from drug problems such as substance use disorder, referring to clinical impairments including mental illnesses and disabilities caused by drugs. The Substance Abuse and Mental Health Services Administration reported the estimated number of illicit drug users increased to 59.3 million in 2020, or 21.4% of the U.S. population, which made drug misuse one of the most concerning public health issues. Opioids are a category of drugs that can be highly addictive, including heroin and synthetic drugs such as fentanyl. Centers for Disease Control and Prevention (CDC) indicated that about 74.8% of drug overdose deaths involved opioids in 2020. The opioid crisis has hit American cities hard, spreading across the U.S. beginning with the west coast, and then expanding to heavily impact the central, mid-Atlantic, and east coast of the U.S. as well as states in the southeast.

In this dissertation, I work on three studies to track the dynamics of the opioid crisis in the U.S. over space and time from a geographic perspective using spatiotemporal data science methods including clustering analysis, time-series models and machine learning approaches. The first study

focused on the geospatial patterns of illicit drug-related activities (e.g., possession, delivery, and manufacture of opioids) in a typical U.S. city (Chicago as a case study area). By analyzing more than 52,000 reported drug activities, I built a data-driven machine learning model for predicting opioid hot zones and identifying correlated built environment and sociodemographic factors that drove the opioid crisis in an urban setting. The second study of my dissertation is to analyze the opioid crisis in the context of the global pandemic of SARS-CoV-2 (COVID-19). In 2020, COVID-19 outbroke and affected hundreds of millions of people across the globe. The COVID-19 pandemic is also impacting the community of opioid misusers in the U.S. The major research objective of Study 2 is to understand how the opioid crisis is impacted by the COVID-19 pandemic and to find neighborhood characteristics and economic factors that have driven the variations before and during the pandemic. Study 3 focuses on analyzing the crisis risen by synthetic opioids (including fentanyl) that are more potent and dangerous than other drugs. This study analyzed the geographic patterns of synthetic opioids spreading across the U.S. between 2013 and 2020, a period when synthetic opioids rose to be a major risk factor for public health.

The significance of this dissertation is that the three studies investigate the opioid crisis in the U.S. in a comprehensive manner and these studies can facilitate public health stakeholders with effective decision making on healthcare planning relating to drug problems. Tracking the dynamics of the opioid crisis by drug type, including modeling and predicting the geographic patterns of opioid misuse involving particular opioids (e.g, heroin and synthetic opioids), can provide an important basis for applying further treatment services and mitigation efforts, and also be useful for assessing current services and efforts.

TRACKING THE DYNAMICS OF THE OPIOID CRISIS IN THE UNITED STATES OVER SPACE AND TIME

by

Zhiyue Xia

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2022

Advisory Committee:
Professor Kathleen Stewart, Chair
Professor Giovanni Baiocchi
Professor Taylor Oshan
Professor Eric Wish
Professor Vanessa Frias-Martinez

Dedication

To my parents,

Ping Zhang and Ming Xia,

my husband,

Mofeng Yang,

and my grandparents for their endless love, support, and encouragement throughout the journey.

Acknowledgments

I want to express my profound gratitude to my advisor, Professor Kathleen Stewart, for her advice, guidance, and help all the way during my Ph.D. Prof. Stewart has provided invaluable feedback on my research and enduring support in both my study and life during these four years. My Ph.D. journey is impossible to be accomplished without Prof. Stewart's consistent support.

I would like to sincerely thank my dissertation committee members, Prof. Giovanni Baiocchi, Prof. Taylor Oshan, Prof. Eric Wish and Prof. Vanessa Frias-Martinez for their willingness to generously provide their insight and expertise to my dissertation.

I am also grateful to my colleagues, enthusiastic CGISers, Dr. Junchuan Fan, Guimin Zhu, Dr. Yao Li, Prof. Yanjia Cao and Prof. Hai Lan for discussing research, helping solve technical issues and assisting in data collection. Thank you to my friends and cohort members, those Drs. and future Drs., Kaihui Song, Ipsita Kumar, Aolin Jia, Yunting Song, Jiaming Lu, Zheng Liu, Diyang Cui, Lei Ma, Allison Bredder, Rebecca Traldi, Meghavi Prashnani, Jiaying He, Xin Xu, Jeffery Sauer, Jamis Bruening, Noel Dyer, Xinyuan Li, Yuhao Wang, Peiqi Zhang, Jiena He, Ruohan Li, Yiming Zhang, Yingrui Zhao, Mengyu Liao, Qianru Liao, Qinglian He, Xiaoshan Yin, and Zhangchi Xu for making these four years jolly and memorable.

I also want to extend thanks to my animal babies, Candy and Coco, who accompany me to overcome the tough time during the pandemic.

This dissertation is funded by an Ann G. Wylie Dissertation Fellowship and the Department of Geographical Sciences' GIS Summer Research Fellowship from the University of Maryland. This dissertation is partly involved in a funded project, National Drug Early Warning System (NDEWS) Coordinating Center at the University of Maryland's Center for Substance Abuse Research (CESAR).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CDC: Centers for Disease Control and Prevention

COVID-19: Coronavirus disease caused by the SARS-CoV-2 virus

SAMHSA: Substance Abuse and Mental Health Services Administration

DEA: Drug Enforcement Administration

NDEWS: National Drug Early Warning System

CDC WONDER: Multiple Cause of Death Data provided by Centers for Disease Control Wide-ranging Online Data for Epidemiologic Research

CLEAR: Chicago Police Department's Citizen Law Enforcement Analysis and Reporting system

STRF: Space-time random forest model

ACC: Predictive accuracy

PAI: Prediction accuracy index

PEI: Prediction efficiency index

GRRF: Guided regularized random forest

AIR: Actual impurity reduction

OOB: Out-of-bag

PDP: Partial dependence plots

STRF_COVID: space-time random forest model accounting for COVID-19 factors

GCN: Graph convolutional networks

Chapter 1: Introduction

1.1 Background and motivation

The opioid crisis spread across the United States beginning from the west coast in the early 2000s, and expanding to heavily impact the central, mid-Atlantic, and east coast of the U.S. as well as states in the southeast by 2014 (Ciccarone, 2017; Dasgupta et al., 2017; Stewart et al., 2017). The estimated number of adolescent and adult illicit drug users increased from 27 million in 2014 to 59.3 million in 2020 (Center for Behavioral Health Statistics and Quality, 2015; Substance Abuse and Mental Health Services Administration, 2021). Opioids are a category of drugs that include heroin as well as synthetic drugs such as fentanyl (National Institute on Drug Abuse, 2019a). The crisis was reflected by increasing numbers of overdoses involving heroin (Hedegaard et al., 2018; Jalal et al., 2018; Stewart et al., 2017). The number of heroin users reached a peak in 2016 with an estimated 948,000 users or 0.4 percent of the U.S. population at that time (Substance Abuse and Mental Health Services Administration, 2019). Fentanyl, a potent synthetic opioid, has also become a threat nationally (Centers for Disease Control and Prevention, 2019a; Marshall et al., 2017; Spencer et al., 2019) as synthetic opioids have been increasingly identified as a leading factor in overdose deaths in at least 23 states since 2015 (Centers for Disease Control and Prevention, 2019b; National Institute on Drug Abuse, 2019b).

This dissertation aims at analyzing the dynamics of the opioid crisis in the U.S. over space and time. To achieve this research goal, three different studies have been undertaken, using various advanced spatiotemporal data science methods including hotspot analysis for identifying drug hot zones, space-time modeling for understanding the changing patterns of drug overdose deaths, and time series analysis for identifying the relationship between the spread of two types of drugs (i.e., heroin and synthetic opioids). Space-time modeling of changing patterns of certain repeated events (e.g., homicides, drug use, and gun violations) that occurred in urban and suburban areas has been studied by researchers (Hodgkinson and Andresen, 2019; Mohler, 2014; Piza and Carter, 2018; Shiode et al., 2015; Zhao and Tang, 2017). Existing geographical research on spatiotemporal patterns of repeated events has included forecasting hotspots based on historical data, and finding correlations between the spatial patterns of the events and surrounding environmental characteristics, among other topics. Recent studies, for example, have used network-based models to detect and forecast street-level crime hotspots by using historical crime data (Shiode and Shiode, 2020; Y. Zhang and Cheng, 2020). While other research studies investigated correlations between spatial patterns and environmental variables extracted from multiple sources of data such as social media data (Vomfell et al., 2018), remote sensing imagery (Yang et al., 2020) and Google Street View(He et al., 2017). Machine learning methods have also been used extensively for predicting hot zones of repeated events, although few studies have combined spatiotemporal patterns of historical data together with multiple underlying driving factors in machine learning models in order to enable the model to capture the changing patterns over space and time during the learning process.

The first study in this dissertation develops a spatiotemporal analysis workflow that uses a machine learning approach to analyze the geospatial patterns of drug activities in a U.S. metropolitan area at the spatial granularity of census block groups. In this study, drug-related crime data was used as a proxy for locations where drug activities are likely to occur. As reported previously, drug-related problems are an important public health issue in the U.S. Research on the spatial and temporal patterns of drug activities and methods that support detecting clusters of particular drugs are useful for distinguishing hot zones of those drugs that in turn provide a basis for assessing, for example, the availability of substance use treatment facilities in these locations (Abraham et al., 2018; Cao et al., 2019; Yarbrough et al., 2019). Geospatial analysis has been extensively used in examining the spatial patterns of drug activities. Previous research on geospatial analysis relating to substance use and the opioid crisis topics has investigated topics, including, for example, the dynamics of the drug overdose epidemic involving heroin, synthetic opioid, methamphetamine, methadone and cocaine in the U.S. between 1979 and 2016 (Jalal et al., 2018), and the spatial pattern of deaths involving heroin across the U.S. from 2000 through 2014 (Stewart et al., 2017). Researchers from the University of New South Wales examined the geographic patterns of injecting locations in Sydney, Australia for drug users and reported that nearly all drug users (96%) in their study injected drugs in public spaces including streets, vehicles and public toilets (Darke et al., 2001). A study on analgesic overdose fatalities in New York City found an association between drug patterns and built environment and neighborhood socioeconomic factors (Cerdá et al., 2013). Another study investigated the spatial patterns of adolescent drug use in Cincinnati, Ohio, where researchers analyzed student drug use survey data and found significant clustering patterns of high risk for drug use in downtown Cincinnati, and Butler and Clermont Counties (Chaney and Rojas-Guyler, 2015). A previous study applied spatiotemporal analysis of social media data to investigate drug-related mental health problems at ZIP code level (Cao et al., 2020). Drug-related crime data that records crime incidents including drug transactions, and delivery and possession charges could provide location information on where drugs are likely to be present.

In the first study, I used a data-driven approach to analyze locations of drug-related crimes and the underlying driving factors for these drug activities over time. This was accomplished by incorporating space and time into a random forest machine learning model. To gain insights into the relationship between the frequent locations of drug use and drug-related crime in Chicago over time, I also compared the spatial patterns of deaths involving opioids with crime incidents involving drugs between 2016 and 2019. For this study, I used narcotic crime data as a proxy for locations where drugs were likely to be present. I analyzed more than 52,000 reported incidents of drug-related crime at block group level in Chicago, IL between 2016 and 2019. Chicago is a major metropolitan city in the U.S. with a population of over 2.7 million in 2018. I designed a space-time random forest model that accounted for spatiotemporal autocorrelation in the patterns of drug-related crimes using three different metrics in order to identify a set of possible underlying drivers for these patterns and track what changes over time have occurred.

The second study in this dissertation was motivated by the sharp increase in the number of opioid-involved deaths in the United States during the global pandemic of SARS-CoV-2 (COVID-19). COVID-19 quickly became the most concerning public health issue in 2020, affecting hundreds of millions of people across the globe (World Health Organization, 2020). Evidence has shown that during the COVID-19 pandemic, public health issues including mental health and opioid use disorder were especially exacerbated by COVID-19 due to stress arising from social isolation, job losses, worry about sickness, and lack of access to healthcare among other reasons (Chen et al., 2021; Holland et al., 2021; Hossain et al., 2020; National Institute of Environmental Health Sciences, 2020; Sato et al., 2022; Sun et al., 2020). Opioid treatment service deliveries were frequently disrupted during the time of the pandemic (Becker and Fiellin, 2020; Hossain et al., 2020) and recent studies have shown that over 30 states in the U.S. reported increased opioid-involved fatalities during the pandemic (American Medical Association, 2020; National Center for Health Statistics, 2021; National Institute on Drug, 2022)

This study focuses on analyzing opioid-involved mortality in the context of the COVID-19 pandemic in the U.S. from a geospatial perspective. The number of opioid-involved deaths that occurred in the U.S. increased by 37.6% in 2020, likely linked to two factors, the increasing trend in opioid-involved deaths generally in the U.S., and the pandemic. In this study, counterfactual analysis is combined with a spatiotemporal machine learning model to estimate the number of opioid-involved deaths in the U.S for 2020 under the assumption that the COVID-19 pandemic did not happen. The counterfactual analysis also estimates the possible additional increase in opioid-involved deaths in 2020 that were likely related in some way to the pandemic. Counterfactual analysis refers to comparing what actually has happened with what would happen if no intervention was involved (Collins et al., 2004; White, 2006) and is used here for analyzing the increased in opioid-involved deaths in 2020 can be attributed to the COVID-19 pandemic.

In the third study, I focused more on the synthetic opioid crisis and investigated the rapid spread of drug overdose deaths involving synthetic opioids across the U.S. from 2013 to 2020. Synthetic opioids are substances that are synthesized in a laboratory and the original purpose of pharmaceutical synthetic opioids was for pain management (Drug Enforcement Administration (DEA), 2020). Fentanyl is a type of synthetic opioid that is much more potent than other opioids (e.g., approximately 50-100 times stronger than morphine) ((U.S. Drug Enforcement Administration, 2022). The fentanyl-related substances in the U.S. are primarily trafficked from Mexico and China, and India is also a source country for manufacturing fentanyl (DEA Intelligence Report, 2020). Previous research has reported that the increase in synthetic opioid-involved deaths in the U.S. has been proportionately related to the amount of illicit fentanyl obtained by law enforcement, but not related to the prescribing rates for pharmaceutical fentanyl (Gladden et al., 2016; O'Donnell et al., 2017; Peterson et al., 2016). These findings indicated that the U.S. synthetic opioid epidemic has been driven by the increase in the number of deaths involving illicitly manufactured fentanyl and fentanyl-related substances (Centers for Disease Control and Prevention, 2022a).

The third study investigates the spread of synthetic opioid deaths across the U.S. since 2013. This study analyzed spatiotemporal patterns of synthetic opioid-involved deaths (including fentanyl-related deaths), and the resulting patterns were compared to patterns of heroin-involved deaths between 2013 and 2020. Time-lagged correlation analysis was implemented to identify the relationships between synthetic opioid-involved deaths and heroin-involved deaths. In this study, in order to model the spread of synthetic opioid-involved deaths and incorporate spatial relationships between counties into the model, graph-structured data on synthetic opioid-involved deaths were generated based on both spatial adjacency and human mobility.

1.2 Dissertation outline and research objectives

This dissertation seeks to enhance understanding of the dynamics of the opioid crisis in the U.S. from a geographic perspective. The overall objective of this dissertation is to track the changing patterns of opioid-involved overdose deaths over space and time using advanced spatial data science methods.

This dissertation is composed of five chapters. Chapter 1 describes the background of the opioid crisis in the U.S. and the motivations for three studies that comprise this dissertation. Chapters 2, 3, and 4 correspond to three different studies: 1) analyzing geospatial patterns of drug-related crime incidents in a major U.S. metropolitan area over space and time; 2) spatiotemporal analysis of opioid-involved mortality in the context of the COVID-19 pandemic in the U.S. and 3) tracking the spatial dynamics of the synthetic opioid epidemic in the U.S.

Research objectives for Chapters 2, 3, and 4

The research objective for Chapter 2 is to analyze geospatial patterns of drug-related crime incidents in a major U.S. metropolitan area using a random forest model that incorporates space and time into the model. Three research objectives are addressed by this study:

1) Analyze and expose the spatiotemporal patterns for heroin and synthetic drug-related crimes in a major U.S. metropolitan area, the City of Chicago, IL, over the period 2016 to 2019, and investigate the space-time trends of the patterns for both types of drug-related crimes to identify the locations where particular drugs were likely to be present in the city.
2) Identify how different factors including built environment and sociodemographic factors, were associated with the heroin and synthetic drug activity patterns in Chicago using a data-driven approach and random forest model.
3) Extend a random forest model to incorporate both spatial and temporal aspects by introducing three new lag variables to capture changing trends (e.g., emergence and disappearance) in drug activity patterns at census block group level.

Chapter 3 analyzes the increase in opioid-related deaths during the COVID-19 pandemic in 2020, and compares the spatiotemporal trend in opioid-involved deaths for 2020 with the trends in the numbers of deaths during the previous three years (2017-

2019). This study implements a counterfactual analysis of the numbers of deaths for the entire nation at county level as well as for the City of Chicago at ZIP code level. The research objectives for this study are:

1) Investigate the spatiotemporal patterns of opioid-involved deaths for all U.S. counties during January-December 2020 and compare these patterns with opioid-involved deaths for the previous three years (2017-2019).
2) Use the counterfactual analysis to estimate opioid deaths for 2020 under the assumption that the COVID-19 pandemic did not happen and compare the estimated deaths with the actual reported number of deaths.
3) Analyze the increase in opioid-involved deaths in 2020 in the context of associated factors including demographic, economic factors and COVID-19 pandemic-related factors using a space-time random forest model.
4) Generate a ZIP-code level vulnerability index of opioid overdose occurrence in the context of the pandemic for a U.S. metropolitan area, the City of Chicago.

Chapter 4 focuses on tracking the spatial dynamics of the synthetic opioid crisis in the U.S., between 2013 and 2020 and compares the pattern of synthetic opioid-involved deaths during this period with the patterns of deaths due to heroin-involved overdoses. The research objectives addressed in this study include:

1) Investigate the relationship between the spread of synthetic opioids and other key opioids (e.g., heroin) using time-lagged cross correlation analysis

and compare the spatiotemporal patterns of synthetic opioid-involved deaths and heroin-involved deaths between 2013 and 2020.

2) Analyze the role of possible contributing factors (e.g., human mobility among states and counties, and spatial adjacency of counties) to the spatial spread in the number of deaths associated with the synthetic opioid crisis (deaths involving fentanyl) using a graph convolutional neural network.

Chapter 5 concludes the dissertation by emphasizing the major findings and summarizing the contributions of the three studies. Future research directions relating to the spatial dynamics of the opioid crisis and spatial data science methods applied in the area of health geography are also included in this chapter.

Chapter 2: Analyzing geospatial patterns of drug-related crime incidents in a major United States metropolitan area over space and time

2.1 Abstract

The opioid crisis has hit American cities hard, and research on spatial and temporal patterns of drug-related activities including detecting and predicting clusters of crime incidents involving particular types of drugs is useful for distinguishing hot zones where drugs are present that in turn can further provide a basis for assessing and providing related treatment services. In this study, we investigated spatiotemporal patterns of more than 52,000 reported incidents of drug-related crime at block group granularity in Chicago, IL between 2016 and 2019. We applied a space-time analysis framework and machine learning approaches to build a model using training data that identified whether certain locations and built environment and sociodemographic factors were correlated with drug-related crime incident patterns, and establish the top contributing factors that underlaid the trends. Space and time, together with multiple driving factors, were incorporated into a random forest model to analyze these changing patterns. We accommodated both spatial and temporal autocorrelation in the model learning process to assist with capturing the changes over time and tested the capabilities of the space-time random forest model by predicting drug-related activity hot zones. We focused particularly on crime incidents that involved heroin and

synthetic drugs as these have been key drug types that have highly impacted cities during the opioid crisis in the U.S.

2.2 Introduction

Research on the spatial and temporal patterns of drug activities and detecting clusters of particular drugs is useful for distinguishing hot zones of those drugs that in turn provides a basis for assessing the availability of substance use treatment facilities in these locations (Abraham et al., 2018; Yarbrough et al., 2019). For this study, the research objective was to identify factors related to built environment and sociodemographic characteristics of communities that were related to the changing patterns of drug activities involving particular drugs in a city, and to build a space-time random forest model, that could shed light on the importance of the different factors, and model these changing patterns over space and time so that mitigation steps might be taken. We investigated the spatiotemporal patterns of drug-related crimes using narcotic crime data as a proxy for locations where drugs were likely to present. We analyzed the spatial patterns of more than 52,000 reported incidents of drug-related crime, including over 16,000 heroin and synthetic drug-related crimes at block group granularity in Chicago, IL between 2016 and 2019. Chicago is a major metropolitan city in the U.S. with a population of over 2.7 million in 2018. We designed a space-time random forest model that accounted for spatiotemporal autocorrelation in the patterns of drug-related crimes in order to identify a set of possible underlying drivers for these patterns and track what changes have occurred. These drivers included specific types of locations (e.g., vacant buildings and alleys), additional built environment factors (e.g., road network density and street intersection density), and sociodemographic factors (e.g., level of education, income, and percentage of owner occupancy in a neighborhood). Here, built environment refers to the properties of human-modified places such as streets, buildings, parks and transportation systems, and characterizes city structure (the United States Environmental Protection Agency, 2019). Public health problems including mental health and substance use disorder have been shown to be related to built environment factors (Cerdá et al., 2013a; Srinivasan et al., 2003). Sociodemographic variables such as education, income and household status have also been found to be correlated with drug activities (Lipton et al., 2013; Nechuta et al., 2018; Vilalta, 2010). The space-time random forest model was used to identify built environment and sociodemographic factors that were most associated with the drug activity locations. Spatial and temporal autocorrelation were accommodated in the model learning process to capture the changing trend of drug activity patterns over successive time periods.

2.3 Related works

Crime dataset records including the location and date-time of drug crime incidents for delivery, possession, and manufacture of drugs, have been intensively used for investigating geographic patterns of drug activities. In a previous study, researchers analyzed crime data as well as recorded calls for police service to examine the relationships between drug activities, social disorder and crime, with results that showed significant spatial links between them (Weisburd and Mazerolle, 2000). Another study investigated the geographical relations between alcohol outlets, drug markets and violence using arrest data on drug possession and trafficking to map estimated drug markets in Boston (Lipton et al., 2013). In Mexico, an analysis of the spatial dynamics of drug arrests related to marijuana and cocaine found sociodemographic factors such as college education, housing conditions, and female-headed households were positively correlated with arrests involving marijuana, but no sociodemographic correlates were significantly established for cocaine patterns (Vilalta, 2010).

Hotspot analysis has assisted in identifying where events are densely concentrated, and has been used as a basis for predicting spatial patterns of repeated events in urban environments (Albright et al., 2019; Chainey et al., 2008). For example, a previous study used hotspot analysis to examine the homicide patterns in Chicago from 1960 to 1995 (Ye and Wu, 2011). Prior studies investigated spatial patterns of drug activities and understand any associations with surrounding built environment characteristics (Cerdá et al., 2013a; Chaney and Rojas-Guyler, 2015b; Darke et al., 2001). Drug-related crime data that records crime incidents including drug transactions, delivery and possession can be analyzed to provide insights on locations where drug use may be present. In this paper, we have incorporated space and time into a random forest model to analyze drug-related crime incidents and their underlying factors over time. An earlier study described how the Chicago Police Department used a ‘heat list’ that included approximately 400 individuals who were forecast to be potentially involved in crime (Lum and Isaac, 2016). They pointed out that police-recorded data was biased and could be attributed to the police's determination of where to patrol and search. To examine drug-related crime patterns, and to gain further insights about locations of drug-related activities, we compared the spatial patterns of deaths involving opioids and drug-related crimes between 2016 and 2019, and based on this analysis, investigated the relationships between drug activities in Chicago, a major metropolitan city in the U.S.

For this study, we designed a space-time random forest model that ingested data from multiple geospatial data sources to investigate drug activity patterns and their associated spatiotemporal characteristics in an urban setting. We accommodated both spatial and temporal autocorrelation in the model learning process to assist with capturing the changes over time and tested the model's ability to capture trending changes by predicting future drug incident locations. Random forest, a machine learning model, combines a large number of decision trees, using selected features to split tree nodes and repeated sampling of different subsets of observations to train the models (Breiman, 2001). The output for random forest regression is the mean value of all regression trees, while the majority decision of all classification trees is the output for random forest classification tasks. Random forest model, with the capability of processing massive datasets and handling multicollinear relationships within multi-source datasets, has been used for drug-related public health research (Ancuceanu et al., 2019a; Fernández-Delgado et al., 2014a; Kamel Boulos et al., 2019a). For example, a random forest model was used to detect individuals with substance use disorder based on a set of behavior and health characteristics (Jing et al., 2020). In previous research, random forest models were often implemented without using spatial and temporal relations between model variables. Recent studies have begun to address spatial dependencies into random forest models. For example, a recent study proposed a geographical random forest model, attempting to include spatial heterogeneity in a random forest model by disaggregating a global model into several local sub-models (Georganos et al., 2019) to understand local variations.

2.4 Data

2.4.1 Drug-related crime in Chicago

In this study, a public safety dataset provided by the Chicago Data Portal[1] was used for accessing drug-related crime data in Chicago between 2016 and 2019. Chicago Data Portal extracted the public safety dataset from the Chicago Police Department's Citizen Law Enforcement Analysis and Reporting (CLEAR) system. For the reported incidents, the police recorded the type of crime such as possession, delivery and manufacture of drugs and also described the drug types. In order to discover the spatial patterns of heroin and synthetic drug-related crimes, the incidents were categorized by drug types. With the increased presence of fentanyl in U.S. cities since 2016, we

[1] Chicago Data Portal: https://data.cityofchicago.org/

assumed that the category of synthetic drugs would likely include incidents involving fentanyl (Hedegaard et al., 2019).

To capture community-level characteristics in the model, we aggregated all the variables including the frequency of drug-related crime incidents, sociodemographic and built environment factors at U.S. Census block group level. Block group is the smallest geographic unit used by the U.S. Bureau of Census to publish sample data from the American Community Survey (ACS), and we selected this as the unit of analysis for this study. To model communities by block group, 5-year ACS estimates were used to help smooth any uncertainty created by using this fine level of spatial granularity. Using block groups helped us to account for varying community characteristics and reveal patterns at this scale. We defined the drug-related crime rate as the frequency of drug crimes during a time period aggregated by block group units and adjusted by block group population. Chicago has 2210 block groups. Typically, a block group in Chicago has a population of 800-10,000 individuals. It should be noted that there were eight block groups with no associated population counts including, for example, O'Hare International Airport and Chicago Midway International Airport. As our analysis examined built environment and sociodemographic characteristics of locations where populations were living and working, these eight block groups were not included in the model dataset. A complete list of factors we have collected in this study as well as literature that have previously established a link between drug activities and these factors are presented in Table A.1 in Appendix A.

Between 2016 and 2019, drug-related crimes occurred across most of downtown Chicago (Figure 2.1). For this period, 52,567 drug-related crime incidents occurred in 1,901 block groups out of 2,210 block groups in total. The total number of drug-related crime incidents (across all categories) decreased by 12% in 2017, and then increased by 21% in 2019 (Table 2.1). The statistics showed that the number of heroin-related incidents was quite consistent from 2016 to 2017 and then increased by 18% over 2018 and 2019. For the same period, incidents involving synthetic drugs consistently increased, with a noticeably high growth rate, 94%, from 2016 to 2019.

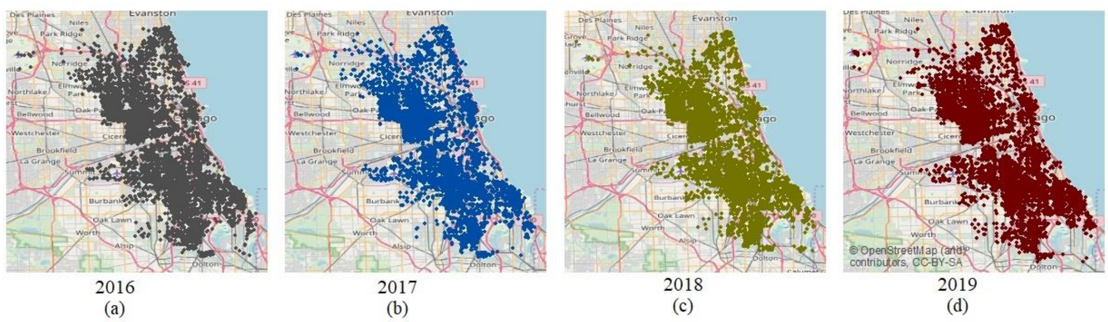

Figure 2.1 Drug-related crimes for all categories of drugs in Chicago for 2016, 2017, 2018 and 2019.

Table 2.1 Summary of drug-related crimes in Chicago for 2016, 2017, 2018 and 2019

Year	2016	2017	2018	2019
Number of incidents				
Total drug-related crimes	13318	11677	13495	14077
Heroin-related incidents	3500	3449	3807	4065
Synthetic drug-related incidents	269	309	396	521

2.4.2 Deaths involving opioids in Chicago

To determine the degree to which the spatial patterns of drug-related crime incidents were associated with locations where drug activities frequently occurred, and illustrate that drug-related crimes were not necessarily being driven by other crime-related aspects (e.g., policing strategies), we analyzed the spatial patterns of opioid-involved mortality and compared these patterns to the reported locations of drug-related crimes. Data on opioid-involved mortality between 2016 and 2019 for Chicago were extracted from the Medical Examiner Case Archive dataset accessed from Cook County Open Data[2]. The Cook County, IL Medical Examiner's Office recorded the time and location of deaths, and determined causes and manners of death cases under its jurisdiction. Deaths involving all categories of opioids in Chicago consistently increased between 2016 and 2019 (Table 2.2). The number of heroin-involved deaths reached a peak in 2017. During the four-year period, deaths involving fentanyl increased in the Chicago area by 69%. A noticeable increase in synthetic drug-related crime incidents was also noted for this same period.

Table 2.2 Deaths involving opioid in Chicago for 2016, 2017, 2018 and 2019

Year Number of deaths	2016	2017	2018	2019
Total deaths involving opioids	743	794	801	879
Deaths involving heroin	418	542	489	473
Deaths involving fentanyl	405	465	624	685

[2] Cook County Open Data: https://datacatalog.cookcountyil.gov/

2.4.3 Built environment and sociodemographic data

We collected built environment data for the study region from the United States Environmental Protection Agency (EPA) Smart Location Database released in 2013[3]. This dataset summarized indicators associated with urban design, location efficiency, transit and demographics, including, for example, street intersections per square mile, road network density and walkability scores. The complete list of variables in the EPA Smart Location Dataset are accessible in their user guide[4]. We also collected supplementary sociodemographic variables from U.S. Census at the granularity of block group including educational factors e.g., percentage of population with college-level education or higher, economic factors e.g., percentage of occupied houses, and social factors, e.g., the percentage of full-time employment. The American Community Survey (ACS) 5-year estimates dataset at block group for 2017 was used for accessing and calculating these sociodemographic variables. In total, 104 candidate factors were collected, including 61 built environment factors and 43 sociodemographic factors (Table A.1 in Appendix A).

2.4.4 Spatial data

The reported crimes dataset for Chicago included descriptions of locations where each incident occurred. The frequencies of all locations for drug-related crime reports

---

[3] Smart Location Mapping: https://www.epa.gov/smartgrowth/smart-location-mapping

[4] EPA Smart Location user guide: https://www.epa.gov/smartgrowth/smart-location-mapping

between 2016 and 2019 were analyzed and the most frequent locations for reported heroin and synthetic drug-involved crimes were computed. Based on this analysis, seven geospatial locations were identified as *key locations*. These included gas stations, vacant lots, abandoned buildings, parking lots, alleys, parks, and high schools. Data layers for sites relevant to drug-related crimes including data on vacant lots and abandoned buildings, and a boundary map of park property were collected from the Chicago Data Portal. Locations of gas stations were obtained from the ESRI Business Analyst[5] dataset. Street alley networks were sourced from Chicago WBEZ[6]; public high school locations were obtained from the Chicago Data Portal, and private high school locations from the Cook County Open Data[7]. We calculated spatial variables from these layers and aggregated the values by U.S. census block group.

2.5 Geospatial patterns of drug-related crime incidents and opioid-involved deaths

Clustering analysis was implemented to determine the spatial patterns of all drug-related crimes, crimes by individual drug type (i.e., heroin and synthetic drugs), and for opioid-involved deaths. Global Moran's I statistic was used to measure the spatial autocorrelations for both the drug-related crimes and the opioid-involved deaths (Li et al., 2007; Moran, 1950).

[5] ESRI Business Analyst https://www.esri.com/en-us/arcgis/products/arcgis-business-analyst/
[6] Chicago WBEZ https://www.wbez.org/
[7] Cook County Open Data https://datacatalog.cookcountyil.gov/

The spatial patterns of both the reported drug-related crime incidents and opioid-involved-deaths were significantly clustered, as illustrated by positive Global Moran's I values of 0.424 and 0.309, respectively, and p-values smaller than 0.001. We examined the correlation between the drug-related crime rates and the number of opioid-involved deaths (adjusted by block group population) (Figure 2.2). Figure 2.2 shows a map of frequency of drug-related crimes and the number of opioid-involved deaths by ZIP code area for the City of Chicago. The Pearson's correlation coefficient between these two variables was 0.61 with the p-value smaller than 0.001 for a two-tailed test. Based on these results, during this period of study, the spatial distribution of drug-related crime incidents and opioid-involved deaths showed a significant positive relation suggesting that the locations of drug-related crimes were indicative of where drug use was also happening in the city.

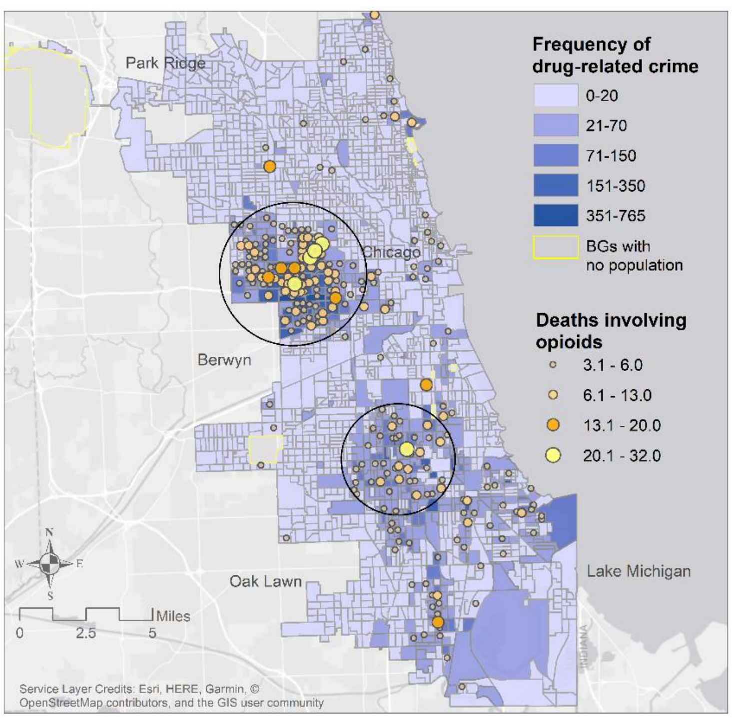

Figure 2.2 Frequency of drug-related crime incidents adjusted by population (count/population*1000), and number of deaths involving opioids adjusted by population (count/population*1000) by block group between 2016 and 2019.

Anselin's Local Moran's I statistic was developed by Anselin (1995) and is commonly used to implement clustering analysis (Liu and Wang, 2017; Neutens et al., 2013). We used Anselin's Local Moran's I statistic to identify changing patterns of heroin and synthetic drug-related crime incidents over time Block groups that were in a significant high-high value cluster were then identified as hotspot block groups. In 2016 and 2017, two major clusters were found for reported crimes involving heroin (Figure 2.3a). Hotspot 1 (upper hotspot) was consistent throughout 2016 and 2019 (Figure 2.3a). Heroin-related crime incidents that occurred in hotspot 1 increased by 22% during the four-year period. By contrast, hotspot 2 (lower hotspot), gradually diminished during the four years and in 2019, impacted only two block groups (Figure 2.3a). During the four-year period, the heroin hotspots diminished over time and the synthetic drug hotspots took over in those block groups. The clusters of synthetic drug-related crimes changed from a scattered pattern in 2016 to being concentrated in two distinct hotspots by the end of 2019 (Figure 2.3b). These two distinct synthetic drug hotspots were in similar locations to the heroin hotspots of 2016 and 2017, but included a higher number of block groups.

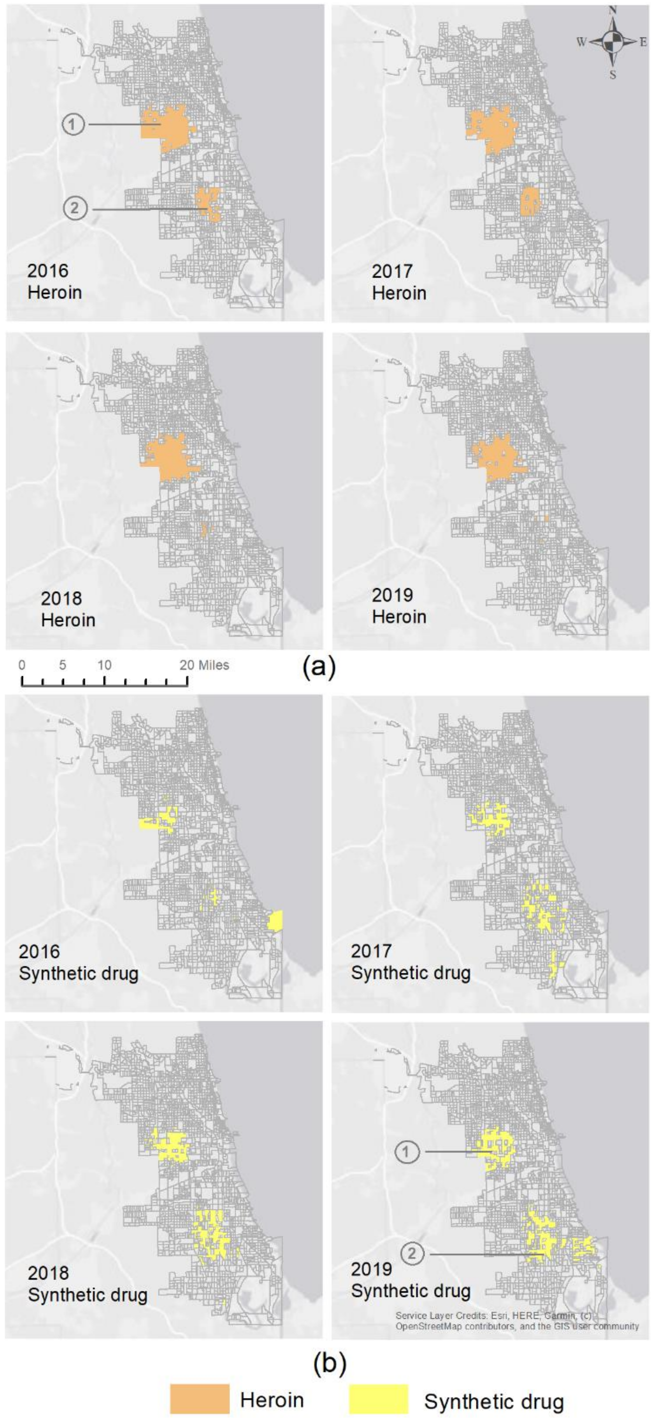

Figure 2.3 Spatial clustering of drug-related crime incidents involving (a) heroin and (b) synthetic drugs in Chicago for 2016, 2017, 2018 and 2019.

2.6 Building a space-time random forest model

2.6.1 Incorporating space and time into a random forest model

To build a random forest model that was capable of capturing the changing patterns of heroin and synthetic-drug related crime incidents over time, spatial and temporal lag variables were added to the model in order to detect the spatial dependencies on neighboring block groups, and the relationships between successive time periods.

Temporal patterns of the reported crimes were used to guide the selection of time periods over which to detect spatial change. To select the temporal granularity for analyzing heroin-involved crime incidents, we aggregated monthly incidents by census block group and used clustering to determine the hotspot block groups. A time series of heroin-related incidents was created based on the monthly count of hotspot block groups. To smooth out short-term fluctuations and distinguish the overall trend, a moving average technique was applied to the heroin crime hotspot block group time series data (Cryer and Chan, 2008). We selected the largest time interval between highest and lowest values in the time series – five months – as the temporal length to process the spatial pattern of heroin-related crime. Using this interval helped us to reliably detect the changing pattern of incidents, even as the pattern fluctuated. For synthetic drug incidents, the number of incidents by block group was much smaller, as there were less than 50 monthly synthetic drug-related incidents in all 2202 block groups during the study period. This number was too small to implement clustering analysis and therefore, we created a time series based on the raw count of synthetic drug-related incidents instead of by hotspot block groups. In this time series, the longest time interval between a highest and lowest number of incidents was nine months, and that was used as the temporal granularity to identify the changing patterns for synthetic drug activities.

We trained the random forest model using different combinations of factors including sociodemographic factors and built environment factors, as well as three variables that related to change over space and time. This included two lag variables (a time-lagged variable and a spatiotemporally lagged variable) and a trend variable. The time-lagged variable was a binary variable identified based on whether a given block group was in a hot spot of drug-related crime during the previous time period. The spatiotemporally lagged variable referred to the count of block groups in queen adjacent neighborhoods that belonged to a hotspot during the previous time period. The trend variable was computed based on the time intervals calculated to capture the changing patterns of clusters (Chi and Zhu, 2008), and tracked any increases or decreases in the numbers of neighboring block groups that belonged to a hotspot during the previous time period. As a final step, model prediction accuracy (i.e., percentage of block groups correctly predicted) was computed.

2.6.2 Including key locations of drug-related crime incidents in the model

To analyze locations for drug-related crime incidents and develop a foundation for building a machine-learning classifier, we summarized the most frequent locations for both heroin and synthetic drug-related crime incidents. According to the crime data provided by the Chicago Data Portal, there were 180 different location descriptions in total. While numerous types of locations were recorded, not all of these locations were useful for our analysis. Sidewalks and streets, for example, were the two most frequent locations for drug-related crime however, these are ubiquitous features in a city and so we did not utilize either of these features in our model. Similarly, specific locations at a residence (e.g., porch or yard), and vehicles were also recorded as frequent locations for drug-related crimes, however, these were also not used for this research. Instead, we used neighborhood features that were more uniquely identifiable including alleys, vacant buildings, vacant lots, parking lots, gas stations, parks, and high schools. To incorporate these key locations in the model, we calculated seven variables from key location layers, including the counts of gas stations, vacant lots, vacant buildings, and high schools, as well as alley density and area of park properties in each block group.

2.6.3 Model training, validating and out-of-sample testing

A random forest model was used to gain insights into the contributions of built environment and sociodemographic factors in classifying hotspots of both heroin and synthetic drug-related crimes and to capture the changing patterns of drug-related activities over space and time. The modeling was performed in R, an open-source ecosystem that provides multiple packages for machine learning, and used R packages ‘ranger’, ‘RRF’, ‘pdp’ and ‘rgdal’ (Bivand et al., 2008; Deng and Runger, 2013; Diggle, 2013; Wright and Ziegler, 2017).

Built environment and sociodemographic factors were used as input variables in the random forest classifier to identify whether a block group belonged to a hotspot or not. Bagging, also called bootstrap aggregating, is an ensemble algorithm that was used for selecting samples for each tree in the random forest. Typically, each bootstrap sample subsets approximately 63% of the training data while leaving out about 37% of the data (Breiman, 1996). The out-of-bag (OOB) error rate is the misclassification error rate of the estimates fitting the trees whose bootstrap samples do not have this observation. In this study, the OOB error rate was used to evaluate the fitting accuracy of the random forest model.

To test the model’s ability of predicting future spatial locations of heroin and synthetic drugs, an out-of-sample test using an independent test dataset was implemented to evaluate the model forecasting performance. We trained the space-time random forest model to associate the spatiotemporal patterns of drug-related crime with potential predictor variables using the data for the first three years (2016 ~ 2018). The drug-related crime data was aggregated by the computed temporal intervals (5 months for heroin-related crime and 9 months for synthetic drug-related crime) to generate spatial patterns. The training model used these aggregated patterns as output, and used one month as an offset during the training process.

Heroin data for the last five months of 2019 and synthetic drug-related crime data for the last nine months of 2019 were used as the independent test datasets in the out-of-sample test. Predictive accuracy (ACC) was used to evaluate the model's ability to forecast future drug hotspots. ACC has been widely used to assess the predictive performance of machine learning classifiers (Chen et al., 2017; He and Garcia, 2009) and is calculated as:

$$ACC = \frac{TP + TN}{TP + FP + TN + PN} \quad (1)$$

where TP (true positive) is the number of actual hotspot block groups that are correctly predicted; TN (true negative) is the number of actual non-hotspot block groups that are correctly predicted; FP (false positive) and FN (false negative) are the numbers of block groups that are incorrectly predicted.

Following the approach by Chainey et al. (2008), we computed a prediction accuracy index (PAI) that returned the density of drug-related crimes in the predicted block groups as compared to the density of drug-related crimes over the whole study area. We also computed a prediction efficiency index (PEI) that measured the ratio of PAI and maximum possible PAI that a model can achieve (Hunt, 2016). PAI and PEI were used to evaluate both the effectiveness and efficiency of the hotspot prediction models.

2.6.4 Feature selection: guided regularized random forest (GRRF)

For this research, a guided regularized random forest (GRRF) was used for feature selection (Deng and Runger, 2013). GRRF models use a variable importance score calculated from a preliminary random forest model to guide feature selection in the regularized random forest. Regularization aims at selecting high-quality feature subsets to avoid overfitting problems by penalizing inputting new features into the model (Deng and Runger, 2012). A feature with a higher importance score in the preliminary random forest is penalized less in GRRF. GRRF selects a compact feature subset and reduces redundancy among the selected features.

For this research, we collected 104 candidate factors in total, including 61 built environment factors and 43 sociodemographic factors. GRRF was used to select two compact variable subsets for heroin model and synthetic drug model separately. For heroin-related crime hotspot classification, input variables selected by GRRF included 21 built environment factors and 25 sociodemographic factors, as well as three spatial and temporal lag variables (Table A.2 in Appendix A). For synthetic drug-related crime hotspot classification, selected factors included 17 built environment factors and 23 sociodemographic factors. The three spatiotemporal lag variables were also included (Table A.3 in Appendix A).

2.6.5 Model optimization: handling class imbalances, grid search for model hyperparameters and weighting key location features at splitting nodes

To handle any class imbalance problems that arose due to the presence of spatial clusters or hotspots, two techniques, oversampling minority classes and underdamping majority classes (i.e. using bagging to repeatedly generate random subsets for each decision tree) have been recognized by researchers as efficient resampling methods (Japkowicz and Stephen, 2002; Kotsiantis et al., 2006; Liu et al., 2009). An imbalanced class problem existed in our synthetic drug training dataset, for example, for the first time period, January through September 2016, 2164 block groups belonged to the majority class (i.e., block groups that were not in a hotspot) while only 39 block groups belonged to the minority class (i.e., block groups belonging to a hotspot). We tested oversampling the minority class at multiple ratios of α (α = 100%, 200%, 300%, 400% and 500%) and repeated underdamping the majority class at ratios of β (β = 50%, 25%, 12.5%, 6.25%, and 3.13%) during the bootstrapping process. We compared the classification accuracy of these 25 (5 x 5) combinations and found that when α=200% and β = 25%, the model had the best prediction performance. For the heroin training dataset, the class imbalance problem was present but not quite as strong as with synthetic drug case. Resampling the heroin training dataset did not help to improve model performance in terms of prediction accuracy, so neither underdamping nor oversampling was implemented for the heroin training dataset.

The random forest model uses hyperparameters that need to be set, including numbers of drawn candidate variables at each split (*mtry*), the number of trees (*ntree*), the sample size of observations for each tree, and the minimum number of samples for each node (Probst et al., 2019). Two hyperparameters, *mtry* and *ntree* have been shown to influence prediction performance (Bernard et al., 2009; Biau and Scornet, 2015). Tuning hyperparameters can improve the accuracy of the random forest model to some extent. A model tuning method, grid search, was used for searching for the best combination of two hyperparameters, i.e., the number of trees (*ntree*) and numbers of drawn candidate variables at each split (*mtry*) in our random forest model (Probst et al., 2019). The typical default setting in a random forest classifier for these two hyperparameters is *ntree* = 500 or 1000, and *mtry* = $\sqrt{nv}$ where *nv* is the number of input variables. In our model, 54 variables were used for predicting heroin hotspots and 53 variables were used for predicting synthetic drug hotspots. The typical default setting for our classifier was *ntree* = 500 and *mtry* = 7, however this default setting might not always be the best setting. In our models, with a given *mtry*, the error rate of the misclassification became stable when *ntree* reached approximately 200. Therefore, we used the tuning grid for *ntree* from 200 to 1000 with step = 100 and *mtry* ranging from 3 to 14 with step = 1, then, set the grid search process with 5-fold cross-validation and 3 repeats per fold. We found that the best settings of hyperparameters for classifying the heroin-related crime hotspots was when using *mtry* = 8 and *ntree* = 500, and for synthetic drug hotspots when *mtry* = 14 and *ntree* = 400.

Weighting the features that were known to be more informative for detecting drug-related crime hotspots (i.e., predicting dependent variables) more than other variables can increase the contribution of these features and is likely to improve the classification accuracy (Amaratunga et al., 2008; Ma et al., 2011; Malley et al., 2012). In this study, the key locations that were identified from the drug crime records were considered to be more informative than other built environment or sociodemographic variables. Given this, we set higher weights for the key location variables to have a higher probability of these variables being selected in the splitting nodes.

2.6.6 Contributions of variables in the model

In order to understand the relationships between the different factors and the reported drug incident patterns, we analyzed variable importance that revealed the contributions of different factors to identify heroin and synthetic drug-related crime hotspots. To interpret each factor's contribution for classification, we used corrected impurity importance, also namely actual impurity reduction (AIR), to measure variable importance (Nembrini et al., 2018). Corrected impurity importance measures the improvement of the classification rate at splitting nodes without bias (Calle and Urrea, 2011; Nembrini et al., 2018; Strobl et al., 2007). We also constructed partial dependence plots (PDP) for the spatial and temporal lag variables (Friedman, 2001; Greenwell, 2017). PDP used a partial dependence function to measure the marginal effect of the variables on the classification outcome (Greenwell, 2017; Molnar, 2019).

This process provided insights into the relationships of drug-related activity patterns between successive time periods.

2.7. Applying the space-time random forest model to drug-related crimes

2.7.1 Model training and contribution of variables

To train the random forest classifier to learn the changing trends of drug-related crime patterns and reinforce the role of spatiotemporal autocorrelation underlying any of the changes, we fitted and validated the model using the data between 2016 and 2018, and evaluated the model performance in terms of the misclassification rate based on out-of-bag (OOB) error. The final time periods of 2019 ((i.e., August-December 2019 for heroin, April-December 2019 for synthetic drugs) were used to perform an independent test of the model's prediction performance. For classifying heroin crime hotspot clusters, the OOB error 2.41% showing that it correctly classified 97.59 % of block groups during the training process. For synthetic drug hotspots, the OOB error was 4.52% indicating that the classifier correctly classified 95.48% block groups.

In order to measure the contribution of each variable for classifying the hotspots, we used actual impurity reduction (AIR) to evaluate the variable importance. Variables used for classifying heroin crime hotspots (Figure 2.4a) and synthetic hotspots (Figure 2.4b) were ranked by their importance scores. For classifying heroin hotspots, the top two important variables were key location variables, vacant lots (*VacLot)* and vacant buildings (*VacBldg)*. The next two high ranking variables were sociodemographic variables, percentage of low-wage workers (*R_PctLowWage*) and percentage of the population with Bachelor's degree or higher (*PBachelorHigher)*. For classifying synthetic drug-related hotspots, high ranking variables were similar to those for heroin but the rank varied. The top two variables were *VacLot* and *VacBldg* (Figure 2.4b). A different set of sociodemographic factors (e.g. race and ethnicity, median household income, percentage of employment) were also correlated with synthetic drug-related crime locations.

We selected four built environment and sociodemographic factors that were in the top five both for classifying heroin hotspots and synthetic drug hotspots (Figure 2.5). These four variables were visualized by the boxplots for heroin-related crime hotspots, synthetic drug-related crime hotspots and other block groups that belonged to neither a heroin nor a synthetic drug hot spot. There were more vacant buildings and vacant lots in heroin and synthetic drug crime hotspots than in non-hotspot block groups (Figure 2.5a and 2.5b). The percentage of low-wage workers was higher in heroin and synthetic drug hotspots than in other areas (Figure 2.5c). Median *PBachelorHigher* in the other block groups was 19.4% that was much higher than either heroin (5.6%) or synthetic drug hotspots (6.6%) (Figure 2.5d).

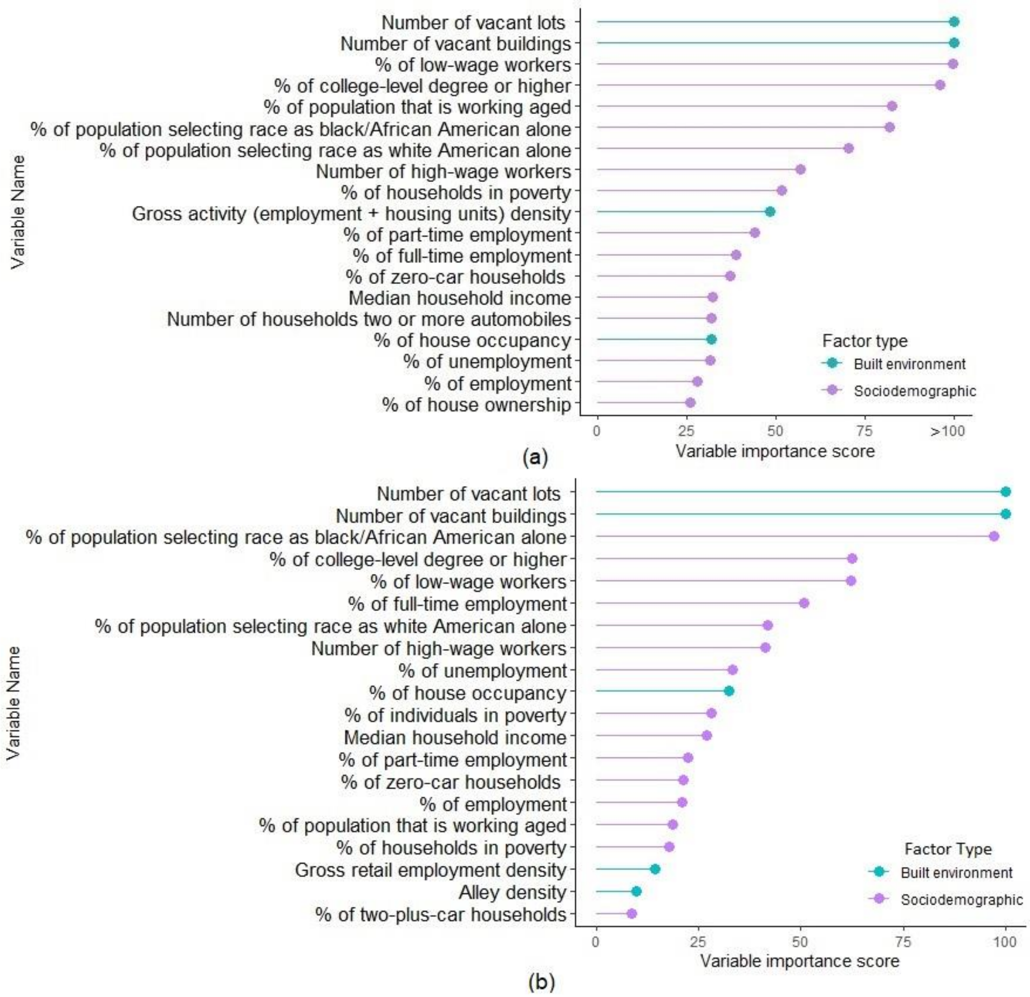

Figure 2.4 (a) top 20 variables with the highest importance score for classifying heroin hotspots; (b) top 20 variables with the highest importance score for classifying synthetic drug hotspots.

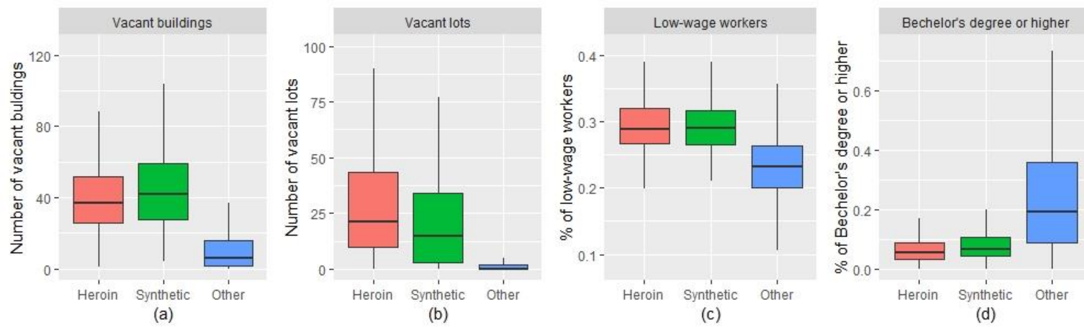

Figure 2.5 Boxplots of (a) number of vacant buildings; (b) number of vacant lots; (c) percentage of low-wage workers and (d) percentage of population with Bachelor's or higher degree for heroin-related crime hotspots, synthetic drug-related crime hotspots and other block groups.

To understand the relationships of drug crime patterns between successive periods *t -1* and *t*, partial dependence plots (PDP) were used to analyze the effect of one or two input features (i.e., spatial and temporal lag variables at *t-1*) on the prediction (i.e., the probability of a block group being classified as a hotspot block group at *t*). The length of a time period was calculated from the previous time series analysis. For example, in the heroin model, when time period *t-1* corresponded to January-May 2016, time period *t* corresponded to Feb-June 2016. For predicting heroin hotspots, the partial dependence plot of *nb_t_1* (number of neighboring hotspot block groups during *t-1* period) showed that being surrounded by more hotpot block groups resulted in a higher probability of a given block being classified as a hotspot block group in a successive time period *t* (Figure 2.6 a). When a given block group at *t-1* was surrounded by a low number of hotspot block groups, this block group was more likely to maintain its status in the successive time period (Figure 2.6 a, b). For predicting synthetic drug hotspots, a given block group tended to become the same status as its surrounding block groups (Figure 6 c, d).

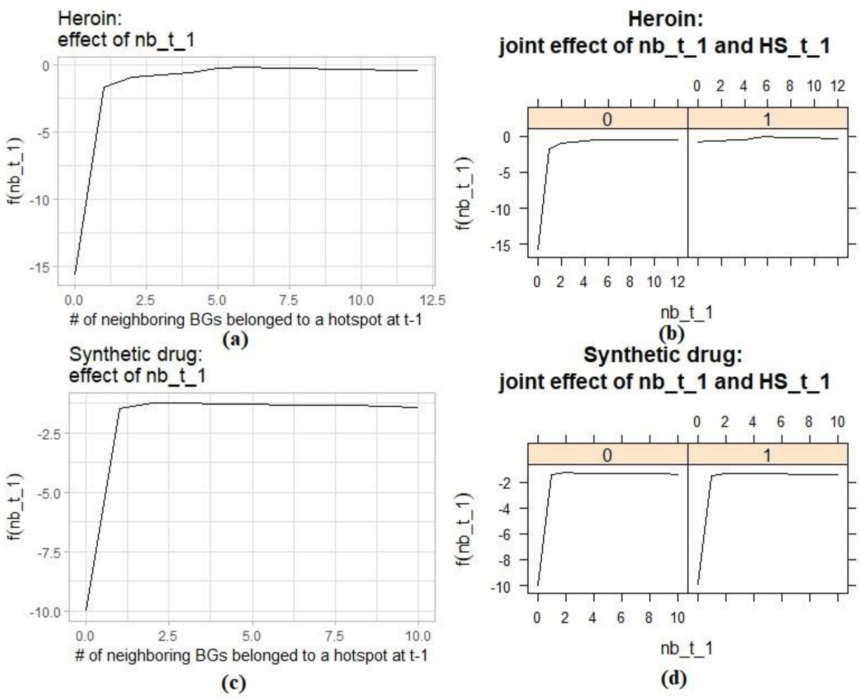

Figure 2.6 Partial dependence plots of classification of heroin-related crime hotspots on selected variables (a) *nb_t_1* (number of neighboring block groups belonged to a heroin hotspot at *t-1*); (b) joint effect of *nb_t_1* and *HS_t_1* (whether this block group belonged to a heroin hotspot at *t-1*, class *0* represented non-hotspot block group and class *1* represented hotspot block group); partial dependence plots of classification of synthetic drug-related crime hotspots on selected variables (c) *nb_t_1*; (d) joint effect of *nb_t_1* and *HS_t_1*.

2.7.2 Predicting heroin and synthetic drug hotspots

To analyze the predictive power of built environment and sociodemographic factors, and also the spatiotemporal variables in the space-time random forest model, we built seven models using different combinations of categories of variables as model input. We trained the model using the data between 2016 and 2018. As stated above, heroin data for the last five months and synthetic drug-related crime data for the last nine months of 2019 were used as independent test datasets to evaluate out-of-sample forecasting accuracy. For this out-of-sample testing, we compared the predicted hotspots with the actual hotspots calculated from the 2019 drug crime data.

To test the model's ability to predict heroin hotspots, built environment and sociodemographic factors assisted in locating the potential hotspots, while spatial and temporal lag variables enabled the classifiers to learn the changing spatial pattern of hotspots (Figure 2.7). Predicting heroin hotspots using only sociodemographic variables (Figure 2.7a) or built environment variables (Figure 2.7b) generated overprediction in the lower hotspot because the model learned information only from previous hotspots but did not capture the disappearing pattern, while only using spatial and temporal lag variables in the model (Figure 2.7 f) resulted in underestimating the size of the upper hotspot. Inputting spatiotemporal autocorrelation variables into the model improved the prediction effectiveness of the model as illustrated by the increase in PAI (Table 2.3), but the variations on prediction efficiency (PEI) of seven models were minimal. The model using the three types of variables (Figure 2.7g) was able to correctly predict 81.4% heroin hotspot block groups and 99.3% non-hotspot block groups, as the overall prediction accuracy (ACC) was 98.3% (Table A.4 in Appendix A).

For predicting synthetic drug hotspots, incorporating spatial and temporal variables enabled the random forest classifier to capture the expanding hotspots (Figure 2.8), while only using spatiotemporal autocorrelation in the model resulted in an overprediction of the size of the hotspot (Figure 2.8f). The built environment (Figure 2.8a) and sociodemographic factors (Figure 2.8b) both played important roles for predicting synthetic drug hotspots but neither of them enabled the model to learn the missing pattern of the small hotspot in the southeastern side. The best model (Figure 2.8g) for predicting synthetic drug-related hotspots successfully identified which block groups were not likely to become a hotspot block group (93.0 % correctly identified), but might underestimate the size of actual hotspots (60.0% correctly identified). The best model with the highest ACC, PAI, and PEI for predicting synthetic drug hotspots, 90.7%, used a combination of all three types of variables (Table 2.4 and Table A.4 in Appendix A).

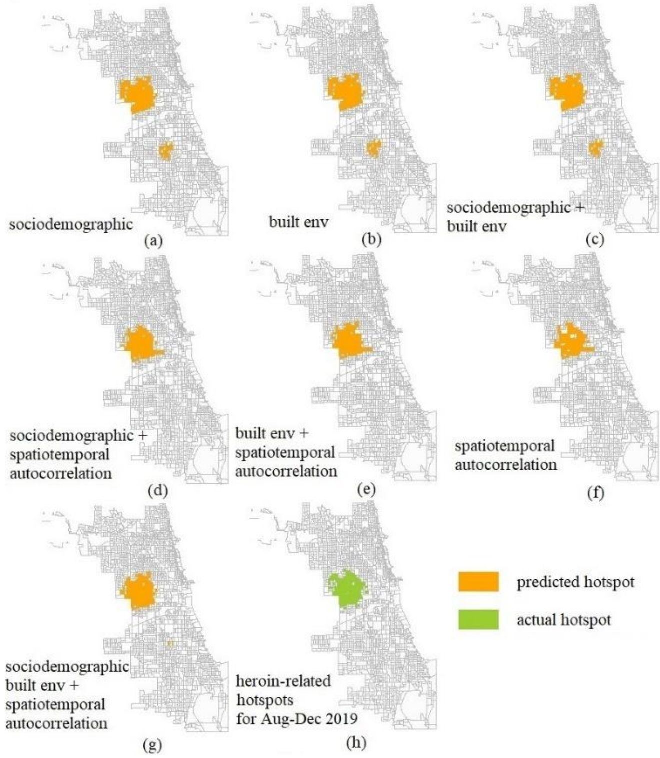

Figure 2.7 Space-time random forest model predicted heroin-related hotspot map using input variables: (a) sociodemographic factors (b) built environment factors (c) sociodemographic and built environment factors (d) sociodemographic factors and spatiotemporal autocorrelation (e) built environment factors and spatiotemporal autocorrelation (f) spatiotemporal autocorrelation (g) sociodemographic, built environment factors and spatiotemporal autocorrelation and (h) actual heroin hotspots calculated from the drug crime data, for August-December 2019 in Chicago.

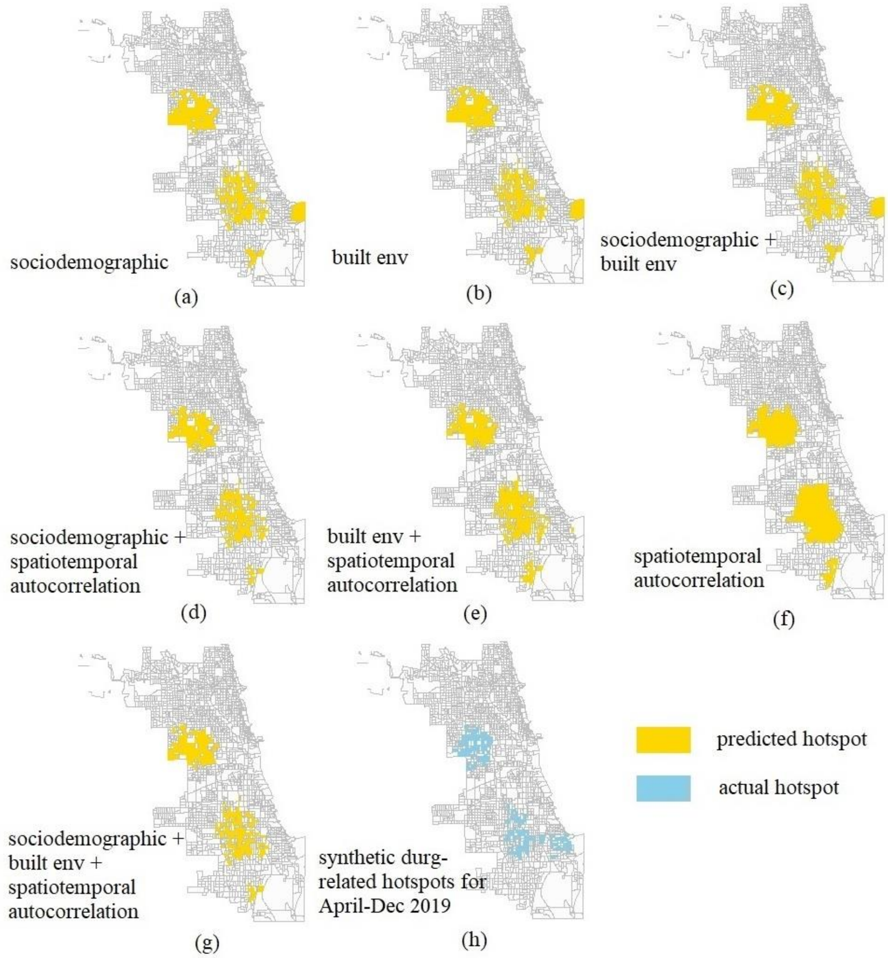

Figure 2.8 Space-time random forest model predicted synthetic drug-related hotspot map using input variables: (a) sociodemographic factors (b) built environment factors (c) sociodemographic and built environment factors (d) sociodemographic factors and spatiotemporal autocorrelation (e) built environment factors and spatiotemporal autocorrelation (f) spatiotemporal autocorrelation (g) sociodemographic, built environment factors and spatiotemporal autocorrelation and (h) actual heroin hotspots calculated from drug-related crime data, for April-December 2019 in Chicago.

Table 2.3 Model evaluation: classification accuracy (ACC), prediction accuracy index (PAI) and prediction efficiency index (PEI)

Drug Type Model	Heroin			Synthetic drug		
	ACC	PAI	PEI	ACC	PAI	PEI
Model 1 (sociodemographic)	97.7%	11.02	0.877	90.5%	4.37	0.548
Model 2 (built env)	97.7%	11.02	0.877	90.5%	4.37	0.548
Model 3 (sociodemographic + built env)	97.7%	11.02	0.877	90.5%	4.24	0.542
Model 4 (sociodemographic + spatiotemporal autocorrelation)	98.4%	14.34	0.877	90.5%	3.90	0.540
Model 5 (built env + spatiotemporal autocorrelation)	98.2%	14.75	0.876	89.1%	4.34	0.508
Model 6 (spatiotemporal autocorrelation)	98.3%	15.29	0.878	87.5%	3.47	0.544
Model 7 (sociodemographic + built env + spatiotemporal autocorrelation)	98.3%	13.44	0.876	90.7%	4.47	0.549

2.8 Discussion

In this research, a space-time random forest model was designed to investigate changing patterns of reported crime incidents involving different drugs in a major metropolitan city (Chicago) over time. One of the outcomes of this study was that we identified a set of factors useful for identifying heroin and synthetic drug hotspots in a city. We found that both heroin and synthetic drug hotspots were more likely to appear in the neighborhoods where there were more vacant buildings and vacant lots comparing to other non-hotspot block groups, while in synthetic drug-related hotspots, there were relatively more vacant buildings and less vacant lots than in heroin-related hotspots. While the characteristics of heroin and synthetic drug-related crime hotspots shared some similarities, we also investigated the possibility of an incident involving both heroin and synthetic drugs, but we found only a few (less than 0.1%) incidents were reported to involve both drugs. We also found that some sociodemographic factors were strong indicators for two types of drug hot spots but the rank of the factors varied. Overall, low income and education were the top two sociodemographic factors for predicting heroin hotspots, while low employment was also an important correlate with synthetic drug-related hotspots. Some of these top factors have also been found to be significantly related to drug activities in previous research. For example, research has established significant sociodemographic and built environment correlates for drug activities, typically including low education levels (McCord and Ratcliffe, 2007a), low income (Vilalta, 2010) and high unemployment rate (Cerdá et al., 2013a).

Incorporating spatial and temporal relationships into a random forest model enabled the model to learn the changing trends of the spatial patterns. In our research, incorporating space and time allowed the model to capture the concentration of block groups over time (heroin) or the increase in block groups (synthetic drugs) in Chicago during the study time period. We also compared our model results with a previous study that used a random forest model approach to predict crime hotspots (Borges et al., 2017). They used urban features and time-varying features such as timestamps of crime occurrences in a random forest model. We found that our approach based on incorporating spatiotemporal autocorrelation, improved the random forest model's performance in terms of prediction accuracy and efficiency by enabling the model to capturing the changing patterns of drug activities.

For this research, the drug-related crime data accessed from Chicago Data Portal were extracted from the Chicago Police Department's Citizen Law Enforcement Analysis and Reporting (CLEAR) system. We do not have information on policing practices during the study time period nor do we know to what degree the drug crime locations may be influenced by these practices. However, our analysis of opioid-involved mortality between 2016 and 2019 showed a positive association between the locations of opioid-involved deaths and drug-related crimes. Understanding, modeling and predicting the changing patterns of drug activities will further assist in locating substance use treatments and counseling services. It should also be acknowledged that drug-related crimes might be underreported (Mosher et al., 2010). For example, Mosher et al. (2010) indicated that high school administrators might be pressured to underreport or not report school crimes and minority groups had a greater tendency to underreport crime behaviors. Using crime-related data may miss additional locations where drugs were present (but no crime incidents were reported).

2.9 Conclusions

Our approach highlights a promising direction of using machine learning together with spatiotemporal autocorrelation to analyze changing patterns of repeated events – in this case, drug-related incidents – in cities. Incorporating space and time into a machine learning model assists in making more accurate forecasts of changing patterns. Future work could consider combining spatiotemporal heterogeneity with machine learning models by adding spatiotemporal bandwidths during the modeling process to investigate the patterns at varying spatial and temporal scales. In this study, a random forest model was used for a space-time analysis framework. Future research could examine other machine learning approaches such as gradient boosting model to test spatiotemporal approaches using other tools. This study also applied the space-time analysis framework and machine learning technologies to investigate the spatiotemporal patterns at block group level of more than 52,000 drug-related crime incidents, including more than 16,000 incidents involving heroin and synthetic drugs, in Chicago between 2016 and 2019. Our research found that while heroin-related crime incidents dropped slightly in 2017, they rose again in 2018, and continually increased in 2019, staying mostly in the same locations. The pattern of crime incidents involving synthetic drugs evolved from a scattered pattern in 2016 to two distinct hotspot areas in 2019. These hotspots were in the locations of 2016 and 2017 heroin-related hotspots. Understanding where drug-related crimes have been occurring is important as it can be indicative of where drug use is happening. Analyzing the geospatial patterns of different drugs including successfully being able to predict how these patterns have changed over space and time can provide a basis for applying further treatment services and mitigation efforts, and also be useful for assessing current related services and efforts. Identifying built environment drivers for drug hot zones such as key locations including vacant buildings and vacant lots where drug-related crime frequently occurred can help public safety stakeholders with effective decision making relating to particular drugs. Future work could investigate additional key locations in more detail.

Chapter 3: Spatiotemporal analysis of opioid-involved mortality in the context of the COVID-19 pandemic in the United States

3.1 Abstract

The global pandemic of SARS-CoV-2 (COVID-19) has been linked to adversely impacting individuals with opioid use disorder in the United States. This study focuses on analyzing opioid-involved mortality in the context of COVID-19 in the U.S. from a geospatial perspective. We investigated spatiotemporal patterns of opioid-involved deaths during 2020 and compared the spatiotemporal pattern of these deaths with patterns for the previous three years (2017-2019) to understand changes in the context of the COVID-19 pandemic. A counterfactual analysis framework together with a space-time random forest (STRF) model were used to estimate the increase in opioid-involved deaths related to the pandemic. To gain further insight into the relationship between opioid deaths and COVID-19-related factors, we built an STRF model for the City of Chicago, that experienced a steep increase in opioid-related deaths during 2020. High ranking indicators identified by the model such as the number of positive COVID-19 cases adjusted by population and the change in percentage of staying-at-home devices during the pandemic were used to generate a vulnerability index of opioid overdoses during the COVID-19 pandemic in Chicago.

3.2 Introduction

The global pandemic of SARS-CoV-2 (COVID-19) quickly became the most concerning public health issue in 2020, affecting hundreds of millions of people across the globe (Bergquist et al., 2020; World Health Organization, 2022). The COVID-19 pandemic has been reported to be also impacting the community of opioid misusers in the United States (National Institute of Environmental Health Sciences, 2020). National Institute of Environmental Health Sciences stated that nearly half of the Americans reported their mental health were negatively impacted by the COVID-19 pandemic due to various stresses including social isolation, job loss and lack of access to healthcare (National Institute of Environmental Health Sciences, 2020). Opioid treatment service deliveries were often disrupted during the COVID-19 pandemic. Recent studies pointed out that over 30 states reported opioid fatalities were increasing during the pandemic (American Medical Association, 2020).

Research on the spatiotemporal pattern of opioid-involved deaths in the context of COVID-19 is useful for understanding changes in opioid use and misuse during the pandemic and for providing decision support for the ongoing opioid crisis regarding the impacts of COVID-19 on vulnerable individuals (Davis and Samuels, 2020; Ghose et al., 2022). The research objective of this study is to analyze opioid-involved mortalities in 2020 during the COVID-19 pandemic in the U.S., from a geospatial perspective. In this study, we investigate the spatiotemporal patterns of opioid-involved deaths that occurred in all U.S. counties during 2020, and compare the spatiotemporal patterns of these deaths with patterns from the previous three years (2017-2019) to understand spatial changes in opioid-involved deaths in the context of COVID-19.

For this research, we use a counterfactual analysis framework together with a space-time random forest model to estimate opioid-involved deaths under the assumption that the COVID-19 pandemic did not happen in 2020, and compare these results with the actual opioid-involved deaths that did occur in 2020. Counterfactual analysis frameworks have been used for evaluating the impact of the COVID-19 epidemic (Parascandola and Weed, 2001; VanderWeele, 2020), for example, recent research on energy economics applied counterfactual analysis to assess the impact of COVID-19 on global CO2 emission in 2020 (Smith et al., 2021). The random forest model (Breiman, 2001), with strong prediction performance and ability in handling multicollinear relationships between a large group of variables has been used for opioid-related public health research (Ancuceanu et al., 2019b; Fernández-Delgado et al., 2014b; Kamel Boulos et al., 2019b). Previous research has accounted for spatiotemporal autocorrelation in the random forest model (i.e., space-time random forest model) to enable it to capture the changing trends of opioid use patterns in urban areas (Xia et al., 2021).

Sociodemographic factors (e.g., household income, education, age groups) and built environment factors (e.g., house ownership, rent property and house vacancy) have been identified as important drivers of opioid misuse patterns (Cao et al., 2020; Cerdá et al., 2013b; Li et al., 2022; Lipton et al., 2013; Nechuta et al., 2018; Srinivasan et al., 2003; Vilalta, 2010). These factors have also been used in analyses of COVID-19 deaths and mental health issues during COVDI-19 in recent studies (Bustamante et al., 2022; Grekousis et al., 2022; Lee et al., 2022; Mateo-Urdiales et al., 2021). In this study, sociodemographic and built environment factors are used in a space-time random forest model (STRF) to estimate opioid deaths for each county as part of the counterfactual analysis. The model is trained using opioid deaths data between 2017-2019 and used to estimate the number of deaths in 2020.

In the context of the COVID-19 pandemic, human mobility behaviors were impacted by policy inventions such that average home-dwelling times were increased due to stay-at-home orders and warning about infections (Xiong et al., 2020; Zhu et al., 2021). To understand the relationship between an increase in opioid deaths and the COVID-19 pandemic, we have accounted for COVID-19 impacts in the space-time random forest model (STRF_COVID) by including pandemic-related factors (e.g., COVID-19 positive cases and COVID-19 deaths) and social distance metrics (e.g., home dwelling time, percentage of people staying at home) as input variables for the model (Ghose et al., 2022; Huang et al., 2021). We compared the number of observed opioid deaths in 2020 with the estimated number of opioid deaths using both the STRF and STRF_COVID models, to better understand how the increase in opioid deaths in 2020 could be related to sociodemographic, built environment, and COVID-19 pandemic-related factors.

To gain further insights, we analyze opioid-involved deaths in 2020 for the City of Chicago, IL at ZIP-code level in order to understand the impact of COVID-19 on the ongoing opioid crisis in a U.S. metropolitan area. Chicago is located in Cook County, IL where the number of opioid-involved deaths was the highest among all U.S. counties between 2017-2020. There were 3,570 opioid deaths that occurred in Chicago between 2017-2020. In 2020, opioid-involved deaths in Chicago increased by almost 40%. A map of opioid overdose vulnerability index at ZIP-code level was generated for the City using the results of variable importance analysis to assist in evaluating the resilience against an opioid crisis during the pandemic period and provide further insights for drug health stakeholders and practitioners.

3.3 Materials and Methods

3.3.1 Materials

The data on opioid-involved deaths used in this study were accessed from Multiple Cause of Death Data provided by Centers for Disease Control Wide-ranging Online Data for Epidemiologic Research (CDC WONDER)[8]. The International Classification of Diseases 10th Revision (ICD-10) was implemented for death causes coding and classification in the U.S. (CDC, 2021). ICD-10 codes X40-X44 (accidental poisoning), X60-X64 (suicide), X85(homicide), Y10-Y14(undetermined poisoning) were used for extracting all drug-included deaths and ICD-10 codes T40.0 (opium), T40.1 (Heroin), T40.2 (other opioids), T40.3 (Methadone), T40.4 (synthetic narcotics) (including fentanyl), T40.6 (unspecified narcotics) were used for extracting opioid-involved mortalities (Substance Abuse and Mental Health Services Administration, 2018). The

[8] CDC WONDER: https://wonder.cdc.gov/

time series of county-level COVID-19 positive case and death data were accessed from an interactive COVID-19 dashboard operated by Johns Hopkins University (Dong et al., 2020). We also collected opioid-involved deaths, COVID-19 positive cases and mortality data used for ZIP-code level analysis in Chicago from Cook County Open Data[9] and the Chicago Data Portal[10]. Sociodemographic and built environment factors for 2017-2019 were derived the American Community Survey 5-year estimates provided by U.S. Census[11]. We collected unemployment data for all U.S. counties for 2017-2020 from Local Area Unemployment Statistics provided by the U.S. Bureau of Labor Statistics[12]. County-level personal income data for 2017, 2018, 2019 and 2020 were collected from U.S. Bureau of Economic Analysis (BEA)[13]. SafeGraph provided block group level social distance metrics dataset for 2019 and 2020 that was generated from GPS pings from anonymous mobile devices (SafeGraph, 2022). We used this dataset to aggregate four social distance variables at both county level and ZIP-code level, including median home-dwelling time, percentage of devices staying at home, percentage of devices at work, and percentage of home-dwelling time. We also computed the change in these four social distance variables in 2020 compared to 2019 (e.g., change in median home-dwelling time from 2019 to 2020). All factors derived

[9] Cook County Open Data: https://datacatalog.cookcountyil.gov/
[10] Chicago Data Portal: https://data.cityofchicago.org/
[11] American Community Survey, U.S. Census: https://www.census.gov/programs-surveys/acs
[12] Local Area Unemployment Statistics, U.S. Bureau of Labor Statistics: https://www.bls.gov/lau/
[13] Personal Income, U.S. Bureau of Economic Analysis: https://www.bea.gov/data/income-saving/personal-income

from these sources and vairables used for this study are listed in Table B.1 in Appendix B.

3.3.2 Space-time random forest model

A random forest model (Breiman, 2001) is composed of a large number of decision trees that can be used for both classification and regression. For regression tasks, the output of the random forest model is the average value of predictions from all individual trees. In a previous study, we incorporated spatiotemporal autocorrelation into a random forest model (STRF) to enable the model to capture changing patterns of drug-related activities (e.g., possession, delivery and manufacture of illicit drugs) (Xia et al., 2021).

In this study, we used a space-time random forest regression model to estimate the number of opioid deaths that occurred in 2020 during COVID-19. The increase in opioid-involved deaths in 2020 was likely due to two reasons, an increasing trend in opioid-involved deaths over the previous years and the impact of the COVID-19 pandemic. An increase in opioid-involved deaths can be captured by the STRF model. To accommodate spatiotemporal autocorrelation in the model, we computed three variables related to changes in the number of opioid deaths over space and time. This includes two lag variables (a time lag variable and a spatiotemporal lag variable) and one trend variable (Xia et al., 2021). The time lag variable refers to the number of opioid deaths in a county in the previous year. The spatiotemporal lag variable refers to the number of opioid deaths that occurred in neighboring counties in the previous year. The trend variable is based on the number of opioid deaths in the previous two years and tracks the increase or decrease in opioid deaths in a county. To estimate the number of opioid-involved deaths in 2020 and capture the accelerating increase in opioid-involved deaths during the stay-at-home period, we applied two components of spatiotemporal autocorrelation variables in the STRF models, including yearly and period-based variables relating to three COVID-19 periods. During March-May 2020, when the level of illness was very high, 42 states and territories in the U.S. placed mandatory state-at-home orders, impacting 2,355 (73%) of 3,233 U.S. counties (Moreland, 2020). Opioid-involved deaths also peaked during this period (Figure 3.2a). In this study, for the nationwide analysis of opioid deaths by U.S. counties, we included the period-based spatiotemporal autocorrelation variables (i.e., the time lag, spatiotemporal lag and trend variables) for three different periods in 2020 (January-February, March-May, and June-December) during the COVID-19 pandemic.

3.3.3 Counterfactual analysis

The counterfactual analysis framework together with a space-time random forest model was used in this study to estimate the increase in opioid-involved deaths during the COVID-19 pandemic. Firstly, we trained the STRF to estimate opioid-involved deaths between 2017-2019. The input variables included 28 demographic, economic and built environment factors, as well as 12 spatiotemporal autocorrelation variables computed based on opioid-involved death data for the previous two years (Table B.1 in Appendix B). Take one year as an example, to estimate the opioid-involved deaths in the U.S.

during 2017, the independent variable in the STRF was the opioid-involved deaths for each county in 2017. The dependent variables included demographic, economic, and built environment variables for 2017, plus the spatiotemporal autocorrelation variables (i.e., opioid-involved deaths for each county in 2016, opioid-involved deaths in the surrounding counties in 2016, changes in opioid-involved deaths from 2015 to 2016). Then, demographic, economic, and built environment variables for 2020 plus spatiotemporal autocorrelation variables were input into the STRF to estimate the opioid-involved deaths for 2020. We compared the estimation of opioid-involved deaths for 2020 with the actual deaths that occurred in 2020 to analyze the increase in opioid-involved deaths in the context of COVID-19. The increase in opioid-involved deaths in 2020 was likely linked to two reasons, an already-high (and increasing) number of opioid deaths in the U.S., and the impact of the COVID-19 pandemic. To understand the increase in opioid-involved deaths during the pandemic and assess the impact of this event on opioid-involved deaths, we also generated an STRF model that accounted for certain pandemic factors. These factors included the number of reported COVID-19 positive cases, COVID-19 deaths, social distancing metrics during the COVID-19 (e.g., median home-dwelling time), and the changes in human mobility patterns (e.g., change in median home- dwelling time from 2019 to 2020) for each of the three time periods. In addition, the STRF_COVID model included 30 additional factors related to the COVID-19 pandemic (Table B.1 in Appendix B) that were used as input in this model.

3.3.4 Opioid overdose vulnerability index for a city

To gain further insights into the increase in opioid deaths, we included sociodemographic characteristics as well as human mobility behaviors with the pandemic-related factors and built a local STRF_COVID model at ZIP-code level for the City of Chicago. The pattern of opioid-involved deaths in Chicago was analyzed for four time periods relating to local stay-at-home orders and the changing trends in opioid deaths during 2020 in the City. Period 1 (1/1/2020-3/20/2020) was the period before the stay-at-home order was placed in Chicago on 03/21/2020. Period 2 (3/21/2020-5/30/2020) refers to the period when a stay-at-home order was implemented in the State of Illinois (State of Illinois, 2020). Periods 3 and 4 referred to two time periods after the stay-at-home orders were lifted. These two periods correspond to changes in the increasing rate of opioid-involved deaths in Chicago. We analyzed the patterns of opioid-involved deaths during these four periods and compared with patterns from previous years. We examined the opioid death patterns and COVID-19 positive case patterns in Chicago to understand how the increase in opioid deaths was possibly impacted by COVID-19. The most important variables for estimating opioid-involved deaths were identified by the model. Using these results, a ZIP-code level vulnerability index map for opioid deaths during the COVID-19 pandemic in Chicago was calculated by summarizing the rank of each variable and weighted by variable importance.

3.4. Results

3.4.1 Spatiotemporal patterns of opioid-involved deaths across the United States

Between 2017-2020, 991 U.S. counties had reported data on opioid-involved deaths, the data for the remaining counties were suppressed due to data confidentiality when the total number of deaths in a county was less than ten cases. In 2020, approximately 68,632 opioid-involved deaths occurred in the U.S. Compared to the average annual opioid-involved deaths during 2017-2019, the total number of nationwide opioid-involved deaths increased by 42.7% in 2020. The increase in opioid-involved deaths occurred in 835 counties (i.e., 84.3% of counties with reported data) in 2020 (Figure 3.1). Opioid-involved deaths stayed at the same level or decreased in only 156 counties, corresponding to 15.7% of counties with reported opioid death data. In 2020, the highest opioid-involved deaths occurred in South Atlantic[14], East North Central[15] and Middle Atlantic[16] areas (Figure 3.1). Opioid-involved deaths increased in these three regions by 42.8% (South Atlantic), 25.0% (East North Central) and 28.2% (Middle Atlantic). Compared to the average annual deaths in the previous three years (2017-2019), the largest increase in opioid-involved deaths in 2020 occurred in the Pacific[17]

[14] According to U.S. Census Bureau Regions and Divisions, South Atlantic Division including Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia and West Virginia
[15] East North Central Division: Indiana, Illinois, Michigan, Ohio and Wisconsin
[16] Middle Atlantic Division: New Jersey, New York and Pennsylvania
[17] Pacific Division: California, Washington, Oregon, Alaska and Hawaii

(89.7%), East South Central[18] (57.9%) and Mountain[19] (52.9%) regions (Figure 3.2b). During 2020, the number of nationwide opioid-involved deaths was higher than the average number of deaths for the previous three years (Figure 3.2a). The number of opioid-involved deaths peaked in May 2020 with an estimated 7,308 deaths, increasing by 81.1% over the same period during 2017-2019.

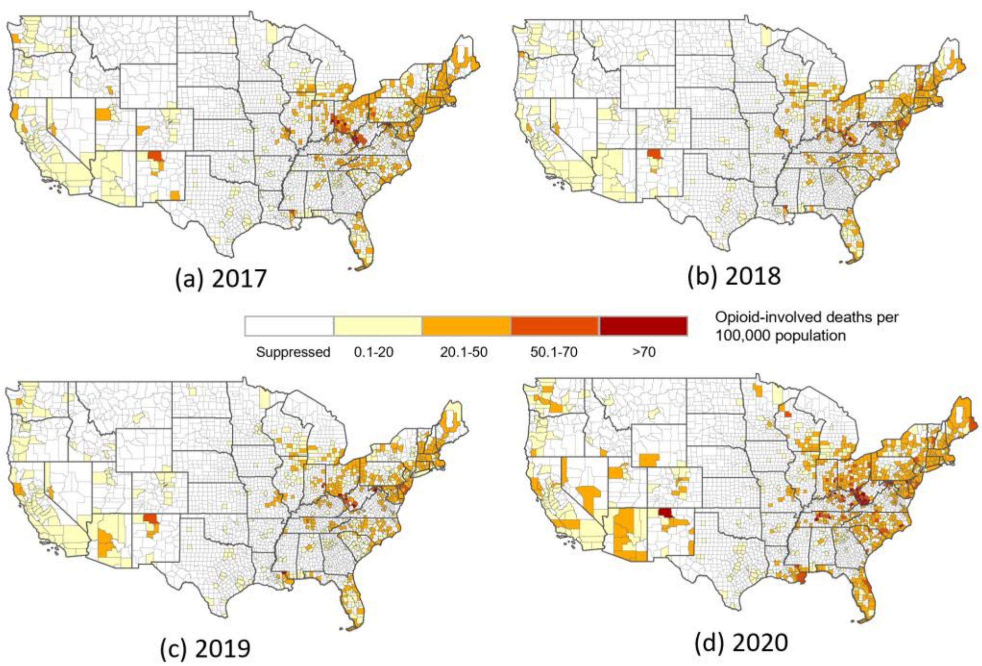

Figure 3.1 Opioid-involved deaths in the U.S. counties normalized by 100,000 population for (a) 2017, (b) 2018, (c) 2019 and (d) 2020.

---

[18] South Central Division: Alabama, Kentucky, Mississippi and Tennessee
[19] Mountain Division: Arizona, Colorado, Idaho, New Mexico, Montana, Utah Nevada and Wyoming

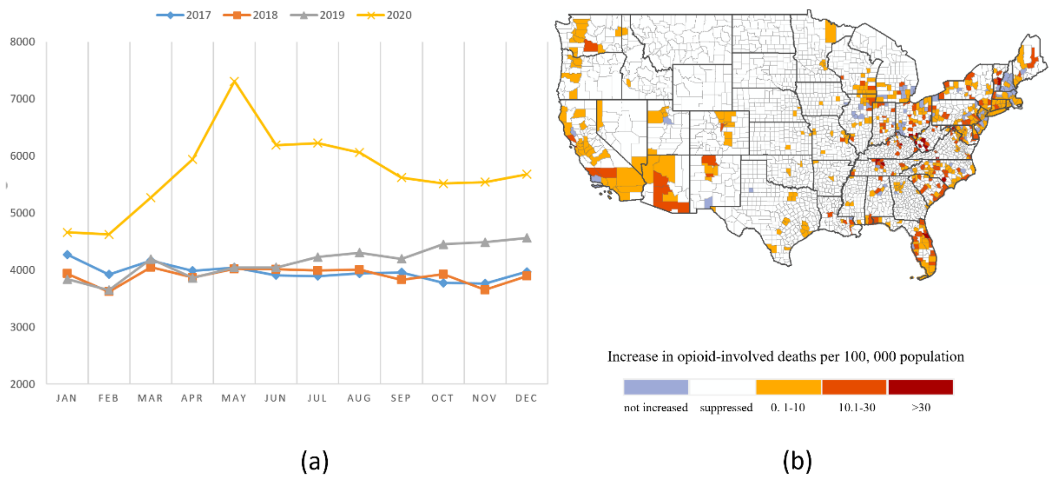

Figure 3.2 (a) Temporal pattern of opioid-involved deaths in the U.S. for 2017-2020; (b) Increase in opioid deaths by county in 2020 compared to the average number of deaths by county in the previous three years (2017-2019).

3.4.2 Space-time random forest model results accounting for pandemic-related factors

The STRF model was trained using the data between 2017-2019. The R-squared value of STRF was 0.895. To estimate the opioid-involved deaths in 2020 under the assumption that COVID-19 had not happened, sociodemographic and built environment factors for 2020 as well as the spatiotemporal autocorrelation variables were input into the trained STRF. The estimated opioid-involved deaths computed from the STRF model indicated that 77.0% of opioid-involved deaths could be explained by the model and input variables (Figure 3.3a). However, that left 23.0% of the increase in opioid deaths that were not able to be explained. The results confirm that the number of opioid deaths in the U.S. during 2020 increased under the impact of COVID-19. The highest underestimation of opioid-involved deaths by the STRF model (Figure 3.3 and

Figure 3.4a) occurred in the Pacific (33.3%), Mountain (27.9%) and East South Central (26.2%) regions.

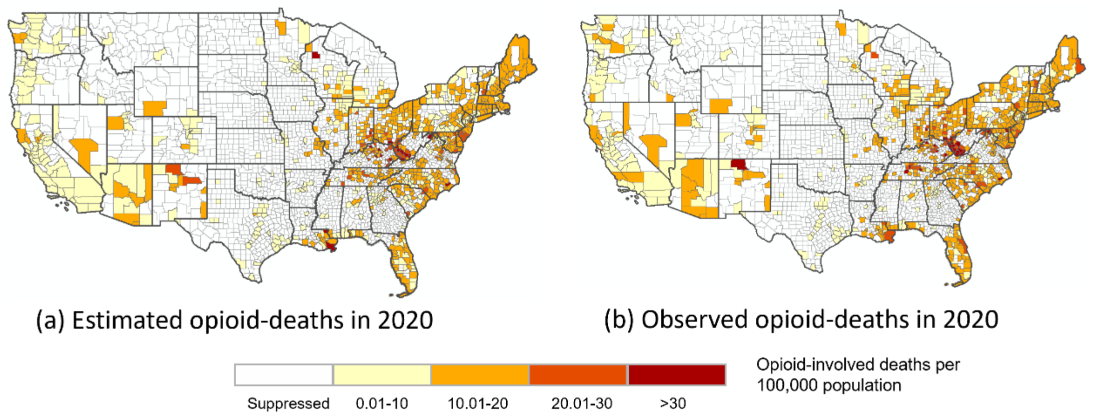

Figure 3.3 (a) Estimated opioid-involved deaths for U.S. counties in 2020 computed by STRF model (i.e., STRF not accounting for COVID-19 factors); (b) Observed opioid deaths in the U.S. in 2020.

To model the impact of COVID-19, we also built a STRF_COVID model using COVID-related, demographic, economic and built environment factors, and spatiotemporal autocorrelation variables as model input variables (Table B.1 in Appendix B). We compared the performance of the STRF_COVID models using the yearly or period-based spatiotemporal autocorrelation variables as well as using both of them in the model. The STRF_COVID model that accommodated both yearly and period-based spatiotemporal autocorrelation variables has the highest R-squared and lowest underestimation of deaths (Table B.2 in Appendix B). The yearly spatiotemporal autocorrelation variables enabled the model to capture a general increase in opioid-involved deaths over the 4 years, while the use of period-based spatiotemporal autocorrelation variables enabled the model to capture the accelerating increase in opioid-involved deaths for each period during 2020. The R-squared value of the STRF_COVID model accommodating two components, yearly and period spatiotemporal autocorrelation was 0.888 (Table B.2 in Appendix B). For this model, 88.6% of opioid-involved deaths in 2020 could be explained by the model and input variables including demographic, economic and COVID-19-related factors as well as spatial autocorrelation variables (Figure 3.4). The increase in opioid-involved deaths in 2020 that cannot be explained by the model was reduced to 11.4%. In this counterfactual analysis, comparing the two STRF model results, we found that accounting for COVID-19 factors reduced the underestimation of opioid-involved deaths by 50.4%.

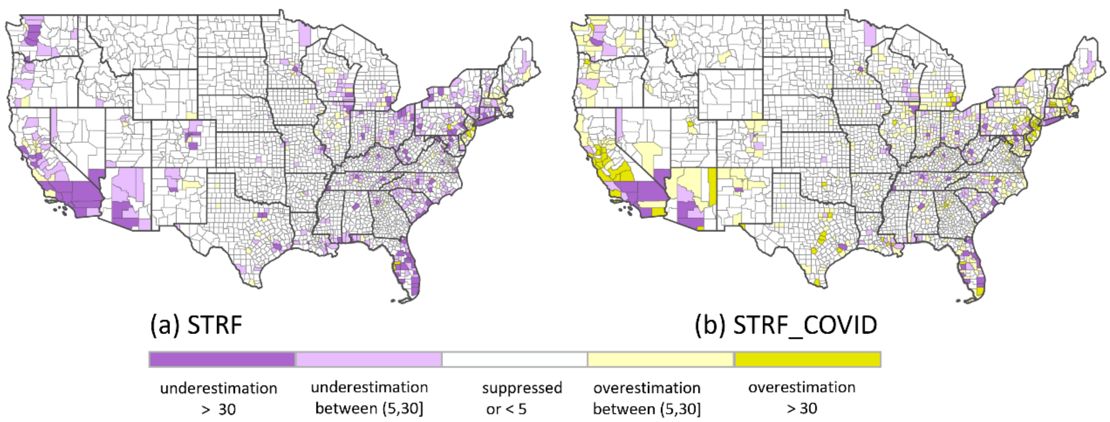

Figure 3.4 Comparison of model performances (i.e., distribution of underestimation and overestimation) for estimating opioid-involved deaths in the U.S. in the context of the COVID-19 pandemic using (a) STRF model that does not account for COVID-19 factors and (b) STRF_COVID that accounts for COVID-19 factors.

3.4.3 Spatiotemporal analysis of opioid overdose vulnerability in Chicago

In the City of Chicago, IL a major U.S. metropolitan area, the total number of opioid-involved deaths that occurred in 2017 was 753. The opioid-involved deaths in Chicago remained the same in 2018, and then increased in 2019 by 14.5%. In 2020, the number of opioid-involved deaths increased by 39.4% with a total number of deaths of 1,201. The pattern of COVID-19 positive cases over time coincided with that of opioid-involved deaths (Figure 3.5). The number of opioid-involved deaths peaked during Period 2, a time when the stay-at-home orders were enacted in the State of Illinois and the level of illness was very high (Figure 3.5a). Overall, the increase in opioid-involved deaths in 2020, compared to the average number of deaths for the previous three years (2017-2019), also reached a peak during this same period (Figure 3.5b).

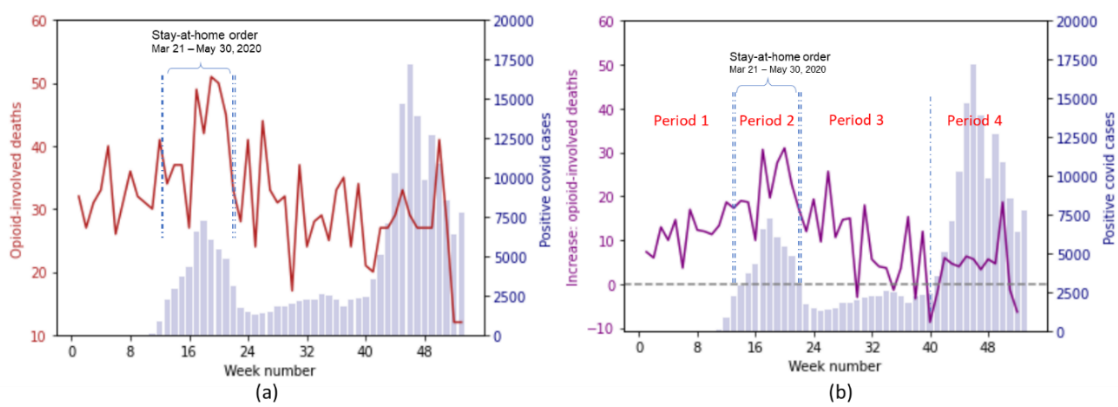

Figure 3.5 (a) Time series of opioid-involved deaths and positive COVID-19 cases in Chicago during 2020. (b) Time series of increase in opioid-involved deaths in Chicago during 2020 compared to average weekly deaths in 2017-2019.

As with other cities in the U.S., human mobility patterns changed over time during 2020 in Chicago. The daily median home-dwelling time in Chicago reached a peak, 921 minutes (i.e., 15 hours and 21 minutes) per day in Period 2, and then decreased to an average value of 680 minutes (i.e., 11 hours and 20 minutes) in Period 3, and remained around 677 minutes (i.e., 11 hours and 17 minutes) in Period 4 (Figure 3.6a). A positive relationship between patterns of median home-dwelling time and opioid deaths was illustrated by a Pearson's correlation index of 0.40 with a p-value smaller than 0.001 (Figure 3.5a and Figure 3.6a). The correlation between opioid-involved deaths and change in daily home-dwelling time in Period 2, 2020 (compared to the same period in 2019) was negative with a Pearson's correlation index of negative 0.49 and a p-value smaller than 0.001 (Figure 3.6b). Our analysis found that ZIP-code areas with greater increases in home-dwelling time were more likely to be associated with lower numbers of opioid-involved deaths.

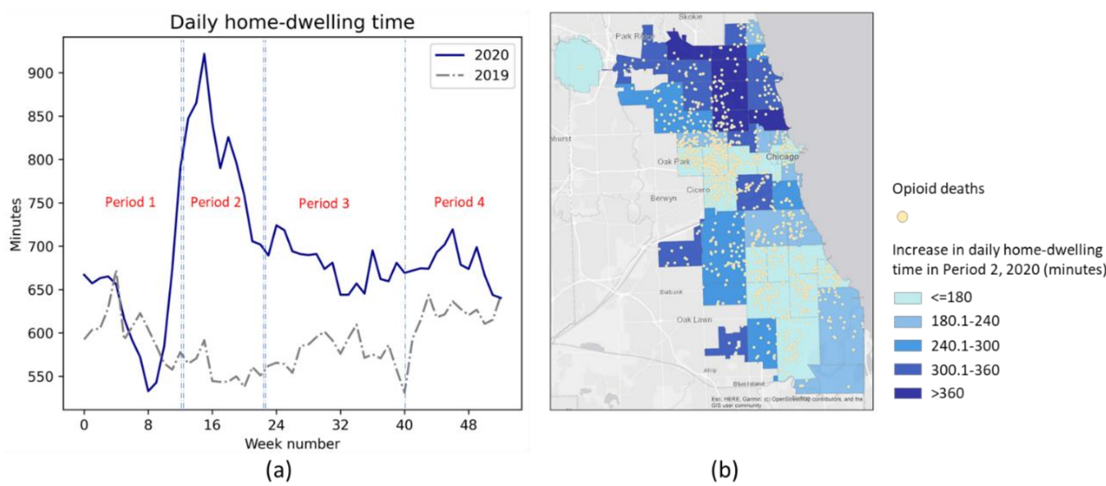

Figure 3.6 Spatiotemporal patterns of human mobility in Chicago during 2020: (a) time-series of median home-dwelling time, (b) change in daily home-dwelling time in Period 2, 2020 compared to the same period in 2019 by ZIP code areas and the distribution of opioid deaths.

The STRF_COVID model was applied to ZIP-code level data for Chicago and the random forest variable importance score was used for identifying high ranking variables (variables with a variable importance score higher than 5) relating to the number of opioid-involved deaths (Figure B.1 in Appendix B). Spatiotemporal autocorrelation variables (e.g., opioid deaths by time period and opioid deaths in neighboring ZIP-code areas in 2019) had high rankings. Other highly-ranked factors included economic variables (e.g., median household income and median property rent), demographic variables (e.g., percentage of population with college-level degree or higher, race and ethnicity), built environment variables (e.g., percentage of property ownership and percentage of vacant property), and COVID-19 variables (e.g., COVID-19 positive cases adjusted by population and change in percentage of completely staying-at-home devices). In total, 24 identified highly ranked indicators (Figure B.1 in Appendix B) were then used to generate an opioid overdose vulnerability index for 2020 at ZIP-code level. Five classes were generated for the vulnerability index using the equal numbers classification method (Figure 3.7). The index shows that in 2020, 56% of opioid deaths and 61% of the increase in opioid deaths in Chicago (compared to the average number of deaths in 2017-2019) occurred around West Side and Far Southeast Side areas of the city (red area on the map) and indeed the top two classes (red and orange areas) representing 76% of opioid deaths and 85% of the increase in deaths were in a broader extent of these same areas in the City (Figure 3.7).

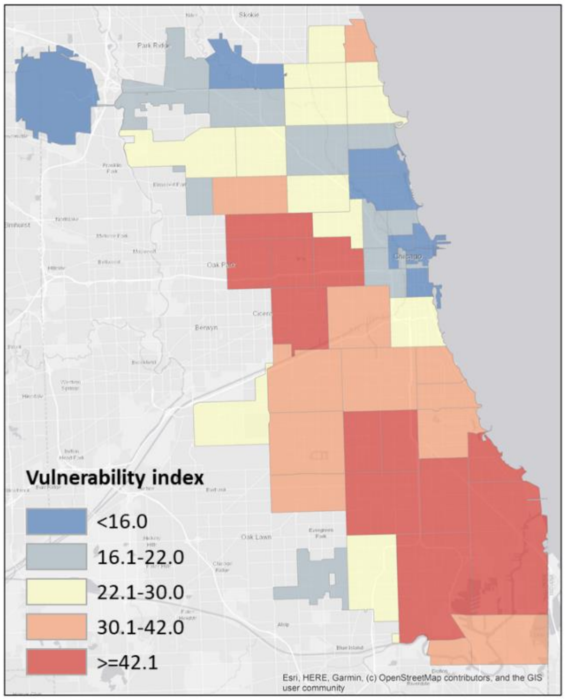

Figure 3.7 ZIP-code level opioid overdose vulnerability index for the period of time during the pandemic for Chicago, IL.

3.5 Discussion

In this study, the spatiotemporal pattern of opioid-involved deaths in the U.S. during the COVID-19 pandemic in 2020 was analyzed and compared to the pattern for each of the previous three years (2017-2019). Previous research on COVID-19 and drug overdose deaths (i.e., opioid-involved deaths and other non-opioid deaths) found that May 2020 was the deadliest month for drug overdose deaths in the U.S. with an estimated 9,192 deaths, representing a 57.6% increase over May 2020 (Friedman and Akre, 2021). Their study noted that particularly large increases in drug overdoses were observed in East South Central, South Atlantic, and Pacific areas in May 2020. Their findings have some similarity with our findings as our results returned the highest number of opioid-involved deaths for the South Atlantic, East North Central and Middle Atlantic regions, and the largest increase in opioid-involved deaths occurred in the Pacific area (89.7%), followed by East South Central (57.9%) and Mountain (52.9%) regions.

A counterfactual analysis using both an STRF and STRF_COVID model was used for estimating the number of opioid-involved deaths under the assumption that COVID-19 had not occurred in 2020. In the counterfactual analysis, the STRF was able to explain 77.0% of opioid-involved deaths in 2020, using demographic, economic, built environment and spatiotemporal autocorrelation factors to account for the numbers of opioid-involved deaths, while STRF_COVID was able to explain 88.6% opioid-involved deaths in 2020, reducing the underestimation by 50.4%. The counterfactual analysis indicated that the largest underestimation of opioid-involved deaths by the STRF model that did not account for COVID-19 was in the Pacific region (33.3%), followed by Mountain (27.9%) and East South Central (26.2%) regions. In an earlier study that used a Poisson model under a counterfactual assumption to analyze the increase in fatal drug overdoses during 2020 in California, the authors found that their estimated number of deaths were lower by 1,766 cases (33.3%) as compared to the actual 5,297 opioid-involved deaths in California in 2020, (Kiang et al., 2022). This underestimation rate in the number of opioid-involved deaths in California is similar to the underestimation rate that we found in our study (i.e., the number of opioid-involved deaths in the Pacific region was also underestimated by 33.3% when COVID factors were not accounted for in STRF model).

To generate an opioid overdose vulnerability index map based on 2020 data, we applied a local STRF_COVID model for the City of Chicago at ZIP-code granularity and identified indicators that were highly ranked for importance in estimating the numbers of opioid overdoses during the pandemic. Vulnerability indices have been used in previous studies to identify community resilience against COVID-19 (Tiwari et al., 2021; Urban Data Visualization Lab, 2022). In our study, the opioid overdose vulnerability index for Chicago was generated to evaluate the resilience against an opioid crisis during the pandemic period at ZIP-code level. The local STRF_COVID model for Chicago accounted for sociodemographic, economic, built environment and COVID-19 factors as well as the spatiotemporal autocorrelation. The highest ranked factors included spatiotemporal autocorrelation factors (e.g., the number of opioid-involved deaths in the previous year), median household income, race and ethnicity, and level of education (i.e., percentage of population with college-level degree or higher). Some of these factors have been found to be related to opioid use in previous studies including income, education level and unemployment rate in previous studies (Cerdá et al., 2017; Ghose et al., 2022; McCord and Ratcliffe, 2007b; Vilalta, 2010; Xia et al., 2021). Demographic factors, race and ethnicity, were also identified as important indicators for estimating opioid-involved deaths during COVID-19 in our study. A research on health disparity also indicated that a higher portion of COVID-19 deaths occurred among non-Hispanic Black residents in Chicago (Pierce et al., 2021). The pandemic had disproportionately impacted racial and ethnic minorities in Chicago.

In this study, to understand the increase in opioid-involved deaths during the COVID-19 pandemic, the counterfactual analysis compared the estimated opioid-involved deaths with the actual number of deaths that occurred in 2020. However, the underlying estimates of opioid-involved deaths assuming that COVID-19 did not occur in 2020 cannot be accurately known. We estimated the number of opioid-involved deaths in 2020 by accounting for the changing patterns in the previous three years and possible drivers for this change. The county-level opioid-involved deaths data used for this study were accessed from CDC WONDER. The data for the counties where the total number of opioid-involved deaths was fewer than ten have been suppressed due to confidentiality (Centers for Disease Control and Prevention, 2022; Quick, 2019). Without this data available, our analysis may have underestimated possible changes in these counties.

3.6 Conclusion

Spatiotemporal analysis of the number of opioid-involved deaths in the context of COVID-19 is important for understanding changes in opioid use and misuse during the pandemic and for the health and well being of populations during the pandemic. This study highlights the relationships between changes in opioid-involved deaths and associated sociodemographic, economic, and built environment factors during COVID-19. A counterfactual analysis was used to estimate how the number of opioid-involved deaths changed for 2020 when the pandemic was ongoing. This type of analysis assists in quantifying the changes in the number of opioid deaths that can be attributed to the

COVID-19 pandemic. Future work could consider in more detail the impacts of closures and state policies (e.g., school closures, restaurant closures) in the counterfactual analysis to further evaluate impacts on the number of opioid-involved deaths during the pandemic. In this study, we also built a model to identify some of the key factors related to opioid-involved deaths for the City of Chicago and created an opioid overdose vulnerability index at ZIP-code level that can be used to enhance awareness of the impact of the pandemic on the opioid crisis for specific areas in the City, and provide further insights for city and county drug health stakeholders and practitioners.

Chapter 4: Tracking the spatial dynamics of the synthetic opioid crisis in the U.S.

4.1 Abstract

Synthetic opioids are the most common drugs involved in drug-involved overdose mortalities in the U.S. The Center for Disease Control and Prevention reported that in 2018, about 70% of all drug overdose deaths involved opioids and 67% of all opioid-involved deaths were accounted for by synthetic opioids. In this study, I investigated the spread of synthetic opioids between 2013 and 2020 in the U.S., and analyzed the relationship between the spatiotemporal pattern of synthetic opioid-involved deaths and another key opioid, heroin, and compare patterns of deaths involving these two types of drugs during this time period. Spatial connections between counties were incorporated in to a graph convolutional neural network model to represent and analyze the spread of synthetic opioid-involved deaths, and in the context of heroin-involved deaths.

4.2 Introduction

Synthetic opioids, including fentanyl, are the most common drugs involved in drug-involved overdose mortalities in the U.S. In 2018, the Centers for Disease Control and Prevention (CDC) reported that about 70% of all drug overdose deaths involved opioids, and 67% of all opioid-involved deaths were attributed to synthetic opioids (Centers for Disease Control and Prevention, 2020a, 2020b). Fentanyl, a very potent synthetic opioid, can be combined with opium or cocaine as a combination product so that fentanyl can be used with or without the intention of users. Fentanyl has become a threat nationally (Centers for Disease Control and Prevention, 2019a; Marshall et al., 2017; Spencer et al., 2019) as synthetic opioids have been increasingly identified as a leading factor in overdose deaths in at least 23 states since 2015 (Centers for Disease Control and Prevention, 2019b; National Institute on Drug Abuse, 2019). Synthetic opioid-involved deaths increased rapidly between 2013 and 2020. In 2020, the number of synthetic opioid-involved deaths was estimated at 56,518, a count that was approximately 18 times higher than in 2013 (Figure 4.1).

Figure 4.1 Time series of both synthetic-involved and heroin-involved deaths in the United States, 2013-202 (data accessed from CDC WONDER)

Previous studies have determined that the increase in synthetic opioid-involved fatalities during 2013-2018 was driven mainly by an increase in fentanyl-involved overdose fatalities (O'Donnell et al., 2020, 2017; Peterson et al., 2016). The National

Forensic Laboratory Information System (NFLIS) of the Drug Enforcement Administration (DEA) reported that the amount of fentanyl law enforcement submissions (i.e., fentanyl seizures submitted and analyzed in State, Federal and local forensic laboratories) across the nation increased by 100 times more in 2019 than the amount reported in 2013. Another earlier study also pointed out that the increase in synthetic opioid-involved deaths from 2013 was correlated with the increase in the amount of fentanyl law enforcement submissions, but was not correlated with the amount of prescribed fentanyl (Gladden et al., 2016). These previous studies indicated that sources of fentanyl that have driven the synthetic opioid epidemic are more likely to be illegally manufactured rather than pharmaceutically produced. Illegal manufactured fentanyl and fentanyl in the U.S. have been shown to be primarily smuggled in from other counties, entering the U.S. from China as well as Mexico (Miller, 2020; Pardo and Reuter, 2020).

Graph neural networks (i.e., graph deep learning models) are neural models that can capture the dependence between nodes of graph from graph-structured datasets. Graph neural networks have been used in recent years for analyzing and forecasting spatiotemporal patterns in a wide array of fields including traffic forecasting, environmental index forecasting and crime rate prediction (Bui et al., 2022; Hu et al., 2022; Li and Zhu, 2021; Zhou et al., 2021). A recent study used graph deep learning models for predicting network-based hotspots of crime events including burglary, assault and theft (Zhang and Cheng, 2020b). However, graph neural networks are still a relatively novel computational approach in the area of health geography using health data for spatial analysis. In this study, I propose two ways to convert data on rates of county-level synthetic opioid-involved deaths to a graph-structured representation using both a spatial adjacency-based graph and a human mobility-based graph in order to incorporate the spatiotemporal correlation into graph neural networks. The graph neural networks were built to estimate whether synthetic opioid-involved deaths would occur in a county or not.

Two major research objectives are pursued in this study. The first research objective is to understand the relationship between the spread of synthetic opioids and the spread of heroin, and compare the spatiotemporal patterns of synthetic opioid-involved deaths and heroin-involved deaths between 2013 and 2020 across the U.S. The second research objective is to analyze the role of possible contributing factors to the spatial spread of deaths associated with the synthetic opioid crisis (e.g., human mobility by county accounting for spatial adjacencies between counties) using a graph learning approach.

4.3 Methods

4.3.1 Data

Synthetic opioid- and heroin-involved data used for this study were accessed from Multiple Cause of Death Data provided by the Centers for Disease Control Wide-ranging Online Data for Epidemiologic Research (CDC WONDER)[20]. In this dataset,

---

[20] CDC WONDER: wonder.cdc.gov

different causes of deaths are classified using the International Classification of Diseases, 10th Revision (ICD-10). The underlying cause of death codes used for extracting drug overdose mortality includes X40–X44 (accidental poisoning); X60–X64 (intentional poisoning); X85 (assault); and Y10–Y14 (poisoning undetermined intent) (Hedegaard et al., 2020). The multiple cause of death code for identifying deaths involving synthetic opioids is T40.4 - synthetic opioids other than methadone (e.g., fentanyl, fentanyl analogs, and tramadol). It should be noted that opioid types in the CDC WONDER mortality dataset were based on ICD-10 codes recorded on death certificates (CDC, 2020), which might lead to underestimation of drug overdose deaths or misclassification of deaths involving particular drug types (Hurstak et al., 2018; Mertz et al., 2014; Ruhm, 2016).

Sociodemographic and built environment features for U.S. counties were derived from the 2013-2020 American Community Survey (ACS) conducted by U.S. Census Bureau[21] and Smart Location Database[22] released by U.S. Environment Protection Agency (EPA) in 2021. The national county boundary file that was used for generating a spatial adjacency-based graph was accessed from the U.S. Census Bureau[23]. The raw human mobility flow dataset used in this study was provided by GeoDS Lab at the University of Wisconsin-Madison (Kang et al., 2020). This human mobility flow dataset (i.e., original-destination matrix (OD matrix)) was generated by tracking the

---

[21] ACS: census.gov/programs-surveys/acs
[22] Smart Location Database: epa.gov/smartgrowth/smart-location-mapping#SLD
[23] U.S. Census: census.gov

movement trajectories of mobile phone users collected by SafeGraph[24] at census block group level and was aggregated to county-level for this study.

4.3.2 Time-lagged cross correlation analysis

To understand the relationship between the changing patterns of synthetic opioid-involved deaths and heron-involved deaths, time-lagged cross correlations (Kilroy et al., 2017; Probst et al., 2012) were used in this study to analyze the time delay between the patterns of numbers of deaths. Time-lagged cross correlation calculated a series of correlation indexes between the time series of synthetic opioid-involved deaths and lagged versions of the time series of heroin-involved deaths using different lags, ranging from one temporal unit to the length of the study period, to identify the highest correlation and the related lag durations. Time-lagged cross correlation can measure the degree of similarity between the time series of synthetic opioid-involved deaths in the U.S. and the lagged version of the time series of heroin-involved deaths for the period from 2013 to 2020 when fentanyl use was on the rise. Time-lagged cross correlation analysis identified the two time series as having a *leader-follower* relationship, i.e., the follower will repeat the leader's patterns. In this study, heroin-involved deaths (i.e., the *leader*) initiated a pattern which was repeated over time by synthetic opioid-involved deaths (e.g., the *follower*).

---

[24] SafeGraph: safegraph.com

4.3.3 Incorporate spatiotemporal patterns into the model

The spatiotemporal autocorrelations between synthetic opioid-involved deaths that occurred in surrounding areas (adjacent counties) over successive years, were incorporated into a graph learning network to generate a model that can capture the spread of synthetic opioid-involved deaths over space and time. A set of spatiotemporal lag variables (e.g., synthetic opioid-involved deaths that occurred in a county for previous years) were calculated and used as node features in the graph learning network model. The heroin-involved deaths that occurred in a county for previous years and the time lag between synthetic opioid-involved deaths and heroin-involved deaths (i.e., the time lag identified by the time-lagged cross correlation analysis), were also incorporated into the graph learning network work model to enable the model to capture the leader-follower patterns during the learning process.

4.3.4 Convert the presence of opioid-involved deaths at county level to graph-structured data

A graph learning network requires the input dataset to be in the format of a graph data structure. In this study, two different graphs were used to represent the conversion of county-level opioid-involved deaths data to graph data, where one graph was based on spatial adjacency relationships (i.e., Queen weight adjacency) and a second graph on human mobility flows. The spatial adjacency-based graph accounted for the spatial correlation between neighboring counties, while the human mobility flow-based graph accounted for the spatial correlation between counties having movement interactions (i.e., number of trips and estimated population flow between two counties). The graph data structure for both graphs included four components, nodes (i.e., U.S. counties), nodes' features (i.e., sociodemographic and built environment factors, spatiotemporal variables related to synthetic opioid-involved deaths and heroin-involved deaths, and any time lags between synthetic opioid deaths and heroin-involved deaths), edges (i.e., spatial adjacency or mobility between counties) and edge weights (i.e., equal weights for the spatial adjacency-based graph and flow-based weights for the mobility-based graph). The edges for the human mobility-based graph were weighted by the value of population flow between each pair of two counties.

4.3.5 Building a graph learning network for opioid-involved deaths

To analyze the power of both spatial adjacency and mobility-based graphs for modeling the spread in the presence of synthetic opioid-involved deaths, we build two graph convolutional neural network (GCN) models using each of these graphs. The input data is the graph-structured data for U.S. counties including nodes, nodes' features, edges and edge weights. The output of the GCN model is a binary variable referring to whether opioid-involved deaths occurred in a county or not. The programming environment for the modeling process in this study was python 3[25] and a library for graph neural network PyG[26] built on top of the deep learning library PyTorch[27]. To

[25] Python: https://python.org/
[26] PyG: https://pyg.org/
[27] PyTorch: https://pytorch.org/

model the general spread of synthetic opioid-involved deaths, the spatiotemporal patterns of synthetic opioid-involved deaths in 2020 (i.e., during the COVID-19 pandemic) were analyzed but were not included in the modeling process. The GCN models were trained using the data on opioid-involved deaths and heroin-involved deaths for 2016-2018. The data for 2019 was used for an independent test to evaluate the model performance by comparing the estimation accuracy (i.e., the rate of correctly estimating the number of opioid-involved deaths that occurred in a county in 2019). For comparing the performances of the graph learning model with other models, I also build a random forest model, using the node's features as the attribute variables in the random forest model and compare the random forest model results with graph learning model results.

4.4 Results

4.4.1 Spatiotemporal patterns of synthetic opioid- and heroin-involved deaths

Synthetic opioid-involved overdose deaths have been rapidly increasing since 2013. The number of synthetic opioid-involved deaths in the U.S. increased from 3,105 in 2013 to 56,518 in 2020 (Table 4.1). Synthetic opioid-involved deaths accounted for 12.4% of all opioid-involved deaths, and this ratio increased to 82.3% in 2020.

Mapping the spatial pattern of synthetic opioid-involved deaths showed that the spread began in the western U.S. (Figure 4.2) in 2013. In 2016 and 2017, the number of synthetic opioid-involved deaths greatly increased nationwide, with the distribution spreading throughout the Northeast and mid-Atlantic regions.

Table 4.1 Synthetic opioid-involved deaths and heroin-involved deaths in the U.S., 2013-2022

Drug type	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Synthetic opioid	3,105	5,544	9,580	19,413	28,466	31,335	36,359	56,518
Heroin	8,259	10,574	12,989	15,469	15,482	14,996	14,019	13,165
All opioid	25,054	28,647	33,092	42,249	47,600	46,802	49,860	68,632

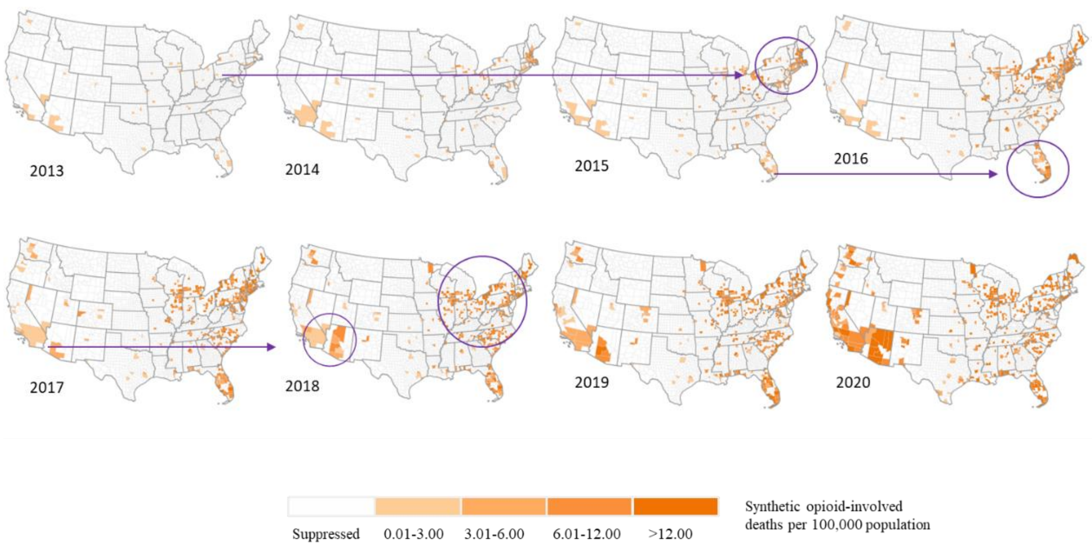

Figure 4.2 Synthetic opioid-involved deaths (adjusted by population) in the U.S. aggregated by counties by year for 2013 to 2020

For heroin-involved overdose deaths, the number peaked in 2017 before decreasing by 3.0% in 2018, 6.5% in 2019 and 6.1% in 2020. In 2016, the total number of synthetic opioid-involved deaths exceeded that of heroin-involved deaths, and were found to occur mostly in regions that had high heroin overdose rates, including the New England, Middle Atlantic and East North Central regions (Figure 4.3). In 2020, synthetic opioid-involved overdose deaths occurred in 795 U.S. counties (Figure 4.2), while heroin-involved overdose deaths occurred in 276 counties. In 2020, there were 271 counties with records of both synthetic opioid- and heroin-involved deaths (Figure 4.3). Overall for this year, the number of counties with synthetic opioid-involved deaths was 288% higher than the number of counties with heroin-involved death. The spatial pattern of deaths related to these two categories of drugs indicated that synthetic opioid-involved overdose deaths occurred in the counties where heroin-involved deaths had occurred, but the distribution of synthetic opioid-involved deaths had expanded to include more counties.

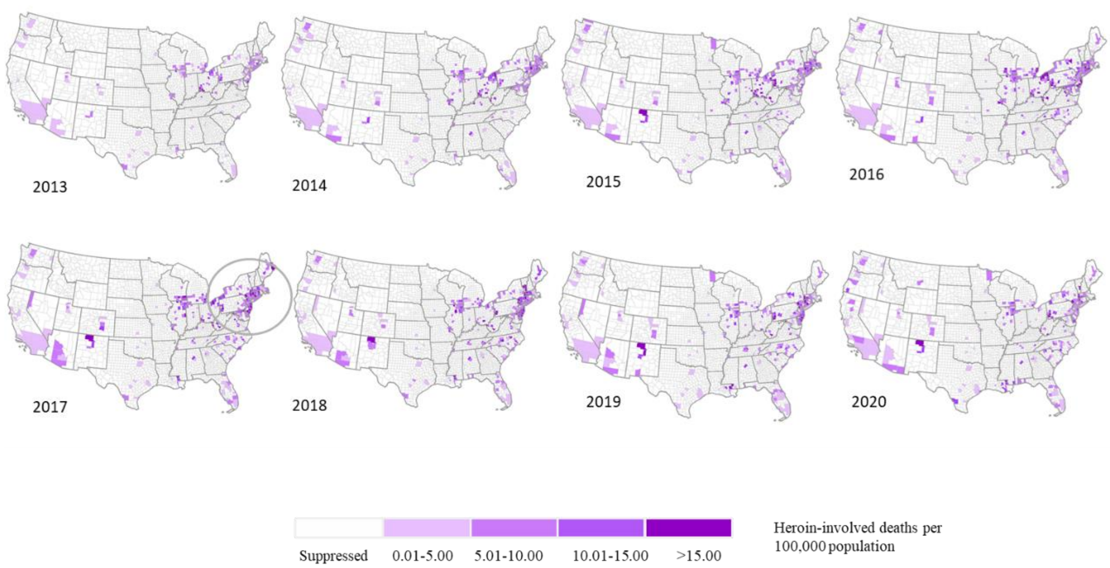

Figure 4.3 Heroin-involved deaths (adjusted by population) in the U.S. aggregated by counties by year for 2013 to 2020

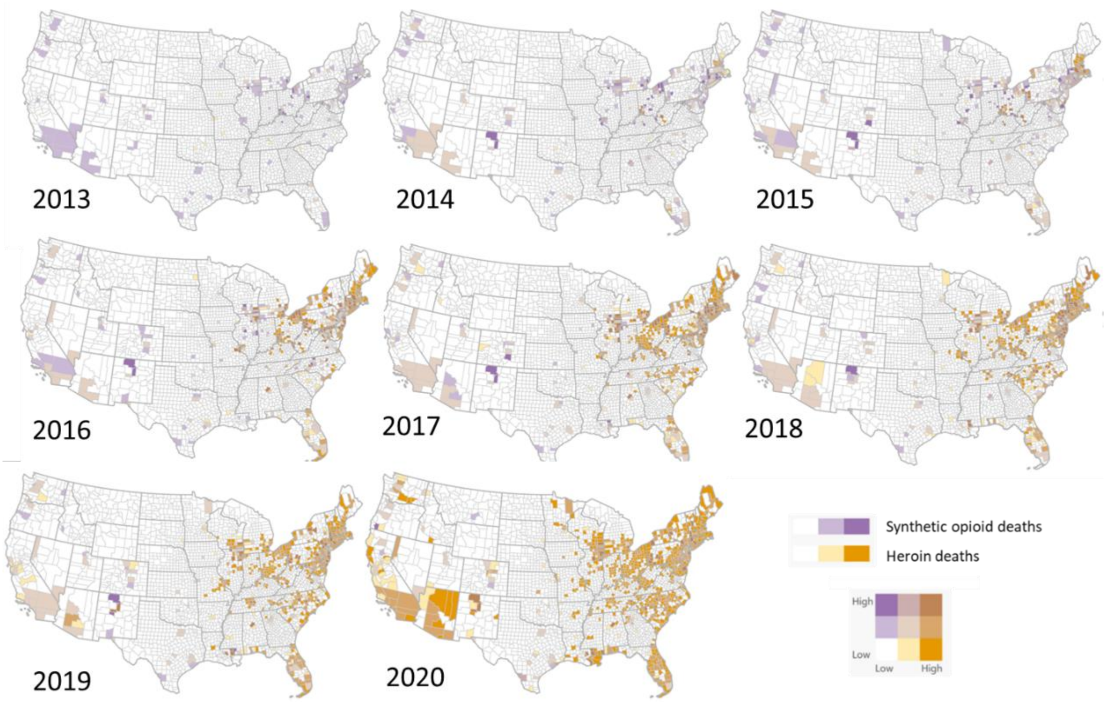

Figure 4.4 Bivariate map of synthetic opioid-involved deaths and heroin-involved deaths (adjusted by population) in the U.S. by year for 2013 to 2020

4.4.2 Time lag between synthetic opioid-involved death and heroin-involved deaths

Time-lagged cross correlation was implemented to analyze the relationship between synthetic opioid- and heroin-involved deaths from 2013 to 2020. I implemented the time-lagged cross correlation at both the nationwide and county level. The national time-lagged cross correlation analyzed the time lag between the monthly time series of synthetic opioid-involved deaths and heroin-involved deaths, where the maximum cross correlation index was 0.951 with a time lag of 16 months. This suggested that the spatiotemporal pattern of heroin-involved deaths (i.e., the presence of deaths in a county) were likely to be followed by synthetic opioid-involved deaths after an estimated 16 months.

This study also implemented a time-lagged cross correlation analysis for each county. Sparse monthly synthetic opioid and heroin-involved mortality data were reported by CDC WONDER due to data confidentiality (i.e., deaths less than 10 within a month in a county were not reported). The yearly time series of synthetic opioid and heroin-involved deaths were used for county-level time-lagged cross correlation analysis. I implemented yearly time-lagged cross correlation analysis for 443 counties with reported time series data for synthetic opioid- and heroin-involved deaths. In 192 counties (i.e., 44.3% out of counties with reported data), the time lag between synthetic opioid-involved deaths and heroin-involved deaths was shorter than one year. The lags were between 1-3 years for 230 counties (i.e., 51.9% out of counties with reported data). In only 21 counties (i.e., 4.7% out of counties with reported data), the lags between synthetic opioid deaths and heroin deaths were longer than four years. The estimated lags between synthetic opioid-involved deaths and heroin-involved deaths were likely to be shorter in high overdose areas including Cook County, IL (lag was within one year), Maricopa County, AZ and Baltimore City, MD (lag was about one year in both these locations), while estimated lags were likely to be longer in other areas including Fairfield County, OH (lag was four years) and Jefferson County, CO (lag was five years).

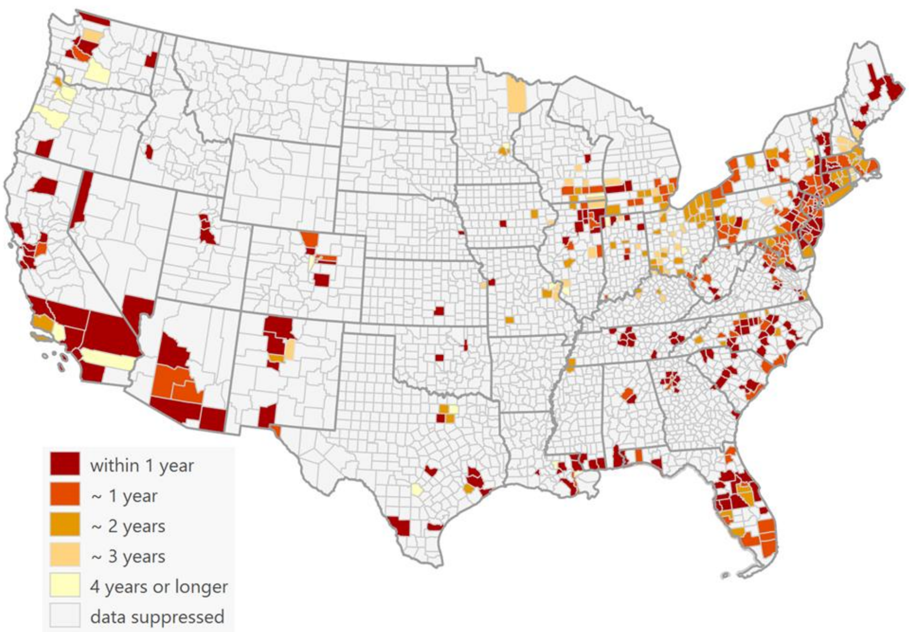

Figure 4.5 Estimated time lags between numbers of synthetic opioid deaths and heroin deaths in the U.S for 2013 to 2020.

4.4.4 Spatial adjacency-based graph and human mobility flow-based graph

Two graphs were generated for structuring the synthetic opioid-involved overdose mortality data at county-level, including a spatial adjacency-based graph and a human mobility flow-based graph. The spatial adjacency-based graph was composed of 3108 nodes (i.e., 3108 U.S. counties), and 18,456 edges connecting neighboring nodes (Figure 4.6). Each edge was assigned an equal weight for the spatial adjacency-based graph (Fotheringham and Rogerson, 2008).

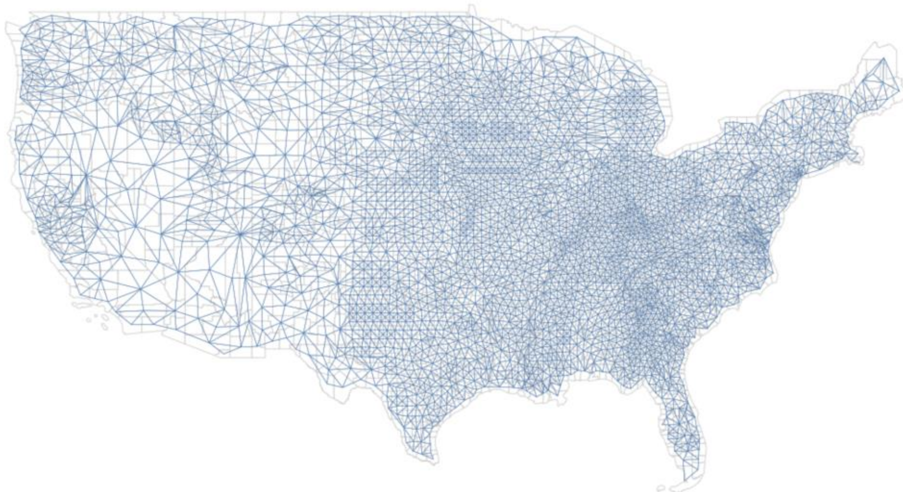

Figure 4.6 Spatial adjacency-based graph data structure for U.S. counties

To generate the human mobility flow-based graph, I aggregated and filtered the human mobility flow dataset (i.e., OD matrix)) with a 75% percentile threshold (i.e., approximately 998 population flow between two counties per year) to extract the major population flow between U.S. counties in 2019. The human mobility flow-based graph contained 3,108 nodes (i.e., 3,108 counties), and 241,876 edges (Figure 4.7). The mobility-based graph have higher number of edges that enable the graph learning neural network model to capture the spatial correlation not only for surrounding counties but also for the farther counties having population flow between the areas.

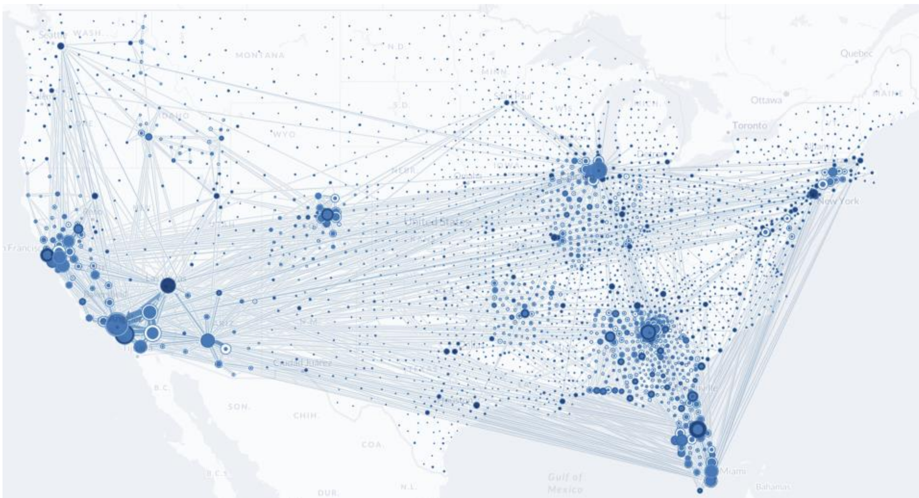

Figure 4.7 Human mobility-based graph data structure for U.S. counties

4.4.5 Estimate opioid-involved deaths using graph convolutional network models

The graph convolutional network (GCN) models were trained using the synthetic opioid-involved mortality data for 2016-2018. The estimated presence of synthetic opioid-involved deaths by county for 2019 computed by the GCNs was compared to the actual presence of synthetic opioid-involved deaths by county that occurred in 2019 (Figure 4.8). The prediction accuracy of the GCN model (i.e., the percentage of counties that were correctly estimated by the model) using the spatial adjacency-based graph was 81.37%. The GCN model using the spatial adjacency-based graph (GCN_1) was able to estimate the approximate cluster locations for synthetic opioid-involved deaths including in Pacific, New England and South Atlantic regions, but GCN1 overestimated in the cluster areas and underestimated the number of estimated synthetic opioid-involved deaths in counties outside of the major clusters.

The prediction accuracy of the GCN model using the mobility-based graph (GCN_2) was 89.86%, which was 9.8% higher than for the GCN model results using the spatial adjacency-based graph. The overall classification accuracy of the random forest model for comparison was 83.56%, which was slightly higher than GCN_1 but 7.5% lower than GCN_2. Comparing to the estimated results by GCNs, the random forest model underestimated the size of major clusters of synthetic opioid-involved deaths in the areas including the pacific regions and northeast regions. The mobility-based graph enabled the graph learning neural network to capture the changing patterns not only from the neighboring areas but also from farer interactive areas, so that GCN_2 can estimate the spreads of the synthetic opioid-involved deaths both inside the major cluster and outside the major clusters.

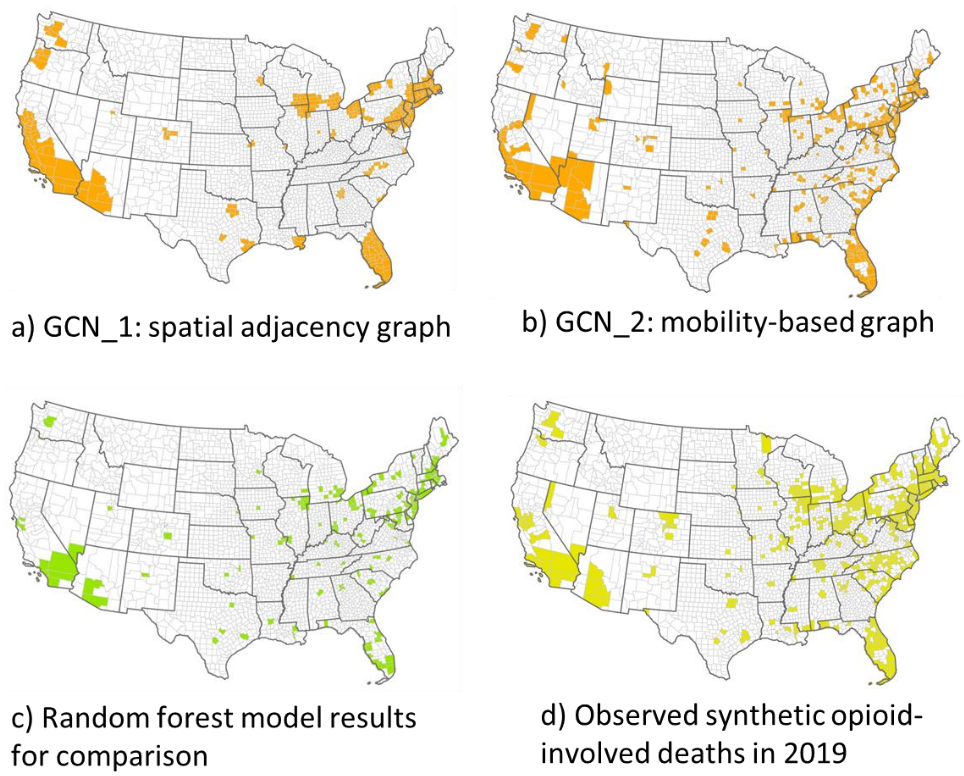

Figure 4.8 a) estimated synthetic opioid-involved deaths for 2019 by GCN using spatial adjacency-based graph; b) estimated synthetic opioid-involved deaths for 2019 by GCN using human mobility-based graph; c) estimated synthetic opioid-involved deaths for 2019 by the random forest model for comparison; d) actual synthetic opioid-involved deaths occurred in the U.S. during 2019

4.5 Discussion

For this research, the spatiotemporal pattern of synthetic opioid-involved deaths in the U.S. at county level for the period 2013 to 2020 was investigated. The synthetic opioid-involved deaths evolved from a scattered pattern in the West Coast of the U.S. in 2013 and increased rapidly from that year. This study found a steep increase in the number synthetic opioid-involved deaths across the U.S. in 2016, and these deaths occurred across the Northeast and mid-Atlantic regions. Previous research on the spatial patterns of heroin overdose deaths in the U.S. found that the heroin-involved death pattern shifted from the west coast in 2020 to New England, the Mid-Atlantic, Great Lakes, and central Ohio Valley areas by 2014 (Hedegaard et al., 2020; Jalal et al., 2018; Stewart et al., 2017). The number of heroin-involved deaths that occurred in the U.S. peaked in 2017. Previous study reported that the number of heroin users arrived peak in 2016 (Substance Abuse and Mental Health Services Administration, 2019). In this chapter, the study found synthetic opioid-involved deaths exceeded heroin-involved deaths in 2016, and synthetic-involved deaths occurred in the regions that used to have high heroin overdose death rates, including New England areas, Middle Atlantic areas, and East North Central areas.

Spatial adjacency relationships have been considered in previous research investigating the geographic patterns of drug overdose deaths. For example, a previous study on rise in drug overdose mortality in the U.S. indicated that drug overdose deaths were likely to be concentrated in large cities that have connections with drug smuggling networks (Ho, 2017). In this chapter, spatial adjacency and human mobility assist in generating graph-structured data accounting for spatial connection between counties. The graph-based data structure for synthetic opioid-involved deaths accounted for the spatial connection between counties and enabled the model to capture the spreading patterns of synthetic opioid-involved deaths over the U.S.

Human mobility has been addressed in previous research on public health issues. Previous research on the 2009 H1N1 outbreak in Mexico considered human mobility and social network information as key elements for modeling the spread of the virus (Frias-Martinez et al., 2011). In this chapter, human mobility was used to generate a graph to represent the spatial interactions between counties. For this study, the mobility data used for this study was originally generated by tracking mobile devices which might underestimate the actual trips between counties.

For this study, synthetic opioid- and heroin-involved overdose mortality data were accessed from Multiple Cause of Death Data provided by Centers for Disease Control Wide-ranging Online Data for Epidemiologic Research (CDC WONDER). Due to the data confidentiality, CDC WONDER suppressed the data with a number of deaths fewer than ten for each spatial unit during the data query (Centers for Disease Control and Prevention, 2022b; Quick, 2019). This study may underestimate possible synthetic opioid- and heroin-involved deaths in these counties and miss the spread of synthetic opioid-involved deaths over these areas.

4.6 Conclusion

This study analyzed the relationship between the spread of synthetic opioids and another significant opioid, heroin, and compared the patterns of synthetic opioid-involved deaths with heroin-involved deaths over space and time. The nationwide analysis identified that the estimated average lag time between the pattern of synthetic opioid-involved deaths and heroin-involved deaths was approximately 16 months. This study incorporated both human mobility and spatial adjacency into the graph learning model to analyze the spatial spread of opioid-involved deaths. The human mobility-based graph enabled the graph learning model to capture the spatial patterns not just for neighboring areas but for areas that have connections between other counties. The potential impact of drug markets on the dynamics of the synthetic opioid crisis could also be analyzed in future research. This study highlighted a modeling approach applying graph deep neural networks in spatiotemporal analysis of synthetic opioid-involved deaths that can be further applied for geo-health research.

Chapter 5: Conclusions

5.1 Summary of findings

This dissertation aims at tracking the dynamics of the opioid crisis in the United States over space and time. The dissertation is composed of three studies that investigate the opioid crisis in the United States at multiple scales including nationwide analysis at county-level, and local analysis at the ZIP code and block group-level, for multiple time periods.

In the first study (chapter 2), space and time, together with multiple driving factors were incorporated into a random forest model to analyze changing patterns of repeated events (i.e., drug activities) in a major metropolitan area. We applied this space-time analysis framework to investigate the spatiotemporal patterns of more than 52,000 reported incidents of drug-related crime in Chicago, IL between 2016 and 2019. Advanced geospatial data analytic methods including machine learning, big data and time series analysis are used to analyze the changing patterns and understand how a set of different factors (e.g., built environment and sociodemographic factors) are related to these changing patterns. We focus especially on drug crime locations involving two drug types, heroin and synthetic drugs, that have been key drugs driving the opioid crisis in the United States.

The second study (chapter 3) focused on an analysis of opioid-involved mortality in the context of the COVID-19 pandemic in the U.S. from a geospatial perspective. For this study, we investigated spatiotemporal patterns of opioid-involved deaths that occurred in the U.S. during 2020 and compared the spatiotemporal pattern of these deaths with patterns for the previous three years (2017-2019) to better understand changes in the number of opioid-involved deaths during 2020 when the impact of the COVID-19 pandemic was especially strong. A counterfactual analysis framework together with a space-time random forest (STRF) model were used to estimate the increase in opioid-involved deaths related at least in part to circumstances during the pandemic. To gain further insight into the relationship between opioid deaths and COVID-19-related factors, I also built a local STRF model for the City of Chicago, a metropolitan area that experienced a steep increase in opioid-related deaths during 2020. High ranking indicators identified by the model were used to generate a vulnerability index of opioid overdose occurrence during the pandemic in Chicago to provide further insights for use by drug health stakeholders and practitioners.

The third study (chapter 4) focused on the dynamics of the synthetic opioid epidemic in the U.S. from 2013 when the number of synthetic opioid-involved deaths began to rise across the U.S. The spatiotemporal patterns of synthetic opioid-involved deaths between 2013 and 2020 were analyzed, and compared to the patterns of heroin-involved deaths for the same period. Graph neural network models were implemented to track the spread of synthetic opioid-involved deaths over space and time. This research also presented two novel ways to convert the presence of synthetic opioid-involved deaths at county level into graphs using a spatial adjacency-based graph and a human mobility flow-based graph. The two graphs accounted for the spatial connections between counties and enabled the model to capture the spread of synthetic opioid-involved deaths across the nation.

The major contributions of this dissertation

Contribution 1: Chapter 2 identified different factors including built environment such as vacant lots, vacant buildings and alley density and sociodemographic factors including education level and percentage of low wage workers that were most associated with the heroin and synthetic drug patterns, that are useful for understanding the contributing factors to the dynamics of the opioid crisis as well as the possible underlying drivers for where drug activities occur.

Contribution 2: Space and time are both important and were incorporated by accommodating spatiotemporal autocorrelation variables (lag variables and trend variables) into the machine learning model to track the changing patterns of drug activities in chapter 2. Previous (non-geospatial) machine learning models often do not take spatiotemporal autocorrelation or space-time factors into account.

Contribution 3: Chapter 2 also investigated the spatiotemporal patterns for heroin and synthetic drugs during the opioid crisis in a U.S. city, Chicago. This study found that while heroin-related crime incidents dropped slightly in 2017, they rose again in 2018, and continually increased in 2019, staying mostly in the same locations.

The pattern of synthetic drugs evolved from a scattered pattern in 2016 to two distinct hotspot areas in 2019. These hotspots were in the locations of 2016 and 2017 heroin-related hotspots.

Contribution 4: Chapter 3 investigated the spatiotemporal patterns of opioid-related deaths in U.S. counties during 2020 and compared the patterns with pre-COVID-19 pandemic time (2017-2019). The largest increase in opioid-involved deaths in 2020 occurred in the Pacific [28] (89.7%), East South Central [29] (57.9%) and Mountain[30] (52.9%) regions. The counterfactual analysis indicated that only 77% of opioid-involved deaths in 2020 can be explained by demographic, economic and built environment factors and the increasing trend of the opioid-involved deaths during the opioid crisis, and 23% of opioid-involved deaths in 2020 were likely can be attributed to the COVID-19 pandemic.

Contribution 5: Chapter 3 generated a ZIP-code level vulnerability index of opioid overdose occurrence in the context of the pandemic for a U.S. metropolitan area. The index showed that in 2020, 56% of opioid deaths and 61% of the increase in opioid deaths in Chicago (compared to the average number of deaths in 2017-2019) occurred around the West Side and Far Southeast Side areas of the city (top one class out of five classes) and indeed the top two classes representing 76% of opioid deaths and 85% of the increase in deaths were in a broader extent of these same areas in the City. This

[28] Pacific Division: California, Washington, Oregon, Alaska and Hawaii
[29] South Central Division: Alabama, Kentucky, Mississippi and Tennessee
[30] Mountain Division: Arizona, Colorado, Idaho, New Mexico, Montana, Utah Nevada and Wyoming

analysis assists in evaluating the resilience against an opioid crisis during the pandemic period and provides further insights for drug health stakeholders and practitioners.

Contribution 6: Chapter 4 analyzed the relationship between the dynamics of the synthetic opioid epidemic and spatial patterns of heroin overdose deaths. This study compared the patterns of synthetic opioid-involved deaths and heroin-involved deaths across the U.S. between 2013 and 2020. The average lag between the patterns for synthetic opioid-involved deaths and heroin-involved deaths was found to be approximately 16 months, i.e., the spatiotemporal pattern of heroin-involved deaths was likely to be followed by synthetic opioid-involved deaths after an estimated 16 months.

Contribution 7: Chapter 4 came up with an analysis frame work to incorporate human mobility and spatial adjacency into a graph neural network model to capture the spread of synthetic opioid-involved deaths, and the changing patterns of the presence of synthetic opioid deaths in U.S. counties. The mobility-based graph enabled the model to learn the changing patterns not only from the surrounding locations but also from the locations that have mobility flows between other counties.

5.2 Future work

To achieve the research objective, this dissertation combined spatial and temporal data science methods including hotspot analysis, time series analysis, counterfactual analysis, time-lagged cross correlation analysis, statistical learning models and deep learning models to understand, analyze and track the dynamics of the opioid crisis over space and time. Future research could consider incorporating spatiotemporal heterogeneity into statistical learning models (e.g., a random forest model) by setting a spatiotemporal bandwidth to investigate the changing patterns at varying scales. For example, for analyzing spatiotemporal patterns for location *A*, the most appropriate spatial bandwidth could be neighboring areas within ten kilometers and the best temporal bandwidth (i.e., lag) could be two years. This spatiotemporal bandwidth is likely to be different for another location *B*. In Chapter 2, space and time were incorporated into the random forest model. Multi-scale spatiotemporal analysis could also be considered in future research. The machine learning model for state-, county-, and block group-level analysis could be implemented, and the driving factors for the ranking of the important factors identified by a multi-scale model could vary from the single-scale analysis.

This dissertation identified spatial and temporal patterns of heroin and synthetic drug patterns in a U.S. metropolitan area (Chicago, IL) and detected clusters of opioid misuse in this area over time. These results are useful for providing fundamental information helpful for decision-making related to the opioid crisis, for example, assisting with assessments of the availability of substance use treatment facilities to ensure that treatment is available in locations of higher drug use. Future research could investigate how to extend and generalize the space-time analysis to other urban and suburban areas adjusted to different city structures. Future research could also consider developing the space-time models of opioid-involved deaths over time into an alert system that can learn the patterns of existing opioids and track the dynamics of the opioid crisis. This could help mitigate the spread of new drugs and provide decision support during the opioid crisis.

Appendices

Appendix A. Supplementary materials for Chapter 2

Table A.1 Summary of factors established to be linked to drug activities in literature and related candidate variables collected in this study

Factors	Related measures	Related variables collected in this study	References
Income inequality	Low income, medium income, and high income	Percentage of low-wage workers, percentage of individuals in poverty status, percentage of medium -wage workers, percentage of high-wage workers	Dasgupta et al., 2017; Lipton et al., 2013; Yarbrough et al., 2019
Low education level	College education	Percentage of population with college-level degree or higher	Nechuta et al., 2018
High unemployment rate	Employment status	Total number of workers, population at working age, employment, unemployment, gross employment density, full-time employment, part-time employment	Cooper et al., 2016; Lipton et al., 2013; McCord and Ratcliffe, 2007; Yarbrough et al., 2019
	Employment by category	Service jobs, industrial jobs, entertainment jobs, retail jobs, office jobs, education jobs, health care jobs, public administration jobs	
Demographics	Population	Total population, residential density, population density	Chaney and Rojas-Guyler, 2015; Iyiewuare et al., 2017; Lipton et al., 2013; Marotta et al., 2019; Marshall et al., 2017; Reisner et al., 2015; Vilalta, 2010
	Household status	Total households, median household income, jobs per household, household workers per job percentage of household in poverty status, number of households own two or more automobiles, number of households own one automobile, number of households own zero automobile	
	Gender	Percentage of female population, percentage of male population	
	Race and ethnicity	Percentage of white Americans, percentage of African Americans, percentage of Asian Americans, Percentage of Hispanic/Latino Americans	
Community environment	Road network	Total road network density, street intersection density, network density of auto-oriented links, network density of pedestrian-oriented links, network density of multi-model links, density of auto-oriented intersections, density of	Cerdá et al., 2013; Marshall et al., 2017; Srinivasan et al., 2003; Sutter et al., 2019;

		pedestrian-oriented intersections, density of multi-model intersections	Weisburd and Mazerolle, 2000
	Land use factors	Walkability index, total land area, total water area, total area, protected conservation area, area of unprotected land	
Community characteristics	Occupancy	Percentage of house occupancy, number of vacant buildings, number of vacant lots	Cerdá et al., 2017, 2013; Darke et al., 2001; Lipton et al., 2013; Martins et al., 2015; McCord and Ratcliffe, 2007; Moore et al., 2018; Sutter et al., 2019; Visconti et al., 2015
	House ownership	Percentage of house ownership, percentage of rental properties	
	Activity density and diversity	Gross house unites and employment density, gross retail employment density, gross office employment density, gross industrial employment density, gross service employment density, gross entertainment employment density, gross education employment density, gross health care employment density, gross public administration employment density, employment entropy, employment and household entropy, commute trip productions and trip attractions equilibrium index, regional diversity	
Key locations	Frequent locations for drug-related crimes	Alley density, number of vacant lots, number of vacant buildings, number of gas stations, number of parking lots, total area of park properties, number of high schools	Key locations were discussed in the Data section
Note: some variable names listed in the table refer to a group of variables (e.g., intersection density referring to intersection density of all intersections, three-leg intersection density and intersection density of intersections having four or more legs).			

Table A.2 List of variables in the random forest model for predicting heroin-related activity hot zones

Variable name	Category	Description
GasStation	Built environment	Number of gas stations in each U.S. census block group
VacLot	Built environment	Number of vacant lots in each U.S. census block group
VacBldg	Built environment	Number of vacant buildings in each U.S. census block group
ParkingLot	Built environment	Number of parking lots in each U.S. census block group
AlleyDen	Built environment	Alley density
ParkArea	Built environment	Area of park propertied in each U.S. census block group
HighSch	Built environment	Number of high schools in each U.S. census block group
POccupyHouse	Built environment	% of house occupancy
POwner	Built environment	% of house ownership
PRenter	Built environment	% of house rent
D3a	Built environment	Total road network density
D3amm	Built environment	Network density of auto and pedestrian links
D3apo	Built environment	Network density of pedestrian links

D3b	Built environment	Street intersection density
D3bpo3	Built environment	Pedestrian-oriented intersection (three legs) density
D1A	Built environment	Gross residential density
D1C	Built environment	Gross employment density
D1C8_Off10	Built environment	Gross office employment density
D1D	Built environment	Gross activity density (employment and house units)
D2R_JOBPOP	Built environment	Regional Diversity (ratio of jobs/population)
TotPop	Sociodemographic	Total population in each U.S. census block group
TotHH	Sociodemographic	Total households in each U.S. census block group
PMale	Sociodemographic	Percentage of male population
PFemale	Sociodemographic	Percentage of female population
PWhite	Sociodemographic	% of population selecting race as white American alone
PBlack	Sociodemographic	% of population selecting race as black/African American alone
PAsian	Sociodemographic	% of population selecting race as Asian American alone
Phispanic	Sociodemographic	% of population selecting race as Hispanic/Latino American alone
PBachelorHigher	Sociodemographic	% of population with college level degree
MedianHHIncome	Sociodemographic	Median household income
PPovertyHouse	Sociodemographic	% of household in poverty status
PPovertyIndv	Sociodemographic	% of population in poverty status
PEmploy	Sociodemographic	% of full time and part time employees
Punemploy	Sociodemographic	% of unemployment
PFulltimeEmploy	Sociodemographic	% of full-time employment among all employment
PParttimeEmploy	Sociodemographic	% of part-time employment among all employment
P_WRKAGE	Sociodemographic	% of population that is working aged
PCT_AO0	Sociodemographic	% of zero-car households
PCT_AO1	Sociodemographic	% of one-car households
AUTOOWN2P	Sociodemographic	Number of households that own two or more automobiles
PCT_AO2P	Sociodemographic	% of two-plus-car households
WORKERS	Sociodemographic	Number of workers
R_HIWAGEWK	Sociodemographic	Number of high wage workers
R_PCTLOWWA	Sociodemographic	% low wage workers
E5_IND10	Sociodemographic	Industrial jobs
E8_SVC10	Sociodemographic	Service jobs
HS_t_1	Spatial and temporal lag variables	Time-lagged variable (whether this BG belonged to a hotspot at t-1 period)
nb_t_1	Spatial and temporal lag variables	Spatiotemporally lagged variable (number of neighboring BGs belonged to a hotspot at t-1 period)
trend	Spatial and temporal lag variables	Trend variable (increase or decrease in the number of neighboring block groups that belonged to a hotspot at t-1 period)

Table A.3 List of variables in the random forest model for predicting synthetic drug-related activity hot zones

Variable name	Category	Description
GasStation	Built environment	Number of gas stations in each U.S. census block group
VacLot	Built environment	Number of vacant lots in each U.S. census block group
VacBldg	Built environment	Number of vacant buildings in each U.S. census block group
ParkingLot	Built environment	Number of parking lots in each U.S. census block group
AlleyDen	Built environment	Alley density
ParkArea	Built environment	Area of park propertied in each U.S. census block group
HighSch	Built environment	Number of high schools in each U.S. census block group
POccupyHouse	Built environment	% of house occupancy
POwner	Built environment	% of house ownership
PRenter	Built environment	% of house rent
AC_UNPR	Built environment	Total land area (not include park or conservation area)
D3b	Built environment	Street intersection density
D3bmm3	Built environment	Auto- and pedestrian-oriented intersection (three legs) density
D1A	Built environment	Gross residential density
D1C5_Ret10	Built environment	Gross retail employment density
D1D	Built environment	Gross activity density (employment and house units)
D2A_EPHHM	Built environment	Employment and household entropy
TotPop	Sociodemographic	Total population in each U.S. census block group
TotHH	Sociodemographic	Total households in each U.S. census block group
PMale	Sociodemographic	Percentage of male population
PFemale	Sociodemographic	Percentage of female
PWhite	Sociodemographic	% of population selecting race as white American alone
PBlack	Sociodemographic	% of population selecting race as black/African American alone
Pasian	Sociodemographic	% of population selecting race as Asian American alone
Phispanic	Sociodemographic	% of population selecting race as Hispanic/Latino American alone
PBachelorHigher	Sociodemographic	% of population with college level degree
MedianHHIncome	Sociodemographic	Median household income
PPovertyHouse	Sociodemographic	% of household in poverty status
PPovertyIndv	Sociodemographic	% of population in poverty status
PEmploy	Sociodemographic	% of full time and part time employees
Punemploy	Sociodemographic	% of unemployment
PFulltimeEmploy	Sociodemographic	% of full-time employment among all employment
PParttimeEmploy	Sociodemographic	% of part-time employment among all employment
P_WRKAGE	Sociodemographic	% of population that is working aged
PCT_AO0	Sociodemographic	% of zero-car households
PCT_AO2P	Sociodemographic	% of two-plus-car households
R_HIWAGEWK	Sociodemographic	Number of high wage workers
R_PCTLOWWA	Sociodemographic	% low wage workers
E8_RET10	Sociodemographic	Retail jobs
E8_OFF10	Sociodemographic	Office jobs
HS_t_1	Spatial and temporal lag variables	Time-lagged variable (whether this BG belonged to a hotspot at t-1 period)
nb_t_1	Spatial and temporal lag variables	Spatiotemporally lagged variable (number of neighboring BGs belonged to a hotspot at t-1 period)

trend	Spatial and temporal lag variables	Trend variable (increase or decrease in the number of neighboring block groups that belonged to a hotspot at t-1 period)

Table A.4 Classification confusion matrixes, accuracy for predicting hotspot and non-hotspot classes and overall prediction accuracy (ACC)

Drug type	**Heroin**					**Synthetic drug**				
Model 1 (sociodemographic)	pred obs	hotspot	other	class accuracy	ACC	pred obs	hotspot	other	class accuracy	ACC
	hotspot	106	12	89.8%	97.7%	hotspot	96	56	63.2%	90.5%
	other	38	2046	98.2%		other	154	1896	92.5%	
Model 2 (built env)		hotspot	other				hotspot	other		
	hotspot	106	12	89.8%	97.7%	hotspot	96	56	63.2%	90.5%
	other	38	2046	98.2%		other	154	1896	92.5%	
Model 3 (sociodemographic + built env)		hotspot	other				hotspot	other		
	hotspot	106	12	89.8%	97.7%	hotspot	96	56	63.2%	90.5%
	other	38	2046	98.2%		other	154	1896	92.5%	
Model 4 (sociodemographic + spatiotemporal autocorrelation)		hotspot	other				hotspot	other		
	hotspot	92	26	78.0%	98.4%	hotspot	93	59	61.2%	90.5%
	other	10	2074	99.5%		other	151	1899	92.6%	
Model 5 (built env + spatiotemporal autocorrelation)		hotspot	other				hotspot	other		
	hotspot	88	30	74.6%	98.2%	hotspot	97	55	63.8%	89.1%
	other	10	2074	99.5%		other	186	1864	90.9%	
Model 6 (spatiotemporal autocorrelation)		hotspot	other				hotspot	other		
	hotspot	87	31	73.7%	98.3%	hotspot	111	41	73.0%	87.5%
	other	7	2077	99.7%		other	234	1816	88.6%	
Model 7 (sociodemographic + built env + spatiotemporal autocorrelation)		hotspot	other				hotspot	other		
	hotspot	96	22	81.4%	98.3%	hotspot	91	61	60.0%	90.7%
	other	15	2069	99.3%		other	144	1906	93.0%	

Appendix B. Supplementary materials for Chapter 3

Table B.1 Variables for analysis in this study

Factor type	**Variable derived from multiple data sources**
Demographic	
Population	Total population
Age	Percentage of population gor age groups (0-17, 18-24, 25-34, 35-44, 45-54, 55-64 and over 65)
Gender	Percentage of female, percentage of male

Race	Percentage of white, black, Asian, Hispanic
Education	Low education percentage (percentage of population 25 years and over without schooling completed), high education percentage (percentage of population 25 years and over with college degree or higher)
Economic	
Unemployment	Unemployment rate for 2017, 2018, 2019 and 2020
Change in unemployment	Change in unemployment rate 2019-2020, 2018-2019, 2017-2018, 2016-2017
Income	Average personal annual income at county-level, median household income
Change in personal income	Change of personal income in 2020 (comparing to 2019), 2019, 2018, 2017
Poverty	Percentage of households at poverty level, percentage of individuals at poverty level
Property	Median house value, median rent
Built environment	
Property ownership and occupancy	Percentage of occupation, percentage of vacant houses, percentage of rental units, percentage of ownership
Pandemic-related variables	
COVID-19 positive cases	Monthly COVID-19 positives cases at county level
COVID-19 deaths	Monthly COVID-19 deaths at county level
Social distancing metrics for periods relating to COVID-19	Median home-dwelling time, percentage of completely home device, percentage of work behavior device, median percentage of home time for periods relating to COVID-19
Change in human mobility patterns from 2019 to 2020	Change in median home-dwelling time, change in percentage of completely home device, change in percentage of work behavior device, change in median percentage of home time for periods relating to COVID-19
Spatiotemporal autocorrelation	
Time lag variables	The number of opioid deaths in a county in the previous year for the entire year and periods relating to COVID-19
Spatiotemporal lag variables	The number of opioid deaths that occurred in neighboring counties in the previous year for the entire year and periods relating to COVID-19
Trend variables	Changes in the number of opioid deaths in year t-1 (the previous year) compared to year t-2 in a county (yearly and period-based).

Table B.2 STRF_COVID model performance

	Split no. used for independent test	STRF_COVID1 (using yearly and period-based spatiotemporal lag variables)	STRF_COVID2 (using yearly spatiotemporal lag variables)	STRF_COVID3 (using period-based spatiotemporal lag variables)
Out-of-bag (OOB) R-squared	1	0.892	0.832	0.878
	2	0.868	0.802	0.854
	3	0.900	0.829	0.889
	4	0.900	0.829	0.889
	5	0.878	0.812	0.860
Average R-squared		0.888	0.821	0.874
% opioid-involved deaths explained by the model		88.6%	87.1%	86.1%

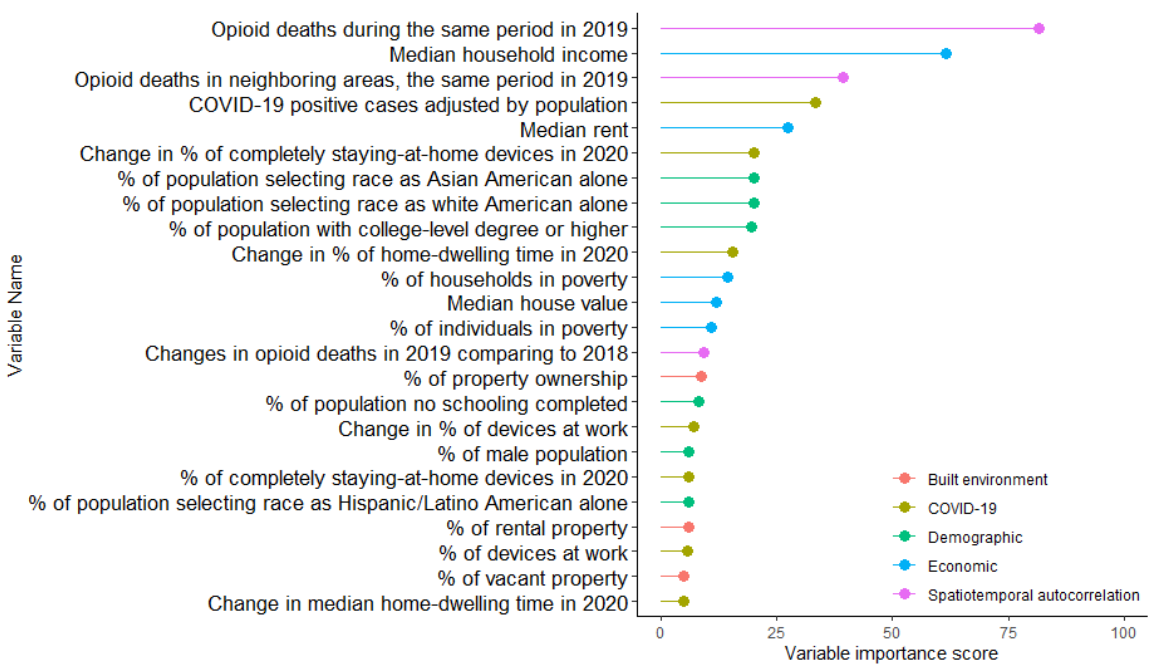

Figure B.1 Important variables for estimating opioid-involved deaths during opioid death accelerating peak time

Bibliography

Abraham, A.J., Andrews, C.M., Yingling, M.E., Shannon, J., 2018. Geographic Disparities in Availability of Opioid Use Disorder Treatment for Medicaid Enrollees. Health Serv. Res. 53, 389–404. https://doi.org/10.1111/1475-6773.12686

Albright, D.L., McDaniel, J., Kertesz, S., Seal, D., Prather, K., English, T., Laha-Walsh, K., 2019. Small area estimation and hotspot identification of opioid use disorder among military veterans living in the Southern United States. Subst. Abuse 0, 1–7. https://doi.org/10.1080/08897077.2019.1703066

Amaratunga, D., Cabrera, J., Lee, Y.-S., 2008. Enriched random forests. Bioinformatics 24, 2010–2014. https://doi.org/10.1093/bioinformatics/btn356

American Medical Association, 2020. Issue brief: Reports of increases in opioidrelated overdose and other concerns during COVID pandemic.

Ancuceanu, R., Dinu, M., Neaga, I., Laszlo, F.G., Boda, D., 2019a. Development of QSAR machine learning-based models to forecast the effect of substances on malignant melanoma cells. Oncol. Lett. 17, 4188–4196. https://doi.org/10.3892/ol.2019.10068

Ancuceanu, R., Dinu, M., Neaga, I., Laszlo, F.G., Boda, D., 2019b. Development of QSAR machine learning-based models to forecast the effect of substances on malignant melanoma cells. Oncol. Lett. 17, 4188–4196. https://doi.org/10.3892/ol.2019.10068

Anselin, L., 1995. Local Indicators of Spatial Association—LISA. Geogr. Anal. 27, 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Becker, W.C., Fiellin, D.A., 2020. When Epidemics Collide: Coronavirus Disease 2019 (COVID-19) and the Opioid Crisis. Ann. Intern. Med. 173, 59–60. https://doi.org/10.7326/M20-1210

Bergquist, S., Otten, T., Sarich, N., 2020. COVID-19 pandemic in the United States. Health Policy Technol. 9, 623–638. https://doi.org/10.1016/j.hlpt.2020.08.007

Bernard, Heutte, Adam, 2009. Influence of Hyperparameters on Random Forest Accuracy, in: Influence of Hyperparameters on Random Forest Accuracy.

Biau, G., Scornet, E., 2015. A Random Forest Guided Tour. ArXiv151105741 Math Stat.

Bivand, R., Pebesma, E.J., Gómez-Rubio, V., 2008. Applied spatial data analysis with R, Use R! Springer, New York.

Borges, J., Ziehr, D., Beigl, M., Cacho, N., Martins, A., Sudrich, S., Abt, S., Frey, P., Knapp, T., Etter, M., Popp, J., 2017. Feature engineering for crime hotspot detection, in: 2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). Presented at the 2017 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced

Trusted Computed, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), pp. 1–8. https://doi.org/10.1109/UIC-ATC.2017.8397586

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., 1996. Out-Of-Bag Estimation.

Bui, K.-H.N., Cho, J., Yi, H., 2022. Spatial-temporal graph neural network for traffic forecasting: An overview and open research issues. Appl. Intell. 52, 2763–2774. https://doi.org/10.1007/s10489-021-02587-w

Bustamante, G., Guzman, V., Kobayashi, L.C., Finlay, J., 2022. Mental health and well-being in times of COVID-19: A mixed-methods study of the role of neighborhood parks, outdoor spaces, and nature among US older adults. Health Place 76, 102813. https://doi.org/10.1016/j.healthplace.2022.102813

Calle, M.L., Urrea, V., 2011. Letter to the Editor: Stability of Random Forest importance measures. Brief. Bioinform. 12, 86–89. https://doi.org/10.1093/bib/bbq011

Cao, Y., Stewart, K., Factor, J., Billing, A., Massey, E., Artigiani, E., Wagner, M., Dezman, Z., Wish, E., 2020. Using socially-sensed data to infer ZIP level characteristics for the spatiotemporal analysis of drug-related health problems in Maryland. Health Place 63, 102345. https://doi.org/10.1016/j.healthplace.2020.102345

Cao, Y., Stewart, K., Wish, E., Artigiani, E., Sorg, M.H., 2019. Determining spatial access to opioid use disorder treatment and emergency medical services in New Hampshire. J. Subst. Abuse Treat. 101, 55–66. https://doi.org/10.1016/j.jsat.2019.03.013

CDC, 2021. ICD - ICD-10 - International Classification of Diseases, Tenth Revision [WWW Document]. URL https://www.cdc.gov/nchs/icd/icd10.htm (accessed 4.26.22).

CDC, 2020. Multiple Cause of Death 1999-2019 [WWW Document]. URL https://wonder.cdc.gov/wonder/help/mcd.html (accessed 1.2.21).

Center for Behavioral Health Statistics and Quality, 2015. Behavioral Health Trends in the United States: Results from the 2014 National Survey on Drug Use and Health 64.

Center for Disease Control and Prevention, 2019a. Fentanyl | Drug Overdose | CDC Injury Center [WWW Document]. URL https://www.cdc.gov/drugoverdose/opioids/fentanyl.html (accessed 8.30.19).

Center for Disease Control and Prevention, 2019b. Synthetic Opioid Overdose Data | Drug Overdose | CDC Injury Center [WWW Document]. URL https://www.cdc.gov/drugoverdose/data/fentanyl.html (accessed 8.30.19).

Centers for Disease Control and Prevention, 2022a. Synthetic Opioid Overdose Data | Drug Overdose | CDC Injury Center [WWW Document]. URL

https://www.cdc.gov/drugoverdose/deaths/synthetic/index.html (accessed 10.9.22).
Centers for Disease Control and Prevention, 2022b. General Help for CDC WONDER [WWW Document]. URL https://wonder.cdc.gov/wonder/help/main.html#What%20is%20WONDER (accessed 5.31.22).
Centers for Disease Control and Prevention, 2020a. Synthetic Opioid Overdose Data | Drug Overdose | CDC Injury Center [WWW Document]. URL https://www.cdc.gov/drugoverdose/data/fentanyl.html (accessed 1.1.21).
Centers for Disease Control and Prevention, 2020b. Understanding the Epidemic | Drug Overdose | CDC Injury Center [WWW Document]. URL https://www.cdc.gov/drugoverdose/epidemic/index.html (accessed 8.31.20).
Cerdá, M., Gaidus, A., Keyes, 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, 103–112. https://doi.org/10.1111/add.13543
Cerdá, M., Ransome, Y., Keyes, K.M., Koenen, K.C., Tardiff, K., Vlahov, D., Galea, S., 2013a. Revisiting the Role of the Urban Environment in Substance Use: The Case of Analgesic Overdose Fatalities. Am. J. Public Health 103, 2252–2260. https://doi.org/10.2105/AJPH.2013.301347
Cerdá, M., Ransome, Y., Keyes, K.M., Koenen, K.C., Tardiff, K., Vlahov, D., Galea, S., 2013b. Revisiting the Role of the Urban Environment in Substance Use: The Case of Analgesic Overdose Fatalities. Am. J. Public Health 103, 2252–2260. https://doi.org/10.2105/AJPH.2013.301347
Chainey, S., Tompson, L., Uhlig, S., 2008. The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. Secur. J. 21, 4–28. https://doi.org/10.1057/palgrave.sj.8350066
Chaney, R.A., Rojas-Guyler, L., 2015a. Spatial patterns of adolescent drug use. Appl. Geogr. 56, 71–82. https://doi.org/10.1016/j.apgeog.2014.11.002
Chaney, R.A., Rojas-Guyler, L., 2015b. Spatial patterns of adolescent drug use. Appl. Geogr. 56, 71–82. https://doi.org/10.1016/j.apgeog.2014.11.002
Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z., Ma, J., 2017. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. CATENA 151, 147–160. https://doi.org/10.1016/j.catena.2016.11.032
Chen, X., Zou, Y., Gao, H., 2021. Role of neighborhood social support in stress coping and psychological wellbeing during the COVID-19 pandemic: Evidence from Hubei, China. Health Place 69, 102532. https://doi.org/10.1016/j.healthplace.2021.102532
Chi, G., Zhu, J., 2008. Spatial Regression Models for Demographic Analysis. Popul. Res. Policy Rev. 27, 17–42. https://doi.org/10.1007/s11113-007-9051-8

Ciccarone, D., 2017. Fentanyl in the US heroin supply: A rapidly changing risk environment. Int. J. Drug Policy 46, 107–111. https://doi.org/10.1016/j.drugpo.2017.06.010

Collins, J., Hall, N., Paul, L.A., 2004. Causation and Counterfactuals. MIT Press.

Cooper, H.L.F., West, B., Linton, S., Hunter-Jones, J., Zlotorzynska, M., Stall, R., Wolfe, M.E., Williams, L., Hall, H.I., Cleland, C., Tempalski, B., Friedman, S.R., 2016. Contextual Predictors of Injection Drug Use Among Black Adolescents and Adults in US Metropolitan Areas, 1993–2007. Am. J. Public Health 106, 517–526. https://doi.org/10.2105/AJPH.2015.302911

Cryer, J.D., Chan, K.-S., 2008. Time Series Analysis: With Applications in R, 2nd ed, Springer Texts in Statistics. Springer-Verlag, New York. https://doi.org/10.1007/978-0-387-75959-3

Darke, S., Kaye, S., Ross, J., 2001. Geographical injecting locations among injecting drug users in Sydney, Australia. Addiction 96, 241–246. https://doi.org/10.1046/j.1360-0443.2001.9622416.x

Dasgupta, N., Beletsky, L., Ciccarone, D., 2017. Opioid Crisis: No Easy Fix to Its Social and Economic Determinants. Am. J. Public Health 108, 182–186. https://doi.org/10.2105/AJPH.2017.304187

Davis, C.S., Samuels, E.A., 2020. Opioid Policy Changes During the COVID-19 Pandemic - and Beyond. J. Addict. Med. 10.1097/ADM.0000000000000679. https://doi.org/10.1097/ADM.0000000000000679

DEA Intelligence Report, 2020. Fentanyl Flow in the United States [WWW Document]. URL https://www.dea.gov/sites/default/files/2020-03/DEA_GOV_DIR-008-20%20Fentanyl%20Flow%20in%20the%20United%20States_0.pdf (accessed 10.9.22).

Deng, H., Runger, G., 2013. Gene selection with guided regularized random forest. Pattern Recognit. 46, 3483–3489. https://doi.org/10.1016/j.patcog.2013.05.018

Deng, H., Runger, G., 2012. Feature selection via regularized trees, in: The 2012 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. https://doi.org/10.1109/IJCNN.2012.6252640

Diggle, P.J., 2013. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition, 3rd ed. Chapman and Hall/CRC. https://doi.org/10.1201/b15326

Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. 20, 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

Drug Enforcement Administration, 2020. Drug Fact Sheet: Synthetic Opioids [WWW Document]. URL https://www.dea.gov/sites/default/files/2020-06/Synthetic%20Opioids-2020.pdf (accessed 10.9.22).

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014a. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? J. Mach. Learn. Res. 15, 3133–3181.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014b. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? J. Mach. Learn. Res. 15, 3133–3181.

Fotheringham, A.S., Rogerson, P.A., 2008. The SAGE Handbook of Spatial Analysis. SAGE.

Frias-Martinez, E., Williamson, G., Frias-Martinez, V., 2011. An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information, in: 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. Presented at the 2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust (PASSAT) / 2011 IEEE Third Int'l Conference on Social Computing (SocialCom), IEEE, Boston, MA, pp. 57–64. https://doi.org/10.1109/PASSAT/SocialCom.2011.142

Friedman, J., Akre, S., 2021. COVID-19 and the Drug Overdose Crisis: Uncovering the Deadliest Months in the United States, January–July 2020. Am. J. Public Health 111, 1284–1291. https://doi.org/10.2105/AJPH.2021.306256

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 29, 1189–1232.

Georganos, S., Grippa, T., Gadiaga, A.N., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2019. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto Int. 0, 1–16. https://doi.org/10.1080/10106049.2019.1595177

Ghose, R., Forati, A.M., Mantsch, J.R., 2022. Impact of the COVID-19 Pandemic on Opioid Overdose Deaths: a Spatiotemporal Analysis. J. Urban Health Bull. N. Y. Acad. Med. 1–12. https://doi.org/10.1007/s11524-022-00610-0

Gladden, R.M., Martinez, P., Seth, P., 2016. Fentanyl Law Enforcement Submissions and Increases in Synthetic Opioid–Involved Overdose Deaths — 27 States, 2013–2014. MMWR Morb. Mortal. Wkly. Rep. 65. https://doi.org/10.15585/mmwr.mm6533a2

Greenwell, B.M., 2017. pdp: An R Package for Constructing Partial Dependence Plots. https://doi.org/10.32614/rj-2017-016

Grekousis, G., Feng, Z., Marakakis, I., Lu, Y., Wang, R., 2022. Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: A geographical random forest approach. Health Place 74, 102744. https://doi.org/10.1016/j.healthplace.2022.102744

He, H., Garcia, E.A., 2009. Learning from Imbalanced Data. IEEE Trans. Knowl. Data Eng. 21, 1263–1284. https://doi.org/10.1109/TKDE.2008.239

He, L., Páez, A., Liu, D., 2017. Built environment and violent crime: An environmental audit approach using Google Street View. Comput. Environ. Urban Syst. 66, 83–95. https://doi.org/10.1016/j.compenvurbsys.2017.08.001

Hedegaard, H., Bastian, B., Trinidad, J., 2018. Drugs most frequently involved in drug overdose deaths: United States, 2011–2016. Natl. Vital Stat. Rep. Vol 67 No 9 Hyattsville MD Natl. Cent. Health Stat. 2018 14.

Hedegaard, H., Bastian, B., Trinidad, J., Warner, M., 2019. Regional Differences in the Drugs Most Frequently Involved in Drug Overdose Deaths: United States, 2017.

Hedegaard, H., Miniño, A.M., Warner, M., 2020. Drug Overdose Deaths in the United States, 1999–2018 8.

Ho, J.Y., 2017. The Contribution of Drug Overdose to Educational Gradients in Life Expectancy in the United States, 1992–2011. Demography 54, 1175–1202. https://doi.org/10.1007/s13524-017-0565-3

Hodgkinson, T., Andresen, M.A., 2019. Changing spatial patterns of residential burglary and the crime drop: The need for spatial data signatures. J. Crim. Justice 61, 90–100. https://doi.org/10.1016/j.jcrimjus.2019.04.003

Holland, K.M., Jones, C., Vivolo-Kantor, A.M., Idaikkadar, N., Zwald, M., Hoots, B., Yard, E., D'Inverno, A., Swedo, E., Chen, M.S., Petrosky, E., Board, A., Martinez, P., Stone, D.M., Law, R., Coletta, M.A., Adjemian, J., Thomas, C., Puddy, R.W., Peacock, G., Dowling, N.F., Houry, D., 2021. Trends in US Emergency Department Visits for Mental Health, Overdose, and Violence Outcomes Before and During the COVID-19 Pandemic. JAMA Psychiatry. https://doi.org/10.1001/jamapsychiatry.2020.4402

Hossain, M.M., Tasnim, S., Sultana, A., Faizah, F., Mazumder, H., Zou, L., McKyer, E.L.J., Ahmed, H.U., Ma, P., 2020. Epidemiology of mental health problems in COVID-19: a review. F1000Research 9, 636. https://doi.org/10.12688/f1000research.24457.1

Hu, N., Zhang, D., Xie, K., Liang, W., Hsieh, M.-Y., 2022. Graph learning-based spatial-temporal graph convolutional neural networks for traffic forecasting. Connect. Sci. 34, 429–448. https://doi.org/10.1080/09540091.2021.2006607

Huang, X., Lu, J., Gao, S., Wang, S., Liu, Z., Wei, H., 2021. Staying at Home Is a Privilege: Evidence from Fine-Grained Mobile Phone Location Data in the United States during the COVID-19 Pandemic. Ann. Am. Assoc. Geogr. 0, 1–20. https://doi.org/10.1080/24694452.2021.1904819

Hunt, J.M., 2016. Do crime hot spots move? Exploring the effects of the modifiable areal unit problem and modifiable temporal unit problem on crime hot spot stability (Ph.D.). American University, United States -- District of Columbia.

Hurstak, E., Rowe, C., Turner, C., Behar, E., Cabugao, R., Lemos, N.P., Burke, C., Coffin, P., 2018. Using medical examiner case narratives to improve opioid overdose surveillance. Int. J. Drug Policy 54, 35–42. https://doi.org/10.1016/j.drugpo.2017.12.017

Iyiewuare, P.O., McCullough, C., Ober, A., Becker, K., Osilla, K., Watkins, K.E., 2017. Demographic and Mental Health Characteristics of Individuals Who Present to Community Health Clinics With Substance Misuse. Health Serv. Res. Manag. Epidemiol. 4. https://doi.org/10.1177/2333392817734523

Jalal, H., Buchanich, J.M., Roberts, M.S., Balmert, L.C., Zhang, K., Burke, D.S., 2018a. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. Science 361. https://doi.org/10.1126/science.aau1184

Japkowicz, N., Stephen, S., 2002. The class imbalance problem: A systematic study. Intell. Data Anal. 6, 429. https://doi.org/10.3233/IDA-2002-6504

Jing, Y., Hu, Z., Fan, P., Xue, Y., Wang, L., Tarter, R.E., Kirisci, L., Wang, J., Vanyukov, M., Xie, X.-Q., 2020. Analysis of substance use and its outcomes by machine learning I. Childhood evaluation of liability to substance use disorder. Drug Alcohol Depend. 206, 107605. https://doi.org/10.1016/j.drugalcdep.2019.107605

Kamel Boulos, M.N., Peng, G., VoPham, T., 2019a. An overview of GeoAI applications in health and healthcare. Int. J. Health Geogr. 18, 7. https://doi.org/10.1186/s12942-019-0171-2

Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., Kruse, J., 2020. Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. Sci. Data 7, 390. https://doi.org/10.1038/s41597-020-00734-5

Kiang, M.V., Acosta, R.J., Chen, Y.-H., Matthay, E.C., Tsai, A.C., Basu, S., Glymour, M.M., Bibbins-Domingo, K., Humphreys, K., Arthur, K.N., 2022. Sociodemographic and geographic disparities in excess fatal drug overdoses during the COVID-19 pandemic in California: A population-based study. Lancet Reg. Health - Am. 11, 100237. https://doi.org/10.1016/j.lana.2022.100237

Kilroy, S., Flood, P.C., Bosak, J., Chênevert, D., 2017. Perceptions of High-Involvement Work Practices, Person-Organization Fit, and Burnout: A Time-Lagged Study of Health Care Employees. Hum. Resour. Manage. 56, 821–835. https://doi.org/10.1002/hrm.21803

Kotsiantis, S., Kanellopoulos, D., Pintelas, P., 2006. Handling imbalanced datasets: A review.

Lee, S., Lee, C., Xu, M., Li, W., Ory, M., 2022. People living in disadvantaged areas faced greater challenges in staying active and using recreational facilities during the COVID-19 pandemic. Health Place 75, 102805. https://doi.org/10.1016/j.healthplace.2022.102805

Li, H., Calder, C.A., Cressie, N., 2007. Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. Geogr. Anal. 39, 357–375. https://doi.org/10.1111/j.1538-4632.2007.00708.x

Li, M., Zhu, Z., 2021. Spatial-Temporal Fusion Graph Neural Networks for Traffic Flow Forecasting. Proc. AAAI Conf. Artif. Intell. 35, 4189–4196. https://doi.org/10.1609/aaai.v35i5.16542

Li, Y., Miller, H.J., Root, E.D., Hyder, A., Liu, D., 2022. Understanding the role of urban social and physical environment in opioid overdose events using found geospatial data. Health Place 75, 102792. https://doi.org/10.1016/j.healthplace.2022.102792

Lipton, R., Yang, X., A. Braga, A., Goldstick, J., Newton, M., Rura, M., 2013. The Geography of Violence, Alcohol Outlets, and Drug Arrests in Boston. Am. J. Public Health 103, 657–664. https://doi.org/10.2105/AJPH.2012.300927

Liu, C., Wang, T., 2017. Identifying and mapping local contributions of carbon emissions from urban motor and metro transports: A weighted multiproxy allocating approach. Comput. Environ. Urban Syst. 64, 132–143. https://doi.org/10.1016/j.compenvurbsys.2017.01.010

Liu, X., Wu, J., Zhou, Z., 2009. Exploratory Undersampling for Class-Imbalance Learning. IEEE Trans. Syst. Man Cybern. Part B Cybern. 39, 539–550. https://doi.org/10.1109/TSMCB.2008.2007853

Lum, K., Isaac, W., 2016. To predict and serve? Significance 13, 14–19. https://doi.org/10.1111/j.1740-9713.2016.00960.x

Ma, X., Guo, J., Wu, J., Liu, H., Yu, J., Xie, J., Sun, X., 2011. Prediction of RNA-binding residues in proteins from primary sequence using an enriched random forest model with a novel hybrid feature. Proteins Struct. Funct. Bioinforma. 79, 1230–1239. https://doi.org/10.1002/prot.22958

Malley, J.D., Kruppa, J., Dasgupta, A., Malley, K.G., Ziegler, A., 2012. Probability machines: consistent probability estimation using nonparametric learning machines. Methods Inf. Med. 51, 74–81. https://doi.org/10.3414/ME00-01-0052

Marotta, P.L., Hunt, T., Gilbert, L., Wu, E., Goddard-Eckrich, D., El-Bassel, N., 2019. Assessing Spatial Relationships between Prescription Drugs, Race, and Overdose in New York State from 2013 to 2015. J. Psychoactive Drugs 51, 360–370. https://doi.org/10.1080/02791072.2019.1599472

Marshall, B.D.L., Krieger, M.S., Yedinak, J.L., Ogera, P., Banerjee, P., Alexander-Scott, N.E., Rich, J.D., Green, T.C., 2017. Epidemiology of fentanyl-involved drug overdose deaths: A geospatial retrospective study in Rhode Island, USA. Int. J. Drug Policy 46, 130–135. https://doi.org/10.1016/j.drugpo.2017.05.029

Martins, S.S., Sampson, L., Cerdá, M., Galea, S., 2015. Worldwide Prevalence and Trends in Unintentional Drug Overdose: A Systematic Review of the Literature. Am. J. Public Health 105, e29–e49. https://doi.org/10.2105/AJPH.2015.302843

Mateo-Urdiales, A., Fabiani, M., Rosano, A., Vescio, M.F., Del Manso, M., Bella, A., Riccardo, F., Pezzotti, P., Regidor, E., Andrianou, X., 2021. Socioeconomic patterns and COVID-19 outcomes before, during and after the lockdown in Italy (2020). Health Place 71, 102642. https://doi.org/10.1016/j.healthplace.2021.102642

McCord, E.S., Ratcliffe, J.H., 2007a. A Micro-Spatial Analysis of the Demographic and Criminogenic Environment of Drug Markets in Philadelphia. Aust. N. Z. J. Criminol. 40, 43–63. https://doi.org/10.1375/acri.40.1.43

Mertz, K.J., Janssen, J.K., Williams, K.E., 2014. Underrepresentation of Heroin Involvement in Unintentional Drug Overdose Deaths in Allegheny County, PA. J. Forensic Sci. 59, 1583–1585. https://doi.org/10.1111/1556-4029.12541

Miller, J.N., 2020. The War on Drugs 2.0: Darknet Fentanyl's Rise and the Effects of Regulatory and Law Enforcement Action. Contemp. Econ. Policy 38, 246–257. https://doi.org/10.1111/coep.12447

Mohler, G., 2014. Marked point process hotspot maps for homicide and gun crime prediction in Chicago. Int. J. Forecast. 30, 491–497. https://doi.org/10.1016/j.ijforecast.2014.01.004

Molnar, C., 2019. Chapter 5 Model-Agnostic Methods | Interpretable Machine Learning.

Moore, T.H.M., Kesten, J.M., López-López, J.A., Ijaz, S., McAleenan, A., Richards, A., Gray, S., Savović, J., Audrey, S., 2018. The effects of changes to the built environment on the mental health and well-being of adults: Systematic review. Health Place 53, 237–257. https://doi.org/10.1016/j.healthplace.2018.07.012

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. Biometrika 37, 17–23. https://doi.org/10.2307/2332142

Moreland, A., 2020. Timing of State and Territorial COVID-19 Stay-at-Home Orders and Changes in Population Movement — United States, March 1–May 31, 2020. MMWR Morb. Mortal. Wkly. Rep. 69. https://doi.org/10.15585/mmwr.mm6935a2

Mosher, C.J., Miethe, T.D., Hart, T.C., 2010. The Mismeasure of Crime. SAGE Publications.

National Center for Health Statistics, 2021. Drug Overdose Deaths in the U.S. Top 100,000 Annually [WWW Document]. URL https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2021/20211117.htm (accessed 4.25.22).

National Institute of Environmental Health Sciences, 2020. The COVID-19 Pandemic Is Fueling the Opioid Crisis! 4.

National Institute on Drug, 2022. Overdose Death Rates [WWW Document]. Natl. Inst. Drug Abuse. URL https://nida.nih.gov/drug-topics/trends-statistics/overdose-death-rates (accessed 4.25.22).

National Institute on Drug Abuse, 2019a. Opioids [WWW Document]. URL https://www.drugabuse.gov/drugs-abuse/opioids (accessed 11.19.19).

National Institute on Drug Abuse, 2019b. Overdose Death Rates [WWW Document]. URL https://www.drugabuse.gov/related-topics/trends-statistics/overdose-death-rates (accessed 8.30.19).

Nechuta, S.J., Tyndall, B.D., Mukhopadhyay, S., McPheeters, M.L., 2018. Sociodemographic factors, prescription history and opioid overdose deaths: a statewide analysis using linked PDMP and mortality data. Drug Alcohol Depend. 190, 62–71. https://doi.org/10.1016/j.drugalcdep.2018.05.004

Nembrini, S., König, I.R., Wright, M.N., 2018. The revival of the Gini importance? Bioinformatics 34, 3711–3718. https://doi.org/10.1093/bioinformatics/bty373

Neutens, T., Farber, S., Delafontaine, M., Boussauw, K., 2013. Spatial variation in the potential for social interaction: A case study in Flanders (Belgium). Comput.

Environ. Urban Syst. 41, 318–331. https://doi.org/10.1016/j.compenvurbsys.2012.06.007

O’Donnell, J., Gladden, R.M., Goldberger, B.A., Mattson, C.L., Kariisa, M., 2020. Notes from the Field: Opioid-Involved Overdose Deaths with Fentanyl or Fentanyl Analogs Detected — 28 States and the District of Columbia, July 2016–December 2018. Morb. Mortal. Wkly. Rep. 69, 271–273. https://doi.org/10.15585/mmwr.mm6910a4

O’Donnell, J.K., Gladden, R.M., Seth, P., 2017. Trends in Deaths Involving Heroin and Synthetic Opioids Excluding Methadone, and Law Enforcement Drug Product Reports, by Census Region — United States, 2006–2015. N. Y. 66, 7.

Parascandola, M., Weed, D.L., 2001. Causation in epidemiology. J. Epidemiol. Community Health 55, 905–912. https://doi.org/10.1136/jech.55.12.905

Pardo, B., Reuter, P., 2020. Enforcement strategies for fentanyl and other synthetic opioids. Washington, DC: The Brookings Institution.

Peterson, A.B., Gladden, R.M., Delcher, C., Spies, E., Garcia-Williams, A., Wang, Y., Halpin, J., Zibbell, J., McCarty, C.L., DeFiore-Hyrmer, J., DiOrio, M., Goldberger, B.A., 2016. Increases in Fentanyl-Related Overdose Deaths — Florida and Ohio, 2013–2015. MMWR Morb. Mortal. Wkly. Rep. 65, 844–849. https://doi.org/10.15585/mmwr.mm6533a3

Pierce, J.B., Harrington, K., McCabe, M.E., Petito, L.C., Kershaw, K.N., Pool, L.R., Allen, N.B., Khan, S.S., 2021. Racial/ethnic minority and neighborhood disadvantage leads to disproportionate mortality burden and years of potential life lost due to COVID-19 in Chicago, Illinois. Health Place 68, 102540. https://doi.org/10.1016/j.healthplace.2021.102540

Piza, E.L., Carter, J.G., 2018. Predicting Initiator and Near Repeat Events in Spatiotemporal Crime Patterns: An Analysis of Residential Burglary and Motor Vehicle Theft. Justice Q. 35, 842–870. https://doi.org/10.1080/07418825.2017.1342854

Probst, P., Wright, M.N., Boulesteix, A.-L., 2019. Hyperparameters and tuning strategies for random forest. WIREs Data Min. Knowl. Discov. 9, e1301. https://doi.org/10.1002/widm.1301

Probst, W.N., Stelzenmüller, V., Fock, H.O., 2012. Using cross-correlations to assess the relationship between time-lagged pressure and state indicators: an exemplary analysis of North Sea fish population indicators. ICES J. Mar. Sci. 69, 670–681. https://doi.org/10.1093/icesjms/fss015

Quick, H., 2019. Estimating County-Level Mortality Rates Using Highly Censored Data From CDC WONDER. Prev. Chronic. Dis. 16. https://doi.org/10.5888/pcd16.180441

Reisner, S.L., Greytak, E.A., Parsons, J.T., Ybarra, M.L., 2015. Gender Minority Social Stress in Adolescence: Disparities in Adolescent Bullying and Substance Use by Gender Identity. J. Sex Res. 52, 243–256. https://doi.org/10.1080/00224499.2014.886321

Ruhm, C.J., 2016. Drug poisoning deaths in the United States, 1999–2012: a statistical adjustment analysis. Popul. Health Metr. 14, 2. https://doi.org/10.1186/s12963-016-0071-7

SafeGraph, 2022. Social Distancing Metrics | SafeGraph Docs [WWW Document]. SafeGraph. URL https://docs.safegraph.com/docs/social-distancing-metrics (accessed 4.26.22).

Sato, K., Kondo, N., Kondo, K., 2022. Pre-pandemic individual- and community-level social capital and depressive symptoms during COVID-19: A longitudinal study of Japanese older adults in 2019-21. Health Place 74, 102772. https://doi.org/10.1016/j.healthplace.2022.102772

Shiode, S., Shiode, N., 2020. A network-based scan statistic for detecting the exact location and extent of hotspots along urban streets. Comput. Environ. Urban Syst. 83, 101500. https://doi.org/10.1016/j.compenvurbsys.2020.101500

Shiode, S., Shiode, N., Block, R., Block, C.R., 2015. Space-time characteristics of micro-scale crime occurrences: an application of a network-based space-time search window technique for crime incidents in Chicago. Int. J. Geogr. Inf. Sci. 29, 697–719. https://doi.org/10.1080/13658816.2014.968782

Smith, L.V., Tarui, N., Yamagata, T., 2021. Assessing the impact of COVID-19 on global fossil fuel consumption and CO2 emissions. Energy Econ. 97, 105170. https://doi.org/10.1016/j.eneco.2021.105170

Spencer, M.R., Warner, M., Bastian, B.A., Trinidad, J.P., Hedegaard, H., 2019. Drug Overdose Deaths Involving Fentanyl, 2011-2016. Natl. Vital Stat. Rep. Cent. Dis. Control Prev. Natl. Cent. Health Stat. Natl. Vital Stat. Syst. 68, 1–19.

Srinivasan, S., O'Fallon, L.R., Dearry, A., 2003. Creating Healthy Communities, Healthy Homes, Healthy People: Initiating a Research Agenda on the Built Environment and Public Health. Am. J. Public Health 93, 1446–1450. https://doi.org/10.2105/AJPH.93.9.1446

State of Illinois, 2020. Gov. Pritzker Announces Statewide Stay At Home Order to Maximize COVID-19 Containment, Ensure Health Care System Remains Fully Operational [WWW Document]. URL https://www.illinois.gov/news/press-release.html (accessed 7.12.22).

Stewart, K., Cao, Y., Hsu, M.H., Artigiani, E., Wish, E., 2017a. Geospatial Analysis of Drug Poisoning Deaths Involving Heroin in the USA, 2000–2014. J. Urban Health 94, 572–586. https://doi.org/10.1007/s11524-017-0177-7

Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8, 25. https://doi.org/10.1186/1471-2105-8-25

Substance Abuse and Mental Health Services Administration, 2021. Key Substance Use and Mental Health Indicators in the United States: Results from the 2020 National Survey on Drug Use and Health 156.

Substance Abuse and Mental Health Services Administration, 2019. Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health 82.

Substance Abuse and Mental Health Services Administration, 2018. Using International Classification of Diseases (ICD) Codes to Assess Opioid-Related Overdose Deaths. MMWR Morb. Mortal. Wkly. Rep. 65, 1445–1452. https://doi.org/10.15585/mmwr.mm655051e1

Sun, Y., Bao, Y., Kosten, T., Strang, J., Shi, J., Lu, L., 2020. Editorial: Challenges to Opioid Use Disorders During COVID-19. Am. J. Addict. 29, 174–175. https://doi.org/10.1111/ajad.13031

Sutter, A., Curtis, M., Frost, T., 2019. Public drug use in eight U.S. cities: Health risks and other factors associated with place of drug use. Int. J. Drug Policy 64, 62–69. https://doi.org/10.1016/j.drugpo.2018.11.007

Tiwari, A., Dadhania, A.V., Ragunathrao, V.A.B., Oliveira, E.R.A., 2021. Using machine learning to develop a novel COVID-19 Vulnerability Index (C19VI). Sci. Total Environ. 773, 145650. https://doi.org/10.1016/j.scitotenv.2021.145650

United States Environmental Protection Agency, 2019. Basic Information about the Built Environment [WWW Document]. US EPA. URL https://www.epa.gov/smm/basic-information-about-built-environment (accessed 4.29.20).

Urban Data Visualization Lab, 2022. COVID-19 Vulnerability Maps | Urban Data Visualization Lab | University of Illinois Chicago [WWW Document]. URL https://udv.lab.uic.edu/featured-projects/covid-19-vulnerability-maps/ (accessed 6.2.22).

U.S. Drug Enforcement Administration, 2022. Fentanyl [WWW Document]. URL https://www.dea.gov/factsheets/fentanyl (accessed 10.9.22).

VanderWeele, T.J., 2020. Invited Commentary: Counterfactuals in Social Epidemiology—Thinking Outside of “the Box.” Am. J. Epidemiol. 189, 175–178. https://doi.org/10.1093/aje/kwz198

Vilalta, C.J., 2010. The spatial dynamics and socioeconomic correlates of drug arrests in Mexico city. Appl. Geogr. 30, 263–270. https://doi.org/10.1016/j.apgeog.2009.06.001

Visconti, A.J., Santos, G.-M., Lemos, N.P., Burke, C., Coffin, P.O., 2015. Opioid Overdose Deaths in the City and County of San Francisco: Prevalence, Distribution, and Disparities. J. Urban Health 92, 758–772. https://doi.org/10.1007/s11524-015-9967-y

Vomfell, L., Härdle, W.K., Lessmann, S., 2018. Improving crime count forecasts using Twitter and taxi data. Decis. Support Syst. 113, 73–85. https://doi.org/10.1016/j.dss.2018.07.003

Weisburd, D., Mazerolle, L.G., 2000. Crime and Disorder in Drug Hot Spots: Implications for Theory and Practice in Policing. Police Q. 3, 331–349. https://doi.org/10.1177/1098611100003003006

White, H., 2006. Impact evaluation: the experience of the Independent Evaluation Group of the World Bank [WWW Document]. URL https://mpra.ub.uni-muenchen.de/1111/ (accessed 4.26.22).

World Health Organization, 2022. Coronavirus disease (COVID-19) – World Health Organization [WWW Document]. URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed 4.25.22).

World Health Organization, 2020. Coronavirus disease (COVID-19) – World Health Organization [WWW Document]. URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed 12.29.20).

Wright, M.N., Ziegler, A., 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Softw. 77, 1–17. https://doi.org/10.18637/jss.v077.i01

Xia, Z., Stewart, K., Fan, J., 2021. Incorporating space and time into random forest models for analyzing geospatial patterns of drug-related crime incidents in a major U.S. metropolitan area. Comput. Environ. Urban Syst. 87, 101599. https://doi.org/10.1016/j.compenvurbsys.2021.101599

Xiong, C., Hu, S., Yang, M., Luo, W., Zhang, L., 2020. Mobile device data reveal the dynamics in a positive relationship between human mobility and COVID-19 infections. Proc. Natl. Acad. Sci. 117, 27087–27089. https://doi.org/10.1073/pnas.2010836117

Yang, B., Liu, L., Lan, M., Wang, Z., Zhou, H., Yu, H., 2020. A spatio-temporal method for crime prediction using historical crime data and transitional zones identified from nightlight imagery. Int. J. Geogr. Inf. Sci. 0, 1–25. https://doi.org/10.1080/13658816.2020.1737701

Yarbrough, C.R., Abraham, A.J., Adams, G.B., 2019. Relationship of County Opioid Epidemic Severity to Changes in Access to Substance Use Disorder Treatment, 2009–2017. Psychiatr. Serv. 71, 12–20. https://doi.org/10.1176/appi.ps.201900150

Ye, X., Wu, L., 2011. Analyzing the dynamics of homicide patterns in Chicago: ESDA and spatial panel approaches. Appl. Geogr. 31, 800–807. https://doi.org/10.1016/j.apgeog.2010.08.006

Zhang, Y., Cheng, T., 2020a. Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. Comput. Environ. Urban Syst. 79, 101403. https://doi.org/10.1016/j.compenvurbsys.2019.101403

Zhang, Y., Cheng, T., 2020b. Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. Comput. Environ. Urban Syst. 79, 101403. https://doi.org/10.1016/j.compenvurbsys.2019.101403

Zhao, X., Tang, J., 2017. Modeling Temporal-Spatial Correlations for Crime Prediction, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM ’17. Association for Computing Machinery, Singapore, Singapore, pp. 497–506. https://doi.org/10.1145/3132847.3133024

Zhou, H., Zhang, F., Du, Z., Liu, R., 2021. Forecasting PM2.5 using hybrid graph convolution-based model considering dynamic wind-field to offer the benefit of spatial interpretability. Environ. Pollut. 273, 116473. https://doi.org/10.1016/j.envpol.2021.116473

Zhu, G., Stewart, K., Niemeier, D., Fan, J., 2021. Understanding the Drivers of Mobility during the COVID-19 Pandemic in Florida, USA Using a Machine Learning Approach. ISPRS Int. J. Geo-Inf. 10, 440. https://doi.org/10.3390/ijgi10070440