

Methodological approaches for the prediction of opioid use-related epidemics in the United States: a narrative review and cross-disciplinary call to action



CHARLES MARKS, GABRIEL CARRASCO-ESCOBAR, ROCÍO CARRASCO-HERNÁNDEZ, DEREK JOHNSON, DAN CICCARONE, STEFFANIE A STRATHDEE, DAVEY SMITH, and ANNICK BÓRQUEZ

SAN DIEGO; LIMA PERU; AND SAN FRANCISCO

The opioid crisis in the United States has been defined by waves of drug- and local-ity-specific Opioid use-Related Epidemics (OREs) of overdose and bloodborne infections, among a range of health harms. The ability to identify localities at risk of such OREs, and better yet, to predict which ones will experience them, holds the potential to mitigate further morbidity and mortality. This narrative review was conducted to identify and describe quantitative approaches aimed at the “risk assessment,” “detection” or “prediction” of OREs in the United States. We implemented a PubMed search composed of the: (1) objective (eg, prediction), (2) epidemiologic outcome (eg, outbreak), (3) underlying cause (ie, opioid use), (4) health outcome (eg, overdose, HIV), (5) location (ie, US). In total, 46 studies were included, and the following information extracted: discipline, objective, health outcome, drug/substance type, geographic region/unit of analysis, and data sources. Studies identified relied on clinical, epidemiological, behavioral and drug markets surveillance and applied a range of methods including statistical regression, geospatial analyses, dynamic modeling, phylogenetic analyses and machine learning. Studies for the prediction of overdose mortality at national/state/county and zip code level are rapidly emerging. Geospatial methods are increasingly used to identify hotspots of opioid use and overdose. In the context of infectious disease OREs, routine genetic sequencing of patient samples to identify growing transmission clusters via phylogenetic methods could increase early detection capacity. A coordinated implementation of multiple, complementary approaches would increase our ability to successfully anticipate outbreak risk and respond preemptively. We present a multi-disciplinary framework for the prediction of OREs in the US and reflect on challenges research teams will face in implementing such strategies along with good practices. (Translational Research 2021; 234:88–113)

From the Interdisciplinary Research on Substance Use Joint Doctoral Program at San Diego State University and University of California, San Diego; Division of Infectious Diseases and Global Public Health, University of California, San Diego; School of Social Work, San Diego State University; Health Innovation Laboratory, Institute of Tropical Medicine “Alexander von Humboldt”, Universidad Peruana Cayetano Heredia, Lima, Peru; Department of Family and Community Medicine, University of California San Francisco.

Submitted for Publication December 18, 2020; received submitted March 25, 2021; accepted for publication March 25, 2021.

Reprint requests: Annick Bórquez, Division of Infectious Diseases and Global Public Health, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92103. e-mail:

aborquez@health.ucsd.edu.

1931-5244/\$ - see front matter

© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license

(

Abbreviations: US = United States; CDC = Centers for disease control; ORE = opioid use-related epidemics; IDU = injection drug use; PWID = people who inject drugs; HIV = human immunodeficiency virus; HCV = hepatitis C virus; TB = tuberculosis; SUD = substance-use disorders; OUD = opioid use disorders; ARIMA = autoregressive integrated moving average; ESDA = exploratory spatial data analysis; LISA = Local Indicators of Spatial Association; SaTScan = spatial scan statistic

INTRODUCTION

The United States (US) is experiencing one of the most devastating public health crises of modern times as a result of serial and intertwining epidemic waves of opioid use and associated health harms, opioid use related epidemics (OREs) hereafter, defined as increased number of cases of overdoses, and/or infections such as HIV, associated with opioid use for a given locality than would normally be expected.

1–5 Over 750,000 lives have been lost to overdose since 2000 in the US, of which 70,630 occurred in 2019 alone.

6

7 Further, preliminary data indicates that the rate of drug overdose deaths has likely accelerated throughout the COVID-19 epidemic.

8 In addition, a number of opioid use associated HIV outbreaks

9–13 have been identified across the country over the past few years; acute HCV incidence has increased by 350% between 2010 and 2016

14

16 and over 15,000 Hepatitis A virus (HAV) cases have been identified across several states since 2016, mostly among homeless and opioid using communities.

17 As novel and synthetic opioids of high potency enter the market, the magnitude and speed of these epidemics increases and the capacity of the public health system to effectively respond is further reduced. The failure to curb this crisis has in great part been a consequence of lagging public health responses, which have so far been reactive rather than preemptive.

18

To effectively respond to geographically diverse emerging OREs, affecting different populations and involving increasingly potent substances, we argue that a shift in perspective, founded on epidemic preparedness, is needed. Preparedness is an integral component in the control of infectious disease epidemics,

19

20 as well as in the management of natural or man-made disasters,

21–23 but it has not been systematically integrated in the context of OREs despite their exceptional impact on public health. Anticipating such threats requires a strong surveillance infrastructure and the development of forecasting tools, which we categorize into three broad groups responding to the following objectives: early detection, risk assessment and prediction.

Early detection refers to the identification of OREs at the initial stages of the epidemic, leading to the implementation of control measures (such as contact tracing for infectious diseases

24 or distribution of drug testing strips in the context of drug adulteration as seen with fentanyl

25

26), thereby preventing further disease transmission or mortality associated with the epidemic.

Risk assessments determine and rank the susceptibility of specific demographics (eg, by ethnicity) and geographical areas (eg, counties) to experiencing an ORE based on assessing the prevalence of risk factors observed in past epidemics or identified in observational studies. Risk assessment generally result in a score or ranking which indicates how likely it is an ORE occurs in each given locality or among specific populations. Risk assessments can inform the allocation of resources and ensure that local public health stakeholders are prepared for such eventualities.

Predictions estimate the future incidence of opioid use disorders (OUD), overdose, and associated health harms. Models (ie, statistical, mathematical, heuristic) can be developed and trained on data-related to the mechanisms of disease and drug use spread, progression and severity over time, with the capacity to take into account dynamic factors such as population demographics or drug market changes.

27 These models can then be applied to project the dynamics of future ORE occurrence, temporally and geospatially. Whereas risk assessment is concerned with the likelihood that an ORE may occur (for example, the relative likelihood of a county-level fentanyl overdose outbreak over the next three years, as compared to other counties), prediction extends this by attempting to predict the trajectory of

the ORE (for example, the specific annual county-level fentanyl overdose death rate over the next three years for all counties). As such, they provide more specific information about the magnitude and likely course of an epidemic before its emergence or along its development. The ability to predict OREs and their magnitude and severity allows for the assessment of potential interventions to mitigate the harms. If we can make well-informed predictions, we are best equipped to assess the best course of action to prevent future harm.

Fig 1 illustrates the use of these types of tools at different stages of the epidemic, as well as their relative impact on reducing incidence.

This narrative review aimed to identify the different quantitative methods that have been applied towards the early detection, risk assessment and prediction of OREs in the US. Having a birds-eye view of existing approaches, the type of findings/outcomes that they

enable, the data they require and their strengths and limitations, is a necessary step towards maximizing scientific knowledge and promoting coordination. We describe and characterize the main approaches used to date in the context of OREs in the US and share an open repository of studies on our project website (<https://www.emergens-project.com/repository>), which we continue to update, and invite contributions from the research community.

METHODS

We implemented a PubMed search to identify studies aiming to detect or forecast outbreaks of opioid use and related health harms in the US. This included five main components; (1) the key health outcomes of interest: based on previous research, we included OUD, overdose, HIV, hepatitis, tuberculosis, as well as opioid-related admissions; (2) opioid use: to restrict the search to outbreaks that were related to opioid use while making sure we captured all relevant studies, we

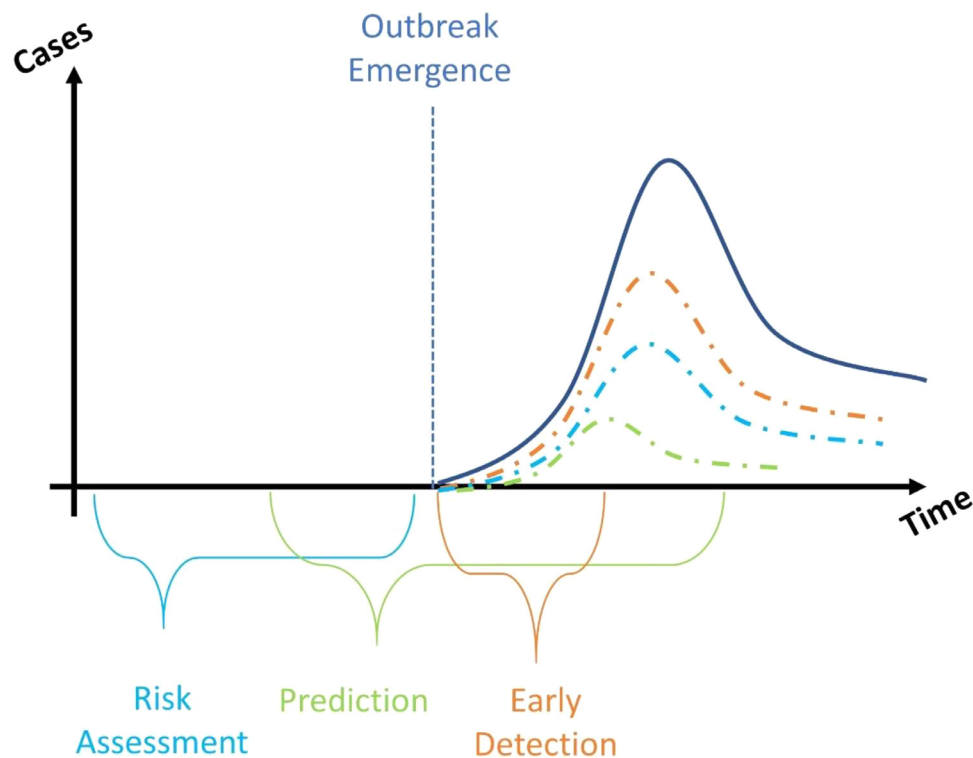


Fig 1. Early detection, risk assessment and prediction of opioid use related epidemics. Legend: The black curve corresponds to the epidemic at baseline, while the dashed orange, blue and green curves correspond to the epidemic in the presence of interventions resulting from early detection, risk assessment and prediction, respectively. The time periods for each of these analyses are also colored in blue, green, and orange, respectively. We hypothesize that accurate prediction would have the strongest prevention impact because it would confer time to plan and implement an appropriate response, followed by risk assessment, which is less specific and therefore less informative, and by early detection, which is highly specific but occurs once the epidemic has started spreading. However, this impact will depend on how this evidence is used by decision makers.

included terms such as “substance use” and “people who inject drugs”; (3) the epidemiological outcome of interest, corresponding to outbreaks, hotspots, clusters and related terms; (4) the study objective, using terms related to “detection”, “prediction” and “risk assessment”; and (5) the location, corresponding to the “United States”. The complete list of search terms is provided in appendix. The search was last updated on January 29th 2021, CM and AB scanned the abstracts and selected relevant articles for full-text consideration by the research team. A snowball approach was used to identify additional articles by scanning the references and citations of each selected article.

For selection, articles were required to address OREs at the population level, contribute to their early detection, risk assessment, and/or prediction, be implemented in the US and be of a quantitative nature. Determination of inclusion in the study was done through discussion with all authors, who had expertise in a range of fields (ie geospatial analyses, phylogenetics, statistics, dynamic modeling, HIV and substance use epidemiology, and computer sciences) and who evaluated the contribution of each study to the detection and/or forecasting of OREs. Authors extracted the following information from the articles within their field of expertise: field of study, objective, health outcome, substance and/or drug, geographic region, geographic unit of analysis, methods and data sources. While our objective was to describe the landscape of methods that can be applied in future ORE prediction efforts, rather than evaluate studies’ quality, we provided an assessment of their limitations when appropriate. Given the breadth of the research question, a narrative review format was chosen.

This study did not involve human subjects as it relied on secondary reporting of aggregated data from multiple studies with no identifiable information.

RESULTS

Our initial search returned 1,265 articles. After full text screening and dialog with the full research team, 46 articles were determined to meet the inclusion criteria for this study. Selected studies are presented in [Table I](#) alongside their corresponding objective and other extracted information. These studies were implemented at the national, state, county, and local (ZIP Code) levels. Health outcomes included OUD, overdose, HCV, HIV, and TB. Substances included: “any substance”, opioids, and novel substances. We describe their potential contribution to policy guidance at different stages of the epidemic

and further disaggregate by type of health outcome (ie, overdose, bloodborne infections and tuberculosis). Studies are organized first by overarching objective (Early Detection, Risk Assessment, Prediction) and then by type of outcome (Overdose, Bloodborne-Infections).

Early detection. The ability to detect OREs is dependent upon the clinical, epidemiological, behavioral and drug market surveillance infrastructure in place as data can be examined to identify unusual patterns. We describe selected studies below and provide further information in

[Table I.](#)

Overdose. In late 2005 to early 2006, it was discovered that heroin was being adulterated with fentanyl and an increase in heroin- and opioid-related deaths in Illinois was detected and reported to the CDC in April of 2006.

[29](#) Friedman found that if public health officials were prospectively analyzing poison call center data for heroin-related exposures, the use of time series analysis — specifically autoregressive integrated moving average (ARIMA) modeling — would have detected this ORE one month earlier, in March of 2006.

[29](#) Similarly, Li et al. found that the location of 311 municipal service calls in Columbus, OH was associated with opioid overdose geographic hotspots.

[33](#) A study conducted in Texas used a combination of data from emergency room visits, the Youth Risk Behavior Survey, poison control cases, and qualitative interviews to characterize a “cheese” heroin use outbreak.

[31](#) Another study implemented in Kentucky used data from a combination of death certificates, overdose deaths from the State Medical Examiners’ Office, emergency room visits, and prescription drug monitoring programs to detect potential opioid use clusters at state level.

[30](#) Rosenblum et al. analyzed law enforcement drug seizure data, finding that increased presence of fentanyl (and analogs) in drug samples was associated with elevated overdose death rates, indicating that the analysis of such data can identify changes in drug markets which may drive OREs.

[27](#)

In the context of internet surveillance, Anwar et al sought to determine if opioid-related Twitter posts could have been used to identify the three waves of the opioid crisis in North Carolina from 2009 to 2017.

Table I. Papers that met study inclusion criteria. They are first organized by overarching objective of the study (early detection, risk assessment, or prediction) and then by outcome type (overdose, bloodborne-infection, tuberculosis). The following information was extracted and presented in the table for each study: methodology (ie, statistical regression, geospatial analyses, etc), overarching study objective, substances measured, health outcome of interest, region in which study took place, unit of analysis (ie, state, county, locality), specific methods employed, and data sources

Paper	Methodology	Objective	Substances	Health Outcomes	Region	Unit of Analysis	Specific Methods Employed	Data Sources
Early detection								
<i>Overdose</i>								
Friedman 29	Statistical regression (SR)	To determine if the 2006 Illinois fentanyl outbreak could have been detected earlier	Heroin, methadone, and "other/unknown opioids"	Toxidrome = symptoms associated to exposure	Illinois	Individual cases atone callcenter	Cumulative sum modeling of the residuals, Pulse analysis method, Autoregressive integrated moving average model	Illinois Poison Center
Maxwell et al. 31	Descriptive statistics	To identify and describe the demographics of a 2008 overdose outbreak	"Cheese" heroin	Overdose deaths. Treatment admissions	Dallas, Texas	County	Student T-test, Chi-sq Test	Youth Risk Behavior Survey, Police arrest database, Poison control center cases
Li et al. 33	SR	To identify geographic "hotspots" of opioid overdose using "311" municipal service requests	Opioids	Overdose deaths	Columbus, Ohio	Neighborhoods	Cross-K-function, k-means, ANOVA	311 Municipal Service Data, Columbus Ohio, Census Data, American Community Survey (ACS)
Rosenblum et al. 27	SR & Epidemiological Surveillance (ES)	To assess the association between drug types identified in drug samples seized by law enforcement in Ohio and overdose rates	Any	Overdose deaths	Ohio	County	Poisson Regression	Ohio Bureau of Criminal Investigation
Anwar et al. 34	SR with internet data (ID)	To explore the potential of using Twitter data to monitor the opioid epidemic	Opioids	Overdose deaths	North Carolina	State	OLS regression, vector autoregression, Granger tests	Twitter, CDC Wonder
Chary et al. 35	SR with ID	To demonstrate that the geographic variation of social media posts mentioning prescription opioid "misuse" strongly correlates with government estimates of prescription opioid "misuse" in the last month	Prescription opioids	Non-medical use of prescription opioids	US	States	Curation of Tweets and Spearman rank correlation coefficient to compare to NSDUH data	Twitter, NSDUH

(continued on next page)

Table I. (Continued)

Paper	Methodology	Objective	Substances	Health Outcomes	Region	Unit of Analysis	Specific Methods Employed	Data Sources
Sarker et al. 36	Machine learning (ML) with ID	To develop and validate an automatic text-processing pipeline for geospatial and temporal analysis of opioid-mentioning social media chatter	Opioids	Opioid overdose deaths, non-medical prescription opioid use, illicit drug use, illicit drug dependence, and illicit drug dependence or "abuse"	Pennsylvania	County	Compared 5 machine learning approaches: naive bayes, decision tree, k-nearest neighbors, random forest, support vector machine, and a deep convolutional neural network. Calculated recall, precision and F1-score. To assess correlation with outcomes, used Pearson and Spearman analyses	CDC Wonder, substate-level NSDUH data, Twitter
Hswen et al. 37	SR with ID	To characterize the relationship between volume of online opioid news reporting and opioid-related deaths in the United States and how these measures differ across geographic and socioeconomic county-level factors	Opioids	Opioid overdose deaths	US	County	Regression analysis	Google News, CDC WONDER, US Census Bureau of Statistics
Bloodborne Infections: HIV and HCV								
Fitzmaurice et al. 28	Descriptive Statistics	To develop a method to identify spatio-temporal clusters of increased HIV diagnoses	IDU	HIV	50 US states and the District of Columbia	State and County	Pearson test for homogeneity	NHSS data
CDC 38	Outbreak investigation	To identify HCV infections most likely to have been acquired recently	IDU	HCV	New York	County	Enhanced epidemiologic surveillance	HCV cases and their socio-demographic data
Alpren et al. 9	Phylogenetics (Pg)	To describe and control an outbreak of HIV among PWID	Opioids (including fentanyl) and stimulants	HIV	Lawrence and Lowell, Massachusetts	City	Genotyping (HIV-TRACE) and epidemiologic investigation, including in depth interviews (ID)	HIV data and medical records from Massachusetts department of health, data from in depth interviews with PWID and stakeholders
Peters et al. 10	Pg	To describe and control an outbreak of HIV among PWID	Opioids (including extended release oxymorphone)	HIV	Scott county, Indiana	County	Genotyping and epidemiologic investigation	HIV data and socio-demographic data from outreach testing

(continued on next page)

Table I. (Continued)

Paper	Methodology	Objective	Substances	Health Outcomes	Region	Unit of Analysis	Specific Methods Employed	Data Sources
Oster et al. 40	Pg	To identify growing clusters of recent, rapid HIV transmission	Does not focus on particular substances, but includes IDU	HIV	27 HIV surveillance jurisdictions	National	Genotyping (HIV-TRACE) and Bayesian molecular clock phylogenetic inference	HIV sequence data from recent diagnoses, medical records from the National HIV Surveillance System
Falade-Nwulia et al. 44	Pg	To identify HCV clusters and factors related to being part of such a cluster among HCV+ PWID	IDU	HCV	Baltimore, Maryland	City	Genotyping, logistic Regression	ALIVE Study
Cai et al. 96	ML with ID	To characterize the Twitter dialogs related to the 2015 Scott County HIV outbreak	Opioids	–	Indiana	State	Natural Language Processing – Biterm Topic Model	Twitter
Tookes et al 43	ES & Pg	To identify HIV outbreaks among PWID	IDU	HIV	Florida	County	Epidemiologic surveillance and genetic analyses (HIV TRACE)	Routine HIV screening at syringe services program and genetic data
Zhou et al 45	Pg	To identify recent HIV infections, detect drug resistance mutations and phylogenetically linked cluster		HIV	North Carolina	State	Multiplexed Primer ID-Next Gen Sequencing approach	HIV diagnostic specimens, Stanford HIV drug resistance database and list of drug resistance mutations from the International Antiviral Society–USA
Tuberculosis CDC 46	Pg	To recognize and control a growing TB transmission cluster related to homelessness and substance use	Alcohol and illicit substances	TB	Duval county, Florida	County	Genotyping, combined with epidemiologic investigation and social network analysis	TB data and medical records from Florida department of health, Duval County record of jails, homeless and outpatient mental health services
Althomsons et al. 47	Pg	To identify TB clusters at risk of developing into outbreaks	Alcohol and illicit drugs	TB	50 states	National	Genotyping, Linear regression and decision tree analysis	CDC National TB Genotyping Service data, medical records from CDC National TB surveillance system, US Census Data
Risk Assessment Overdose Endo et al. 59	Wastewater testing	To semi-quantitatively detect opioid metabolites (eg, morphine glucuronide) at a sub-city community resolution	Opioids	Opioids, naloxone and buprenorphine presence	North Carolina	Urban municipality	Robotic wastewater sampling platform at ten residential manholes and targeted mass spectrometry	Wastewater

(continued on next page)

Table I. (Continued)

Paper	Methodology	Objective	Substances	Health Outcomes	Region	Unit of Analysis	Specific Methods Employed	Data Sources
Brownstein et al. 52	GA	To understand the spatial distribution of opioid “abuse” at the local level to facilitate public health interventions	Prescription opioids	Substance use disorder	New Mexico	3-digit ZIP code	Spatial Scan statistics (SaTScan)	ASI-MV-Connect®, part of the National Addictions Vigilance Intervention and Prevention Program
Pesarsick et al. 54	GA	To investigate local clustering of nonfatal overdose	All	Nonfatal overdoses	Pennsylvania (Westmoreland, Fayette, and Washington counties)	County	Overall clustering using Ripley’s K function and local clustering using Kulldorff spatial scan statistics with the purely spatial Bernoulli model	Address-level Emergency Medical Services data
Albright et al. 50	SR and GA	To determine hotspots of OUD among veterans at county level	Opioids	OUD	Alabama	County	Mixed-effects generalized linear model and hotspot analysis using local indicators of spatial association (LISA) analysis with local Moran’s I	NSDUH and the American Community Survey
Basak et al. 53	GA	To identify hotspots for opioid use and opioid prescription claims	Prescription opioids	Prescription opioid claims, anomalous prescription opioid providers	Virginia, West Virginia, and North Carolina	County	Spatial scan statistics (Poisson version of the Kulldorff scan statistic), network scan statistic and logistic regression	Medicare Provider Utilization and Payment Data; Part D Prescriber Public Use File ACS census datasets,
Marotta et al. 48	GA	To determine spatial patterns in the distribution of county-level rates of overdose deaths in New York State and associations between prescriptions for opioid pain relievers, race, and overdose deaths from 2013–2015	Any drug, all opioids, prescription opioids, heroin, methadone, synthetic heroin excluding methadone	Overdose deaths	New York state	County	LISA: Global and local Moran’s I, Bayesian smoothed rates	New York State Department Of Health
Stopka et al. 49	GA	To conduct spatial epidemiological analyses of novel comprehensively linked data to identify overdose and potentially inappropriate opioid prescribing practices hotspots	Prescription opioids	Overdose deaths	Massachusetts	Zip codes	Local Indicators of Spatial Autocorrelation (LISA); Getis-Ord Gi*	Administrative datasets, Prescription monitoring program, medical claims, vital statistics, medical examiner data

(continued on next page)

Table I. (Continued)

Paper	Methodology	Objective	Substances	Health Outcomes	Region	Unit of Analysis	Specific Methods Employed	Data Sources
Cerda et al. 55	GA	To determine whether prescription opioid poisoning hospital discharges spread across space over time. To determine the locations of hot-spots of prescription opioid-related hospital discharges	Opioids	Hospital discharges for prescription opioid poisoning	California	Zip Codes	Bayesian space-time misalignment Poisson model	Poisoning data from the California Office of Statewide Health Planning and Development
Hernandez et al. 51	GA	To analyze the spatio-temporal nature of the opioid epidemic in Ohio from 2010-2017	Opioids	Overdose	Ohio	Identified clusters	SaTScan, Bayesian Poisson	Ohio Department of Health, ACS
Cao et al. 56	GA	To investigate how Twitter data can be used to detect ZIP level characteristics associated with spatio-temporal patterns of Emergency Department (ED) patients	All drugs	Drug use related ED visits	Maryland	ZIP code	Spatial lag regression with fixed time effect and geographically and temporally weighted regression (GTWR)	ACS, electronic records from ED patients, geo-tagged Twitter data
Bloodborne infections: HIV and HCV								
Des Jarlais et al. 32	Descriptive statistics	To identify potential geographic hot-spots of HIV and HCV transmission	IDU	HIV and HCV	New York City	Zip Code	GIS Maps, descriptive statistics	Mount Sinai Beth Israel drug detoxification and methadone maintenance programs
Van Handel et al. 62	SR	To identify US counties at risk of experiencing an injection drug use (IDU)-related HIV outbreak similar to that in Scott County, Indiana	Access to prescription opioids included as independent variable (IV)	HCV infection as a proxy for IDU and risk of HIV/HCV outbreaks	US	County	Backward stepwise multilevel Poisson regression	NCHS, National Vital Statistics, NCHS Urban Rural Classification, DEA ARCOS, CMS National Provider Identification, ACS, HSIP 2012 Gold, ESRI maps, US Census, SAMHSA data
Rickles et al. 63	SR	To identify Tennessee counties at risk of experiencing an IDU-related HIV outbreak	Prescription opioid rates, heroin- and opioid-related mortality, and substance-related crimes as IVs	HCV infection as a proxy for IDU	Tennessee, US	County	Principal component analysis, factor analysis, stepwise Poisson regression	NCHS, National Vital Statistics, NCHS Urban Rural Classification, DEA ARCOS, CMS National Provider Identification, ACS, HSIP 2012 Gold, ESRI maps, US Census, SAMHSA Data, Tennessee Department of Public Health, Robert Wood Johnson Foundation County Health Rankings and Roadmaps

(continued on next page)

Table I. (Continued)

Paper	Methodology	Objective	Substances	Health Outcomes	Region	Unit of Analysis	Specific Methods Employed	Data Sources
Sharareh et al. 64	SR & GA	To identify Utah counties at risk of an HCV outbreak and identify factors related to IDU	Prescription opioid rates, heroin- and opioid-related mortality as IVs	HCV infection as a proxy for IDU	Utah, US	County	Negative binomial regression, spatial autocorrelation, optimized outlier analysis (GIS)	NCHS, National Vital Statistics, NCHS Urban Rural Classification, DEA ARCOS, CMS National Provider Identification, ACS, HSIP 2012 Gold, ESRI maps, US Census, SAMHSA Data, Utah Department of Public Health, Robert Wood Johnson Foundation County Health Rankings and Roadmaps
Prediction Overdose Sumetsky et al. 67	SR & GA	To predict overdose deaths by county by incorporating epidemic onset, growth rate and "limited capacity" (ie size of susceptible population)	Opioids	Overdose deaths	North and South Carolina	County	Standard log-linear vs log-logistic Bayesian hierarchical Poisson conditionally autoregressive spatial models	National Vital Statistics System (NVSS), Geolytics, US Census County Business Patterns, Healthcare Cost and Utilization Project's State Inpatient Database, State government hospitalization data
Cooper et al 69	SR	To describe dynamics of opioid sub-epidemics and identify states with highest elasticity (ie rate of change over time)	Heroin, semi-synthetic and synthetic opioids	Overdose deaths	US	State	Third degree polynomial	Kaiser Family Foundation analysis of CDC National Vital Statistics
Campo et al. 72	ID	To assess changes in IDU through estimating yearly overdose death rates at state and county levels, as well as monthly estimates at state level	All	Overdose deaths	US	State and county	Machine learning model (Extremely Random Forest)	Google searches, CDC Wonder
Young et al. 70	ID	To predict heroin related emergency hospital visits across nine metropolitan statistical areas in the United States	Prescription and non-prescription opioids	Emergency department admissions for heroin overdoses	Nine metropolitan statistical areas	Metropolitan Statistical Area	Linear mixed model	SAMHSA, Google Trends

(continued on next page)

Table I. (Continued)

Paper	Methodology	Objective	Substances	Health Outcomes	Region	Unit of Analysis	Specific Methods Employed	Data Sources
Perdue et al. 71	ID	To utilizes internet search data to predict novel substance use	Novel substances (Adderall, Salvia, Snus, synthetic marijuana, bath salts)	Prevalence of novel substance use	US and Central Appalachia	National and regional	Pearson's correlation	Monitoring the Future, Google Trends
Pitt et al. 73	Dynamic Modeling (DM)	To predict the impact of public health policies on OUD incidence and associated mortality	Prescription and non-prescription opioids	Prescription opioid and heroin use disorder and associated deaths and QALYs	US	National	Dynamic compartmental model	NSDUH, NESARC, Rand Corporation, Peer reviewed literature
Chen et al. 74	DM	To project effects of interventions to lower prescription opioid "misuse" on opioid overdose deaths from 2016 to 2025	Prescription opioids, illicit opioid	Number of people using prescription opioids non-medically and using illicit opioids. Opioid overdose deaths	US	US	Dynamic compartmental model	NSDUH, CDC WONDER
Ballreich et al. 75	DM	To predict overdose death rates and potential intervention impact up to 2029	Opioids	Overdose deaths	US	US	Dynamic decision analytic Markov model	US Census, CDC WONDER, NSDUH, National Epidemiologic Survey on Alcohol and Related Conditions—III
Bloodborne infections: HIV and HCV								
Fraser et al. 77	DM	To determine the required scale-up of HCV treatment with or without scale-up of HCV prevention interventions to achieve a 90% reduction in HCV chronic prevalence or incidence by 2025 and 2030 in a rural US setting	All drugs (IDU)	HCV	Scott County, Indiana	County	Dynamic compartmental model	State surveillance and contact-tracing data collected during the 2014-2015 HIV-outbreak, drug treatment episode dataset (TEDs) for Indiana, acute HCV case reports from Indiana
Gonsalves et al. 11	DM	To evaluate the response to the 2015 Scott county HIV outbreak among PWID	Prescription opioids (Opana) and heroin	HIV infection	Scott county, Indiana	County	Dynamic compartmental model	Indiana HIV outbreak team, CDC HIV incidence study, peer reviewed literature
Goedel et al. 76	DM	To determine the impact of syringe service programs on the 2015 Scott County HIV outbreak among PWID	Opioids	HIV infection	Scott County, Indiana	County	Agent-Based Model	ACS, CDC Wonder, peer reviewed literature
Marks et al. 79	DM	To predict the impact of opioid agonist therapy scale up on injection drug use initiation	Opioids	Injection drug use	Generic North American setting	City	Dynamic compartmental model	PRIMER study, NSDUH, peer reviewed literature
Bobashev et al. 78	DM	To predict how variations in heroin source-forms, needle type, and injection practices drive IDU-related HIV transmission	Heroin	HIV Infection	US	East and West	Agent-Based Model	Peer-reviewed literature

34 They used a natural language processing (NLP) approach to identify opioid-related tweets. Their findings indicate that the annual number of heroin- and synthetic opioid-related Twitter posts were associated with annual heroin- and synthetic opioid-overdose death rates, respectively.

34 Chary et al used NLP (semantic distance specifically) to identify Tweets related to the non-medical use of prescription opioids (NMPO) from 2012 to 2014 and showed these correlated with 2013-2015 NSDUH state level data on NMPO across time, with the strongest correlation found among those aged 18-25.

35 In a recent study, Sarker et al used machine learning and NLP to identify Twitter chatter related to “self-reported opioid abuse or misuse” in Pennsylvania from 2012 to 2014.

36 They showed the latter was significantly correlated with yearly opioid related overdose deaths at county level across the three years and that it was also correlated with four pertinent measures of drug use from the NSDUH at substate level, although significance was not attained due to small sample sizes. This holds promise for the real time monitoring of opioid use and associated overdose and could therefore contribute to the early detection of overdose outbreaks. Another type of internet data explored has been online news. Hswen et al investigated the correlation between opioid related (Google) news and opioid related overdose deaths across the US and found that geographical variations in these two outcomes were not consistent, suggesting news will unlikely prove useful for the real-time monitoring and detection of opioid related deaths, but instead might reflect or affect public opinion and policy decision making regarding OUD.

37

Bloodborne-infections. Collecting information on substance use characteristics is key to enabling ORE detection and this is particularly true in the context of infectious disease surveillance. For example, Fitzmaurice et al. utilized national HIV surveillance data to identify injection drug use (IDU) related HIV outbreaks using a heuristic approach.

28 They examined three years (2013-5) of HIV diagnosis data, generating an average annual HIV diagnosis rate and corresponding standard error at the state- and county-levels. Then, they examined HIV diagnosis rates for the subsequent year (2016) and generated

“alerts” for all localities whose HIV diagnosis rate was at least two standard deviations greater than that localities average rate for the prior three years. Another study showed that enhanced HCV surveillance implemented in New York state and prioritizing follow-up of positive laboratory markers for HCV infection among persons aged <30 years allowed researchers to identify a cluster of HCV transmission related to IDU among people from one particular county who had attended the same high school.

38 While effective, enhanced surveillance and comprehensive outbreak investigations involving contact tracing and data linkage generally rely on additional funding and are labor intensive.

Several studies applied phylogenetic and/or molecular analyses to the early detection of OREs of HIV and HCV. Such analyses use genetic data of pathogens (either nucleotide sequences or genotypes) to help identify persons involved in the same chain of disease transmission and add value to conventional contact investigation. They may help identifying unknown or unusual transmission settings or factors, uncover inter-jurisdictional transmission and identify additional persons involved in an outbreak.

39 Oster et al. analyzed HIV-1 nucleotide sequences, reported by the USA National HIV Surveillance System.

40 Using the HIV-TRACE software tool

41 they identified 103 growing clusters of rapid HIV transmission between 2013 and 2015, with genetic viral distances lower than 1.5%.

40 This study failed to identify the Scott County, Indiana HIV outbreak associated with injecting Opana for the simple reason that the state of Indiana does not participate in the National HIV molecular surveillance program (further highlighting the importance of epidemiologic surveillance infrastructure). The Scott County outbreak was however retrospectively investigated using phylogenetic analyses as reported in Peters et al and Campbell et al, confirming that nearly all HIV cases belonged to the transmission cluster associated with injecting Opana and were not linked to other HIV infections in Indiana.

10

42 Alpren et al. used phylogenetic analyses to study an HIV outbreak among people who inject drugs (PWID) in Lawrence and Lowell,

Translational Research

Massachusetts by identifying transmission networks.

9 Similarly, routine HIV screening at the first legal syringe services program (SSP) in Florida led to the identification of ten anonymous HIV seroconversions.

43 Through phylogenetic analysis, they were able to identify both drug use and sexual HIV transmission networks linking SSP clients with individuals outside of this group.

43 Falade-Nwulia et al. found that, among HCV-positive PWID in Baltimore, women and HIV-positive individuals were more likely to be associated with an HCV transmission cluster,

44 showing that phylogenetic approaches can help identify individuals at higher risk of infection. Further Zhou et al present a novel molecular analysis to identify recent infections and drug resistance mutations among new HIV diagnoses.

45 Systematically applying this method to new diagnoses could help identify and interrupt new transmission clusters. While the cost of sequencing has decreased over time, it is still not routinely implemented across settings due, in part, to high costs. Interestingly, routine molecular analyses could lead to more cost-effective screening and contact tracing strategies through prioritizing specific transmission clusters.

Tuberculosis. In a CDC outbreak investigation, the review of medical records and interviews identified substance-use as a factor associated with a growing cluster of TB transmission in Florida.

46 In a study of genotype clusters of tuberculosis, Althomsons et al suggested that “routinely reported [genotype] data may identify small clusters that are likely to become outbreaks and which are therefore candidates for intensified contact investigations.”

47 The early identification of these growing clusters is fundamental to targeting prevention campaigns.

Risk assessment. Risk assessments aim to identify risk factors for specific health outcomes and to characterize their distribution in order to identify communities at higher risk of experiencing outbreaks of the health outcome of interest. As could be expected, many of the studies identified under this objective employed geospatial methods.

Overdose. Geospatial approaches have been applied to substance use surveillance data to identify localities at high-risk (ie “hotspots”) of increased overdose

burden. Marotta et al. applied exploratory spatial data analysis (ESDA) to identify hotspot clusters of opioid overdose in the state of New York.

48 ESDA involves first visualizing the data (ie mapping overdose death rates by county), performing a test of global spatial auto-correlation (via Moran’s I), and then identifying clusters of counties applying Local Indicators of Spatial Association (LISA) analysis.

48 Stopka applied a similar approach to identify clusters of opioid overdose and high rates of prescription opioid administration among localities (by ZIP code) across Massachusetts.

49 Albright et al also followed a similar approach to identify hotspots of OUD among veterans in Alabama.

50 Hernandez et al. and Brownstein et al. both applied the spatial scan statistic (SaTScan) approach to analyze the spatial distribution of opioid overdose across Ohio and of opioid use across New Mexico, respectively.

51.

52 Basak et al also applied this method to identify hotspots for opioid use and opioid prescription claims in Virginia.

53 Pesarsick et al investigated local clustering of non-fatal overdose in southern Pennsylvania using Emergency Medical Services (EMS) data from cases in which naloxone was administered and an improvement in the Glasgow coma score was recorded. They identified local clusters through applying the Kulldorf scan statistic.

54 These EMS data allow for near real time risk assessments and service adjustment.

Geospatially driven regression strategies can also be applied to identify locality-specific factors associated with a higher risk of overdose mortality occurring. Cerda et al. applied a spatio-temporal Bayesian Poisson model to examine the relationship between zip code level factors and prescription opioid overdose-related hospitalizations.

55 They found that localities in California with greater pharmacy density, greater prevalence of manual labor jobs, and lower income were all associated with higher rates of prescription opioid hospitalizations.

55 Cao et al also applied geospatial regression strategies to characterize the ZIP code-level relationship in Maryland between drug-related emergency departments visits and series of predictors including geolocated Twitter data on sentiment related to crime, drug use and

depression, ACS data on poverty, unemployment, housing prices, and education level.

56 Their findings suggest that this approach can be used to identify areas at higher risk of experiencing drug use related health harms.

The consistent measurement of outcomes of interest (and factors associated with them) across geographic settings and over time represents a primary limitation to the effective implementation of geospatial analyses. Geospatial analyses often rely on data collected by multiple disconnected organizations with wide variations in data collection and reporting practices. For example, it would certainly be of interest to incorporate the geospatial distribution of synthetic opioid overdose throughout the US into a risk assessment model – unfortunately, due to inconsistent reporting of the involvement of specific drugs in mortality records by local coroner’s offices,

57 such an endeavor is not feasible. A CDC report of overdose deaths found that only 20 of 50 US states from 2013-2017 reported consistent and trustworthy synthetic opioid overdose counts.

58 In addition, it is important to acknowledge that geospatial methods have so far predominantly been used for explanatory rather than predictive purposes, and therefore further validation of predictive adaptations is needed.

Wastewater epidemiology is emerging as a potential useful tool to monitor and characterize geographic differences in substance use patterns. Endo et al present results from a study involving robotic wastewater sampling across residential manholes in an urban municipality in North Carolina and targeted mass spectrometry to detect opioids, including illicit and prescription opioids as well as naloxone and buprenorphine.

59 No correlation was found between total opioid exposure in each catchment area and the number of overdoses, but the latter did correlate with both naloxone and buprenorphine exposure. Both Duvall et al. and Gushgari et al. applied a similar approach to identify the presence of opioids in urban wastewater.

60

61 Gushgari et al. found that opioid presence in wastewater in two US cities was associated with future opioid overdose deaths, indicating it may be a valuable risk assessment tool.

61 Such technology can be used to quantify levels of different opioids in a community, thereby identifying areas at higher risk for overdose, as well as the presence of new or more potent opioids.

Bloodborne infections. One of the most cited studies, published by Van Handel et al

62 as part of a CDC initiative, applied a multi-level Poisson regression to rank US counties at highest risk of an IDU-related HIV outbreak, using HCV incidence as a proxy measure for IDU prevalence. They calculated a risk score for future potential HIV transmission risk in each county based on the model’s coefficients and on the distance to high HIV prevalence counties.

62 They identified 220 counties at highest risk of IDU-related HIV outbreaks and these findings have been used to inform funding allocation for substance use treatment and harm reduction programs. Rickles et al replicated and expanded on the Van Handel risk assessment approach, focusing solely on counties in Tennessee.

63 They incorporated a wider variety of potential indicators and applied both factor analysis and principal component analysis in order to select variables for their final model. The Rickles study found counties in western Tennessee to be at highest risk, whereas the Van Handel study had identified counties in eastern Tennessee, indicating the importance of undertaking locality-specific endeavors to understanding the harms of OREs. Similarly, Sharareh et al expanded on Van Handel et al. to identify small areas (combining a number of ZIP codes) at high risk of HCV infection associated with IDU in Utah.

64 They complemented the study with an optimized outlier analysis, a geographic information system method, to identify HCV hotspots when accounting for HCV rates in neighboring small areas. Their results differed from those of Van Handel, which they argue is likely due to their analysis being more recent (2019 vs 2016) and using different methods for handling regions with limited data.

64

While its impact has been influential and has driven resource allocation, the approach employed by Van Handel et al is subject to important limitations that need be discussed. First the vulnerability score calculated for each county was defined cyclically: they trained their model with HCV as the outcome and then they applied the model on the same data it was trained on to generate vulnerability scores. As such, is unclear if their approach was optimal compared to simply using HCV incidence as the vulnerability score. It is also unclear if HCV incidence represents an accurate proxy of IDU (as partly

dependent on chronic prevalence and its measurement is uneven) and, therefore, if their approach effectively captured the at-risk population in each county. Next, they applied no validation strategy, internal or external, to their findings. Without such validation the accuracy of their risk assessment is unclear, which is important both in the context of resource allocation based on the results and in determining whether stakeholders should implement this strategy again in the future. Importantly Van Handel et al

62 and related studies use the proportion of people who are white as a key model predictor. The opioid crisis has long been characterized primarily affecting white people in the US.

65 However, this was related to preferential pain management treatment (and over-prescription of opioids) among white people.

65 Robinson et al

66 note that using race as a predictive factor, when it is systemic racism driving the observed pattern, risks reinforcing the impact of systemic racism (in this instance, by overestimating the harm of future OREs in localities with a higher proportion of white people and potentially depriving ethnically diverse urban areas from needed resources). Finally, it is unclear if applying a regression strategy without the ability to capture geospatial correlations resulted in biased findings. Given the proliferation of geospatial modeling strategies, identifying the appropriate method to rigorously incorporate geospatial effects is challenging. This concern also further highlights the importance of including a strategy to validate model performance.

At a smaller geographic scale, Des Jarlais et al assessed location of substance use and injection risk behaviors among PWID accessing substance use treatment centers in New York City to determine zip codes considered “hotspots” of HIV and HCV risk.

32 This type of local level study can inform the prioritization of outreach efforts such as mobile HIV/HCV testing or SSP, as well as naloxone distribution efforts. However, they are difficult to standardize as they rely on treatment (or harm reduction) centers’ capacity. Coordination efforts across these centers through a central organization and through dedicated funding mechanisms could address this issue.

Prediction. Prediction goes one step further than risk assessment by quantifying the future incidence of a health outcome of interest among specific populations or geographical locations. This is

challenging in general and particularly in the context of OREs given the rapidly changing dynamics of drug markets and drug use itself. However, estimating future outcomes requires the systematic and rigorous analysis of observed patterns and their underlying mechanisms, thereby leading to a better understanding of OREs.

Overdose. An interesting study led by Sumetsky et al tested the performance of two statistical methods (standard log–linear vs log–logistic Bayesian hierarchical Poisson conditionally autoregressive (CAR) spatial models) in predicting overdose deaths by county in two states from 2001-2014.

67 They generated a county-level “carrying capacity” (ie, the number of people at risk of overdose in given county) based on a set of factors, a well-founded decision given that a majority of individuals are not at risk of overdosing. However, one factor informing this carrying capacity was the proportion of county population that was white, replicating the issue highlighted by Robinson above and potentially limiting its validity to the prescription opioid wave of the opioid crisis. It is important to capture risk factors across different classes of opioids and other drugs as their impact is driven by different mechanisms and affects different populations.

68

A recent study by Cooper et al. used three-degree polynomial functions to investigate fatal overdose dynamics from 2012 to 2016 by state, disaggregating rates by heroin, semi-synthetic and synthetic opioids.

69 They identified states with highest elasticity (ie rate of change over time) for each of the opioid sub-epidemics. Such an approach is useful for characterizing changes in overdose rates over time (ie, elastic vs inelastic) and displays the potential benefit of applying non-linear functions for predictive purposes. Further, calculating overdose rate elasticity may be applied within a broader predictive framework, as counties with greater elasticity may be subject to greater fluctuations in overdose death. Based on the write-up, it is not readily clear how the model coefficients were reached and accuracy of predictive models are not reflected on – making it challenging to evaluate how this study should be adapted for future use. Further, while it is of interest to consider non-linear regression, it is worth investigating whether the equation they used, or a single equation structure generally, could explain

changes in overdose death rates across all localities.

Two studies utilized Google Trends data to predict opioid use trends and related outcomes.

70

71 Google Trends analyzes a portion of Google web searches over a period of time and normalizes the data to compare trends of different search terms from the same region during the same period. Young et al built a linear mixed model regression using opioid related search terms as the independent variable and opioid related emergency department admissions as the dependent variable between 2005 and 2011 in 10 metropolitan statistical areas to then predict opioid related hospital admissions for the next two 3-month periods.

70 Perdue et al also used Google Trends to investigate whether searches for novel drugs were associated with prevalence of use of these novel drugs, using the “Monitoring the Future” survey as their validation dataset, finding that internet searches for novel drugs were associated with novel drug use in teenagers.

71 Young et al display that Google Trends may be effectively used in a predictive analysis. One limitation of their analysis is that it is restricted to metropolitan statistical areas, as opposed to smaller localities such as counties in rural areas that may need to mobilize a preventative response – unfortunately, Google Trends are not available at county level. Recently, Campo et al presented a novel machine learning algorithm informed with Google search terms related to drug use to predict next year’s overdose death rates at county level across the US from 2005 to 2017, as well as monthly estimates at state level across this period.

72 They showed that the model has good accuracy based on mean average error (with variation across states) and that for 2017 (which data was not used for model training), it identified 75 of the top 100 counties based on overdose death rates. Both the Young et al and Campo et al studies display that incorporating Google Trend data to inform predictive models holds the potential to improve overall model performance.

Pitt et al used a compartmental dynamic model of opioid use to estimate future overdose mortality rates at national level, finding that in the absence of further interventions, 510,000 opioid-related deaths would occur between 2016 and 2025.

73 Unlike statistical models, which relate a set of predictors with the outcome of interest through capturing their association, dynamic models explicitly represent the mechanisms generating a specific outcome, such as progression through different stages of opioid use associated with differential risk of fatal overdose, or contact patterns leading to infectious disease transmission. Dynamic models can be split into two main categories: compartmental models, which represent groups of individuals and their “average behaviors” and agent-based models, which explicitly represent each individual and their behavior. Chen et al also developed a compartmental model differentiating non-medical use opioid prescription (with and without disorder) and illicit opioid use, and estimated that in the absence of interventions 700,400 opioid-related deaths would occur between 2016 and 2025.

74 This is higher than estimated by Pitt et al in part due to the assumption that both the incidence of illicit opioid use and its lethality would follow an upward trend based on observed increases in synthetic opioid related deaths. They found that eliminating the non-medical use of prescription opioids would have a marginal effect on the overdose epidemic, indicating multi-pronged interventions are needed. Ballreich et al used a detailed dynamic model including 32 compartments which disaggregate the population by type of opioid used and whether it was initiated through prescription opioid use or not, as well as seven different medication assisted treatment (MAT) states, to simulate the opioid epidemic in the US between 2020 and 2029.

75 They estimated 484,429 opioid related deaths across this time period in the absence of further interventions. Similar to Pitt et al and Chen et al, they found that reducing opioid prescribing (40% reduction) would have a small effect on the epidemic, while scaling up naloxone access (leading to a 19% reduction in the overdose fatality rate) would prevent about 15% of deaths and increasing MAT access (tripling uptake) would have the greatest impact with a 25% reduction in mortality. Each of these compartmental modeling studies have been implemented at the national level, and therefore fail to account for heterogeneity in opioid prescriptions, substance use behaviors and OUD treatment access across the country. As such they can only provide

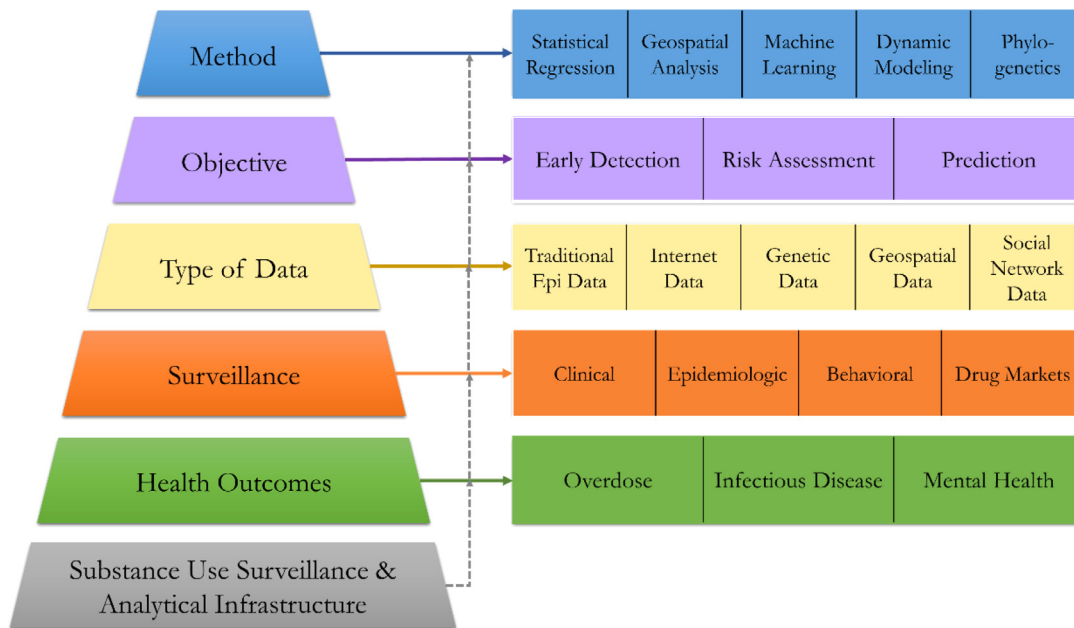


Fig 2. Multidisciplinary framework to enable the early detection, risk assessment and prediction of opioid use related epidemics (OREs). Legend: The diagram should be read from the bottom upwards, with each layer corresponding to a different component determining the choice of method in a step by step manner: 1) health outcome of interest, 2) type(s) of surveillance data available, 3) characteristics of the collected data, 4) objective, 5) analytical method. First, the bottom layer refers to the surveillance and analytical infrastructure – these are pieces that must be in place to collect data and to analyze it. Surveillance, in the context of these studies, is often undertaken by public health agencies and institutions such as the CDC. The next step is identifying available ORE-driven health outcomes. Then, to identify available measures of the outcome and potential predictors, researchers should ask which types of surveillance data are available to them. Clinical data (ie, EMS, hospital records, death records), epidemiologic data (ie, poison call centers, harm reduction services, 311 calls), behavioral data (ie, observational studies, internet data), and drug market data (ie DEA, drug sample testing, wastewater sampling) represent four types of data of importance to consider. After identifying potential data sources, the type of data available should be characterized as it can inform the study objective. We have identified five types of data (traditional epi, internet data, genetic data, geospatial data, and social network data). Prior to selecting the method to employ, we recommend choosing an overarching objective. Failing to do so can lead to confusion amongst the research team about the underlying purpose of a study. For example, certain approaches may be well-suited for risk assessment but not prediction, and a failure to explicitly determine the study objective prior to choosing a method may result in the implementation of an inappropriate analytic approach. Finally, the research team can select the method(s) best suited to their data and objective. In addition, multiple methods can be used in parallel to increase the sensitivity and accuracy of findings.

general insight and guidance about the likely course of the epidemic and appropriate responses. In addition, assumptions regarding future trends in overdose deaths between the different studies lead to significantly different estimates and therefore to different implications for interventions (such as for future treatment need). That said, comparing outputs across modeling studies and identifying the reasons behind discrepancies is an effective way to better understand epidemic dynamics.

Bloodborne infections. While Campo et al used overdose deaths as their outcome of interest, they present this as a proxy for IDU and state their model could be

used to predict or raise alert about the risk of IDU related outbreaks of blood borne infections. Indeed, they highlight that Scott County was identified by their model as the top Indiana County for overdose deaths in 2011 (and 12th highest US county), which coincided with the HCV and HIV outbreaks, suggesting that these could have been prevented or mitigated to a large extent had these signals been attended to. Gonsalves et al applied a compartmental model to retrospectively evaluate the response to the Scott County HIV outbreak.

11 They found that implementing the response in April 2011 (instead of in early 2015), just after HCV outbreaks were identified in several

counties in the state, could have limited the outbreak to 10 cases. Such a model could be used prospectively in the event of an outbreak to inform the response. Goedel et al. applied an agent-based modeling approach, projecting that the proactive implementation of SSP in Scott County would have decreased HIV incidence by 90%.

76 Fraser et al also used data from the Scott County outbreak to predict future HCV incidence and estimate the directly acting antivirals (DAA) treatment coverage needed to achieve a 90% reduction in chronic HCV prevalence and incidence by 2025 and 2029, either in isolation or in combination with MAT and SSP.

77 They found that 159 per 1,000 PWID would need needed to be treated every year and this would be halved if MAT and SSP were scaled up to 50% of PWID. While these studies rely on detailed data collected during a specific infectious disease outbreak, which is not available across settings, they are useful case studies to demonstrate the importance of scaling up harm reduction and treatment services early in the context of OREs and also provide concrete estimates of treatment coverage needed to achieve HCV elimination goals and will be useful to similar settings. Bobashev et al employed an agent-based modeling approach to explore the impact of different types of heroin (powder vs black tar), different types of needles (high vs low dead space), and various injection practices (such as syringe sharing) on HIV incidence.

78 This approach highlights both the feasibility and necessity of incorporating variations in drug type (driven by drug market availability) and route of administration when attempting to predict OREs. In a more experimental fashion, substance use has been modeled as a communicable behavior. Marks et al. estimated the future incidence of IDU in a generic North American setting, assuming different coverages of MAT, which has been associated with a lower rate of assisting others to initiate IDU.

79 They found that a 60% MAT coverage could translate in a 23% reduction in the annual rate of IDU initiation.

IMPLICATIONS

Proposing a framework for ORE surveillance. Based on these findings we propose a **framework for the enhanced surveillance and forecasting of OREs** which integrates the different methodological approaches identified to enable the early detection, risk assessment and prediction throughout the epidemic's course as shown in

Figure 2. To develop a multidisciplinary ORE surveillance and analytic infrastructure, it is essential to consider the multiple health outcomes arising from opioid use, including overdose, infectious diseases and mental health disorders (which we attempted to include in our search but failed to identify relevant studies) and to maximize the different types of surveillance systems associated with each.

Types of surveillance data. Clinical surveillance remains the cornerstone of OREs' early detection and local authorities have merged a variety of data sources to increase the sensitivity of their surveillance efforts. Indeed, the ability to disaggregate routine infectious disease, hospital admissions, EMS records and death certificates by substance use characteristics is crucial to detecting OREs. Enhancing this information with epidemiological surveillance data, including poison call center data, as well as data from substance use treatment and harm reduction organizations can further minimize time to outbreak detection. 28

33 Substance use behavioral surveillance had until recently been grounded on routine population surveys such as the NSDUH, as well as opioid prescription and OUD treatment records.

Recently, internet-based surveillance of opioid use related searches and social media posts has been proposed as an alternative to traditional surveillance approaches, potentially enabling the real time monitoring of substance use trends. It has been shown to correlate well with NSDUH estimates as well as with overdose outcomes. While representativeness and validity remain difficult to assess, especially considering lower internet access among some people suffering from OUD such as those experiencing homelessness, this is a promising field.

Drug market surveillance, through drug seizures, drug testing services and wastewater testing, can provide the often-missing part of the equation. Ideally, these different sources of data would be easily accessible in real-time to allow for continuous "merged-monitoring" analyses.

Translational Research

Table II. Presentation of the overarching methodologies identified in the review, the objectives (early detection, risk assessment, prediction) that they can be employed for, and the strengths and limitations inherent to each methodological approach

Methodology	Objective	Strengths	Limitations
Statistical Regression	Risk Assessment and Prediction	<ul style="list-style-type: none"> • Can identify risk factors which are predictive of OREs and related harms • Can be used to rank localities based on risk of ORE • Can be used to predict harm associated with ORE 	<ul style="list-style-type: none"> • Require consistently measured data across both time and locality • Unable to account for complex pathways and multi-stage processes which may better characterize OREs • Ability to make long-term predictions is generally limited by data availability and structure of data
Geospatial Analysis	Risk Assessment and Prediction	<ul style="list-style-type: none"> • Can account for the geographic spread of OREs • Can account for geographic variations in the relationship between predictive factors and OREs 	<ul style="list-style-type: none"> • Similar limitations to statistical regression, as they are often employed together • Poor reporting in specific localities (ie a county that does not report fentanyl deaths) may lead to underestimates of the importance of the role of geographic spread
Internet Data Analysis	Early Detection, Risk Assessment, and Prediction	<ul style="list-style-type: none"> • Internet data is easily accessible • Internet data can be collected in real-time (ie, social media posts, Google Trends) • Huge quantities of data can be collected 	<ul style="list-style-type: none"> • Quality and applicability of data for study purposes will generally be poor (involves many steps of data cleaning) • Ability to identify data specific communities at high risk is challenging (ie, difficult to identify posts and content corresponding to vulnerable groups such as PWID) • Not representative of populations at highest risk which do not have internet access • Without qualitative and culturally-specific expertise (ie, vernacular and social media norms), identifying relevant data points may be challenging or even impossible
Dynamic Modeling	Prediction	<p>Can be used to:</p> <ul style="list-style-type: none"> • make short and long-term projections of ORE impact • simulate population dynamics at the individual level (ie agent-based modeling) • predict the efficacy of interventions aimed at attenuating harms of OREs • predict the impact of changes to drug markets • Can incorporate qualitative data in model design 	<ul style="list-style-type: none"> • Over-simplification of real-world phenomena • Data thirsty, limiting its application to specific localities or, on the contrary, to large regions, ignoring heterogeneity • Often rely on published research to inform model parameters, thus are subject to bias introduced by other studies • Data used to inform models is not always specific to the locality the model is being applied to

(continued)

Table II. (Continued)

Methodology	Objective	Strengths	Limitations
Phylogenetics	Early Detection	Can be used to identify: <ul style="list-style-type: none"> • locality-specific outbreaks of ORE driven infections • interpersonal spread of ORE driven infections • drug use social networks • Can be used to confirm the connection between cases arising from a given ORE 	<ul style="list-style-type: none"> • Genetic data is challenging and expensive to collect • Generally, only available for specific localities undertaking phylogenetic driven research and prevention initiatives • Results, generally, cannot be generalized to inform ORE detection, assessment, or prediction in other localities

27 How these data are collected or processed, will determine key attributes, such as the availability and granularity of geospatial, socio-demographic, social network or genetic information.

Types of methods that can be employed. Depending on the health outcome of interest, the surveillance data available and their characteristics, different analytic methods can be used to enable the early detection, risk assessment and prediction of OREs. The studies identified have implemented a variety of statistical regression, geospatial, machine learning, dynamic modeling, and phylogenetic and/or molecular analyses. Some of these methods could in theory be applied towards any of the three objectives, as long as the analysis is designed accordingly.

For example, geospatial methods can be used to detect clusters and/or hotspots for both opioid use and associated harms, or to characterize the geographical heterogeneity in these outcomes thereby contributing to early detection and risk assessment efforts. Furthermore, they can estimate the likely future geographical distribution of opioid use related outcomes, thereby providing predictive capacity.

As illustrated through the identified studies, statistical regression methods have been used in the context of early detection, risk assessment and prediction and are often the method of choice given the limited availability of granular geospatial data. The emergence of big data has led to its replacement by machine learning, using other classification and regression algorithms such as random forests, artificial neural networks, and support vector machines.

Other methods such as phylogenetic/molecular analyses are mostly relevant in the context of early detection (or retrospectively to better characterize an outbreak). They have enhanced traditional infectious disease outbreak investigations and they can also allow the identification of rapidly growing clusters of transmission among people who use drugs.

Dynamic modeling, on the other hand, is mostly used for prediction purposes and can provide guidance on

effective intervention strategies. Additionally, modeling substance use as a communicable behavior could provide early information about the potential spread of a new drug in a specific community and at larger scales. Different methods should be combined to enhance the sensitivity and accuracy of early detection or forecasting analyses and to guide control measures through the epidemic's course.

Further, in [Table II](#), we provide further insight on the different types of methods identified in the review as well as details on when to employ them and their potential limitations.

DISCUSSION

Objective and scope. This narrative review described studies illustrating methodologies which can be employed for the early detection, risk assessment or prediction of future OREs. Further, we have developed a multidisciplinary framework for enhancing ORE surveillance and forecasting endeavors. Few efforts have been made to anticipate OREs and appropriate methods for evaluating their predictive performance have largely been absent. However, recent contributions show promising results and these methodologies represent an interdisciplinary toolkit that can be utilized in the efforts to prevent, identify, and mitigate substance use associated health harms.

Here, we highlight the potential of these different approaches, as well as challenges that must be overcome to undertake this research and we propose good practices (see

[Table III](#) for a summary). Importantly, while we attempted to capture all key terms through our search strategy, we note that we might have missed some relevant papers but expect that the selected sample provides a comprehensive overview of the current tools available across different fields. While we attempted to

Table III. Challenges and good practices when undertaking research aimed at the early detection, risk assessment, and/or prediction of opioid use related-epidemics. Note: Challenges are organized into three categories: challenges related to data; challenges related to methods; and challenges related to dissemination of results

Challenge	Good practices
<p>Data</p> <p>Lack of standardized opioid related epidemic clinical, epidemiologic, behavioral and drug market surveillance</p>	<ul style="list-style-type: none"> • Invest in the systematic monitoring of clinical and mortality outcomes, including the collection of substance use indicators • Implement enhanced surveillance protocols for substance use related infectious disease outbreaks, including strategic phylogenetic analyses • Invest in the monitoring of poison center call data and similar resources and make these available to researchers • Enable the systematic collection and sharing of data at substance use treatment and harm reduction services • Explore alternative methods to collect substance use behavioral surveys, including participatory surveillance through online platforms • Implement and continuously refine internet-based substance use surveillance • Invest in drug market surveillance, implement wastewater sampling, free drug testing services for people who use drugs and standardized testing and reporting of drug seizures by DEA • Invest in estimating the prevalence of locality-specific high-risk drug use indicators like injection drug use to better estimate locality “carrying capacity”
<p>Expertise Across a Range of Disciplines is Required</p>	<ul style="list-style-type: none"> • Focus on building multi-disciplinary workforce • Develop and promote funding mechanisms for multi-disciplinary, multi-method research • Utilize qualitative methods to aid in the design of modeling strategies, which are well-equipped to identify fast moving changes in the landscape of the opioid crisis
<p>Limited and delayed access to valuable data sources collected at national level</p>	<ul style="list-style-type: none"> • Enable timely access to local NSDUH data (while the sampling is designed to provide data representative at state level, the local estimates can provide key insight and uncertainty can be handled rigorously) • Enable timely access to local NFLIS data • Enable timely access to local NEMSIS data • Enable timely access to routine national/state infectious disease surveillance data as well as all other surveillance data available
<p>Disorganized management of available data</p>	<ul style="list-style-type: none"> • Create a log of available data sources, including what geographic divisions and localities they are available and for what time divisions (ie month/year) and periods • Focusing on smaller geographic regions (such as counties or states, instead of the entire US) will likely allow for richer set of available data to be used • Develop and share coding pipelines such that when new data is available, analytic data sets can be updated efficiently • Form partnerships with organizations and other research teams to improve access to data
<p>Methods</p> <p>Explanatory and Predictive Modeling are Distinct Methodological Tasks</p>	<ul style="list-style-type: none"> • Place emphasis on predictive performance, not on explanatory findings • Include evaluation of internal validity (such as cross-validation) to ensure model is not overfit. • Include evaluation of external validity to ensure model performance will meaningfully address research question at hand

(continued)

Table III. (Continued)

Challenge	Good practices
Validity of Internet based substance use surveillance is difficult to ascertain	<ul style="list-style-type: none"> • Internet data represents a valuable resource to complement logistically complex household surveys and research efforts should be dedicated to refining current methods to ensure sensitivity and specificity of their findings in this new field
Results from dynamic models making long term predictions (5+ years) are very sensitive to assumptions on trends in behaviors and sizes of at-risk populations	<ul style="list-style-type: none"> • Be very transparent about assumptions for future behavioral/mortality trends • Carry out sensitivity analyses • Compare to findings with similar studies and encourage model comparison exercises • Invest in estimation of population sizes of people who use drugs
Different methods will provide complementary insights about OREs	<ul style="list-style-type: none"> • Use multiple methods providing short and long term as well as small and large geographical scale predictions to inform decision making, including qualitative research
Predictive Models Can Replicate Systemic Racism (and other -isms) Depending on Variable Selection	<ul style="list-style-type: none"> • Explicitly justify or avoid using immutable, identity-based characteristics (such as race) as factors in predictive modeling endeavors • Exception to this is if the modeling approach employed can effectively account for the mediating and moderating pathways by which such identity-based characteristics impact the outcome of interest.
Dissemination Proprietary Predictive Technologies Mean Every Research Team Must “Re-Create the Wheel”	<ul style="list-style-type: none"> • Publicly share code and software for executing predictive technologies whenever possible. • Publicly share data whenever possible. • We have introduced a platform where research teams can share code and data
No Platform for Researchers to Register Their Future ORE-Related Predictions	<ul style="list-style-type: none"> • Disseminate ORE predictions prior to time being predicted (ie, if a research team predicts ORE outcomes for the year 2022, then they should aim to disseminate the predictions in 2021). • Predictions can be disseminated by any means that allows public to confirm the predictions were made in advance • We have introduced a platform where research teams can register predictions for future-oriented predictions

incorporate mental health outcomes in our search, we regret that no eligible studies were identified, likely illustrating the lesser focus on population level prediction in the field of mental health and the increased complexity in identifying particular conditions, aside from suicide. One study by Yao et al did apply machine learning to Reddit data to investigate suicidality among people who use opioids.

80 It was not included because its main objective was to validate their language processing methods, however, these hold potential for suicide prediction or risk assessment efforts.

Methodological limitations. Given the common application of explanatory models within substance use-related research, it is important to note that predictive and explanatory modeling endeavors are distinct pursuits with their own considerations. Statistical and

geospatial regression methods in the context of OREs have primarily been used for explanatory purposes, to identify factors independently associated with overdose mortality or infectious disease transmission. Only a few studies have extended this for predictive purposes. Indeed, it is sometimes the language, rather than the methods, which determines whether these models are used for explanatory or predictive purposes. Explanation and prediction, however, are distinct endeavors requiring distinct methodological considerations.

81 Introducing rigorous predictive methodologies into ORE prediction studies, such as evaluation of predictive performance, cross-validation, as well as steps to avoid common pitfalls, such as overfitting, is a necessary step. Collaboration with researchers in machine learning and bioinformatics may aid with this introduction.

Innovation in the field of dynamic modeling could enhance the use of these methods to predict the spread of substance use as communicable behavior – theoretical model development by Behrens, Caulkins et al display that dynamic modeling is well-equipped to address the mechanisms underlying social communication of substance use behaviors.

82 This would require more complete data on social and substance use networks; fortunately, this information is becoming increasingly available. As suggested by preliminary studies, the use of Google searches, as well as other internet data including social media data, could provide complementary information to anticipate OREs. However, evidence is limited and experience using these data in the context of flu epidemics' predictions has shown that differentiating (health seeking) behaviors from interest is challenging, requiring more complex methods to implement corrections.

83–86

Data constraints. Implementing these methods is dependent on having timely access to data. The epidemiological surveillance infrastructure to monitor substance use and associated harms must be in place, and structural barriers to access requisite data and to communicate findings across local, state, and federal public health agencies must be removed. In the case of Scott County, no advanced analytic techniques were required to identify the IDU-related HIV outbreak, but a failure to communicate information around the outbreak pushed its detection back by months.

87 Additionally, the non-participation of the state of Indiana in the molecular surveillance program limited the efficiency of the response early in the outbreak.

Timeliness of data is crucial in rapidly evolving OREs. Typically, publicly available epidemiological data related to substance use is limited and outdated. For example, as of February 2021, the Centers for Disease Control and Prevention (CDC) was only reporting county-level overdose death rates through the end of 2019.

88 Assuming no other sources of this data are available, our ability to predict future county-level overdose death rates is limited by the fact that the data from the previous two years is unavailable. This is reflected in the historical timeframes of the predictive studies presented here: Sumetsky, Cooper and Campo (each published in 2020) only predicted overdose deaths up to 2017.

67

69

72

Importantly, drug market data collected by law enforcement agencies is typically unavailable or restricted to public health research teams when it provides valuable information to monitor both changes in drug availability and properties.

68

89 Alternative strategies to monitor drug markets, such as wastewater testing have been investigated and hold promise.

90

Epistemological considerations. Detecting and predicting OREs is also dependent on identifying predictor data which can be used to train and inform analytic approaches. As touched on in the results, though, it is important to consider unintended consequences of certain predictor choice. Robinson et al. highlight how predictive endeavors which use identity-based variables such as race may inadvertently reinforce dynamics of systemic prejudice.

66 This is especially important given that the opioid crisis has long been characterized as primarily affecting non-Hispanic white people in the US.

65 Quite reasonably, many studies we identified used the white proportion of the population in each locality as a key predictor in their analytic approaches. The question that is important to ask is whether it appears that race or systemic racism is driving the disparity. It is argued that the opioid crisis initially impacted white populations more severely because they received preferential pain management care and were essentially over-prescribed opioid painkillers.

65 In the past decade, excessive opioid prescribing practices have been curtailed and it has become clear that the opioid crisis is being felt across all racial and ethnic populations in the US.

91

92 Looking to historically available data, white race will continue to be a significant predictor despite likely changes in the dynamics of the opioid crisis – it is, thus, important that researchers and reviewers engage critically with the appropriateness of its inclusion. Broadly, we recommend that researchers explicitly justify the use of immutable, identity-based characteristics (such as race) in predictive modeling endeavors to protect from the risk of replicating systemic disparities.

Further and more generally, we argue that the use of qualitative methods is informative in driving variable selection, as they can provide rationale for the importance of given factors.

93

Knowledge dissemination. Given the variety of methods, data sources, and potential outcomes to explore, the importance of transparency and knowledge-sharing is crucial to the timely development of OREs' prediction tools. As noted by Woelfle et al, an open science approach can accelerate the time to breakthroughs and findings.

94 Given the multitude of disciplines that need to be coordinated to best address OREs and their harms, precious time can be saved by ensuring that research teams are not unnecessarily replicating one another's work. On top of traditional academic publishing, it will be important to leverage non-peer reviewed publication platforms such as arXiv which can allow for the quick dissemination of novel approaches while still protecting individual researchers' rights to their creative works. Furthermore, predictive modeling efforts are only as useful as they are applicable. Open-source solutions can improve the ability of public health department to implement predictive solutions for themselves.

Finally, it is important to note that the potential for developing effective technologies for predicting ORE and for understanding their expected utility is dependent on transparent platforms which allow research teams to make future-oriented predictions. For example, during the COVID-19 pandemic, modeling teams have attempted to project future incidence and mortality, which organizations such as the CDC have continued to share.

95 As such, we have the ability to examine the accuracy of a wide range of COVID-19 predictive models over space (ie various geographic regions) and time and we can make evaluations of their predictive utility. It is crucial that future-oriented predictions of OREs be transparently aggregated and evaluated. We believe that the introduction of such a platform can work to promote inclusion, collaboration and innovation and will be the most effective way to both ascertain the utility of ORE prediction endeavors and to provide timely evidence-based guidance.

CONCLUSION

This is an interdisciplinary field that has not yet coalesced. It is our intention that this review and accompanying resources can begin the process of bringing together the urgent multidisciplinary action needed to stem the harms of ORE in the US and globally. We have launched a website (<https://www.emergens-project.com/repository>) where research on this topic will be aggregated and invite readers to share their publications. The intention of this site is to create a space where stakeholders can build on one another's work and disseminate approaches and findings. This review and subsequent resources can act as a first step to coalescing this field and accelerating advancements that may be readily implemented.

ACKNOWLEDGMENTS

This project and Dr. Borquez's work is supported by NIH/NIDA grant [DP2 DA049295-01](#). Dr Ciccarone's work is supported by NIH/NIDA grant [R01DA037820](#). Dr. Strathdee's work is supported by NIH/NIDA grant [R01DA049644](#). Dr. Johnson's work is supported by NIH/NIDA grant [T32 DA023356](#). This research was supported by funds from the California HIV/AIDS Research Program of the University of California, Grant Number [[OS17-SD-001](#)]. All authors have read the journal's policy on conflicts of interest and have no conflicts of interest to disclose. All authors have read the journal's authorship agreement and confirm that the manuscript has been reviewed and approved by all named authors. No editorial support was required in the preparation of this manuscript.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi: [10.1016/j.trsl.2021.03.018](https://doi.org/10.1016/j.trsl.2021.03.018).

REFERENCES

1. Volkow ND, Blanco C. The changing opioid crisis: development, challenges and opportunities. *Mol Psychiatry* 2020. <https://doi.org/10.1038/s41380-020-0661-4>.
2. Ciccarone D. Fentanyl in the US heroin supply: A rapidly changing risk environment. *Int J Drug Policy* 2017;46:107–11.

<https://doi.org/10.1016/j.drugpo.2017.06.010>.

- Compton WM, Jones CM, Baldwin GT. Relationship between nonmedical prescription-opioid use and heroin use. *Longo DL, ed. N Engl J Med* 2016;374:154–63.

<https://doi.org/10.1056/NEJMra1508490>.

- Kanouse AB, Compton P. The epidemic of prescription opioid abuse, the subsequent rising prevalence of heroin use, and the federal response. *J Pain Palliat Care Pharmacother* 2015;29:102–14.

<https://doi.org/10.3109/15360288.2015.1037521>.

- Jalal H, Buchanich JM, Roberts MS, Balmert LC, Zhang K, Burke DS. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science* (80-), 36; 2018. 2018.

<https://doi.org/10.1126/science.aau1184>.

- Hedegaard H, Miniño AM, Warner M. Drug overdose deaths in the United States, 1999–2019. *NCHS Data Brief* 2020:

<https://www.cdc.gov/nchs/data/databriefs/db394-H.pdf>. Accessed August 1, 2020.

- Wilson N, Kariisa M, Seth P, Smith H, Davis NL. Drug and opioid-involved overdose deaths — United States, 2017–2018. *MMWR Morb Mortal Wkly Rep* 2020;69:290–7.

<https://doi.org/10.15585/mmwr.mm6911a4>.

- Centers for Disease Control and Prevention. *Overdose Deaths Accelerating During COVID-19*; 2020.

<https://www.cdc.gov/media/releases/2020/p1218-overdose-deaths-covid-19.html>.

Accessed February 19, 2021.

- Alpren C, Dawson EL, John B, et al. Opioid Use Fueling HIV transmission in an urban setting: an outbreak of HIV infection among people who inject drugs—Massachusetts, 2015–2018. *Am J Public Health* 2020;110:37–44.

<https://doi.org/10.2105/AJPH.2019.305366>.

- Peters PJ, Pontones P, Hoover KW, et al. HIV infection linked to injection use of oxycodone in Indiana, 2014–2015. *N Engl J Med* 2016;375:229–39.

<https://doi.org/10.1056/NEJMoa1515195>.

- Gonsalves GS, Crawford FW. Dynamics of the HIV outbreak and response in Scott County, IN, USA, 2011–15: a modelling study. *Lancet HIV* 2018;5:e569–77.

[https://doi.org/10.1016/S2352-3018\(18\)30176-0](https://doi.org/10.1016/S2352-3018(18)30176-0).

- Golden MR, Lechtenberg R, Glick SN, et al. Outbreak of human immunodeficiency virus infection among heterosexual persons who are living homeless and inject drugs — Seattle, Washington, 2018. *MMWR Morb Mortal Wkly Rep* 2019;68:344–9.

<https://doi.org/10.15585/mmwr.mm6815a2>.

- Evans ME, Labuda SM, Hogan V, et al. Notes from the Field: HIV infection investigation in a rural area — West Virginia, 2017. *MMWR Morb Mortal Wkly Rep* 2018;67:257–8.

<https://doi.org/10.15585/mmwr.mm6708a6>.

- Suryaprasad AG, White JZ, Xu F, et al. Emerging epidemic of hepatitis C Virus infections among young nonurban persons who inject drugs in the United States, 2006–2012. *Clin Infect Dis* 2014;59:1411–9.

<https://doi.org/10.1093/cid/ciu643>.

- Zibbell JE, Iqbal K, Patel RC, et al. Increases in hepatitis C virus infection related to injection drug use among persons aged

≤30 years - Kentucky, Tennessee, Virginia, and West Virginia, 2006–2012. *MMWR Morb Mortal Wkly Rep* 2015;64:453–8,

<http://www.ncbi.nlm.nih.gov/pubmed/25950251>. Accessed April 10, 2020.

- Centers for Disease Control and Prevention (CDC). *Viral Hepatitis Surveillance United States, 2016*; 2017.

<https://www.cdc.gov/hepatitis/statistics/2016surveillance/pdfs/2016HepSurveillanceRpt.pdf>. Accessed April 10, 2020.

- Network CHA. Update: widespread outbreaks of hepatitis a among people who use drugs and people experiencing homelessness across the United States. 2019.

<https://emergency.cdc.gov/han/HAN00418.asp>. Accessed April 10, 2020.

- Strathdee SA, Kuo I, El-Bassel N, Hodder S, Smith LR, Springer SA. Preventing HIV outbreaks among people who inject drugs in the United States: plus ça change, plus ça même chose. *AIDS* 2020;34:1997–2005.

<https://doi.org/10.1097/QAD.0000000000002673>.

- Herfst S, Böhlinger M, Karo B, et al. Drivers of airborne human-to-human pathogen transmission. *Curr Opin Virol* 2017;22:22–9.

<https://doi.org/10.1016/j.coviro.2016.11.006>.

- Cox NJ, Trock SC, Burke SA. Pandemic Preparedness and the Influenza Risk Assessment Tool (IRAT). In: 2014:119–136. doi:10.1007/82_2014_419

- Liroy PJ, Laskin JD, Georgopoulos PG. Preparedness and response to chemical and biological threats: the role of exposure science. *Ann N Y Acad Sci* 2016;1378:108–17.

<https://doi.org/10.1111/nyas.13173>.

- Choudri BS, Al-Nasiri N, Charabi Y, Al-Awadhi T. Ecological and human health risk assessment. *Water Environ Res* 2020;92:1440–6.

<https://doi.org/10.1002/wer.1382>.

- Lane K, Charles-Guzman K, Wheeler K, Abid Z, Graber N, Matte T. Health Effects of Coastal Storms and Flooding in Urban Areas: A Review and Vulnerability Assessment. *J Environ Public Health* 2013;2013:1–13.

<https://doi.org/10.1155/2013/913064>.

- 24.

Klinkenberg D, Fraser C, Heesterbeek H. The effectiveness of contact tracing in emerging epidemics. *PLoS One* 2006;1:e12.

- Tupper KW, McCrae K, Garber I, Lysyshyn M, Wood E. Initial results of a drug checking pilot program to detect fentanyl adulteration in a Canadian setting. *Drug Alcohol Depend* 2018;190:242–5.

<https://doi.org/10.1016/j.drugalc-dep.2018.06.020>.

- Krieger MS, Goedel WC, Buxton JA, et al. Use of rapid fentanyl test strips among young adults who use drugs. *Int J Drug Policy* 2018;61:52–8.

<https://doi.org/10.1016/j.drugpo.2018.09.009>.

- Rosenblum D, Unick J, Ciccarone D. The rapidly changing US illicit drug market and the potential for an improved early warning system: evidence from Ohio drug crime labs. *Drug Alcohol Depend* 2020;208:107779.

<https://doi.org/10.1016/j.drugalc-dep.2019.107779>.

28. Fitzmaurice AG, Linley L, Zhang C, Watson M, France AM, Oster AM. Novel method for rapid detection of spatiotemporal HIV clusters potentially warranting intervention. *Emerg Infect Dis* 2019;25:988–91.
<https://doi.org/10.3201/eid2505.180776>.
29. Friedman LS. Real-time surveillance of illicit drug overdoses using poison center data. *Clin Toxicol* 2009;47:573–9.
<https://doi.org/10.1080/15563650902967404>
 Accessed March 20, 2020.
30. Slavova S, Costich JF, Bunn TL, et al. Heroin and fentanyl overdoses in Kentucky: epidemiology and surveillance. *Int J Drug Policy* 2017;46:120–9.
<https://doi.org/10.1016/j.drugpo.2017.05.051>.
31. Maxwell JC, Coleman JJ, Feng S-Y, Goto CS, Tirado CF. Cheese: an old drug in a new wrapper. *Drug Alcohol Depend* 2012;126:161–7.
<https://doi.org/10.1016/j.drugalc-dep.2012.05.015>.
32. Des Jarlais DC, Cooper HLF, Arasteh K, Feelemyer J, McKnight C, Ross Z. Potential geographic “hotspots” for drug-injection related transmission of HIV and HCV and for initiation into injecting drug use in New York City, 2011–2015, with implications for the current opioid epidemic in the US. Khudyakov YE, ed. *PLoS One*, 13; 2018. 2018:e0194799.
<https://doi.org/10.1371/journal.pone.0194799>.
33. Li Y, Hyder A, Southerland LT, Hammond G, Porr A, Miller HJ. 311 service requests as indicators of neighborhood distress and opioid use disorder. *Sci Rep* 2020;10:19579.
<https://doi.org/10.1038/s41598-020-76685-z>.
34. Anwar M, Khoury D, Aldridge AP, Parker SJ, Conway KP. Using twitter to surveil the opioid epidemic in North Carolina: an exploratory study. *JMIR Public Heal Surveill* 2020;6:e17574.
<https://doi.org/10.2196/17574>.
35. Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from tweets: estimating misuse of prescription opioids in the USA from social media. *J Med Toxicol* 2017;13:278–86.
<https://doi.org/10.1007/s13181-017-0625-5>.
36. Sarker A, Gonzalez-Hernandez G, Ruan Y, Perrone J. Machine learning and natural language processing for geolocation-centric monitoring and characterization of opioid-related social media chatter. *JAMA Netw Open* 2019;2:e1914672.
<https://doi.org/10.1001/jamanetworkopen.2019.14672>.
37. Hswen Y, Zhang A, Freifeld C, Brownstein JS. Evaluation of volume of news reporting and opioid-related deaths in the United States: comparative analysis study of geographic and socioeconomic differences. *J Med Internet Res* 2020;22:e17693.
<https://doi.org/10.2196/17693>.
38. Centers for Disease Control and Prevention (CDC). Use of enhanced surveillance for hepatitis C virus infection to detect a cluster among young injection-drug users—new York, November 2004–April 2007. *MMWR Morb Mortal Wkly Rep*. 2008;57:517–521. doi:18480744
39. {Centers for Disease Control and Prevention}. *Tuberculosis Genotyping*; 2008.
<https://www.cdc.gov/tb/publications/factsheets/statistics/genotyping.pdf>.
40. Oster AM, France AM, Panneer N, et al. Identifying clusters of recent and rapid HIV transmission through analysis of molecular surveillance data. *JAIDS J Acquir Immune Defic Syndr* 2018;79:543–50.
<https://doi.org/10.1097/QAI.0000000000001856>.
41. Kosakovsky Pond SL, Weaver S, Leigh Brown AJ, Wertheim JO. HIV-TRACE (TRANsmiSSion Cluster Engine): a Tool for Large Scale Molecular Epidemiology of HIV-1 and other rapidly evolving pathogens. Shapiro B, ed. *Mol Biol Evol* 2018;35:1812–9.
<https://doi.org/10.1093/molbev/msy016>.
42. Campbell EM, Jia H, Shankar A, et al. Detailed Transmission network analysis of a large opiate-driven outbreak of HIV infection in the United States. *J Infect Dis* 2017;216:1053–62.
<https://doi.org/10.1093/infdis/jix307>.
43. Tookes H, Bartholomew TS, Geary S, et al. Rapid identification and investigation of an HIV risk network among people who inject drugs —Miami, FL. *AIDS Behav* 2018;24:246–56.
<https://doi.org/10.1007/s10461-019-02680-9>
 2020.
44. Falade-Nwulia O, Hackman J, Mehta SH, et al. Factors associated with phylogenetic clustering of hepatitis C among people who inject drugs in Baltimore. *BMC Infect Dis* 2020;20:815.
<https://doi.org/10.1186/s12879-020-05546-x>.
45. Zhou S, Sizemore S, Moeser M, et al. Near real-time identification of recent human immunodeficiency virus transmissions, transmitted drug resistance mutations, and transmission networks by multiplexed primer ID—next-generation sequencing in North Carolina. *J Infect Dis* 2020.
<https://doi.org/10.1093/infdis/jiaa417>.
46. Centers for Disease Control and Prevention. Notes from the field: tuberculosis cluster associated with homelessness—Duval County, Florida, 2004–2012. *MMWR Morb Mortal Wkly Rep*. 2012;61(28).
47. Althomsons SP, Kammerer JS, Shang N, Navin TR. Using Routinely Reported Tuberculosis Genotyping and Surveillance Data to Predict Tuberculosis Outbreaks. Pai M, ed. *PLoS One* 2012;7:e48754.
<https://doi.org/10.1371/journal.pone.0048754>.
48. Marotta PL, Hunt T, Gilbert L, Wu E, Goddard-Eckrich D, El-Bassel N. Assessing spatial relationships between prescription drugs, race, and overdose in New York State from 2013 to 2015. *J Psychoactive Drugs* 2019;51:360–70.
<https://doi.org/10.1080/02791072.2019.1599472>.
49. Stopka TJ, Amaravadi H, Kaplan AR, et al. Opioid overdose deaths and potentially inappropriate opioid prescribing practices (PIP): a spatial epidemiological study. *Int J Drug Policy* 2019;68:37–45.
<https://doi.org/10.1016/j.drugpo.2019.03.024>.
50. Albright DL, McDaniel J, Kertesz S, et al. Small area estimation and hotspot identification of opioid use disorder among military veterans living in the Southern United States. *Subst Abuse* 2021;42:116–22.
<https://doi.org/10.1080/08897077.2019.1703066>.
51. Hernandez A, Branscum AJ, Li J, MacKinnon NJ, Hincapie AL, Cuadros DF. Epidemiological and geospatial profile of the prescription opioid crisis in Ohio, United States. *Sci Rep*. 2020;10:4341. doi:10.1038/s41598-020-61281-y
52. Brownstein JS, Green TC, Cassidy TA, Butler SF. Geographic information systems and pharmacoepidemiology: using spatial