



Published in final edited form as:

Lancet Public Health. 2021 October ; 6(10): e720–e728. doi:10.1016/S2468-2667(21)00080-3.

Identifying Counties at Risk of High Overdose Mortality Burden Throughout the Emerging Fentanyl Epidemic in the United States: A Predictive Statistical Modeling Study

Charles Marks, MPH^{1,*}, Daniela Abramovitz, MS², Christl A. Donnelly, ScD^{3,4} [Professor], Gabriel Carrasco-Escobar, MS^{2,5}, Rocío Carrasco-Hernandez, PhD², Daniel Ciccarone, MD⁶ [Professor], Arturo González-Izquierdo, PhD⁷, Natasha K. Martin, DPhil^{2,8}, Steffanie A. Strathdee, PhD² [Professor], Davey M. Smith, MD² [Professor], Annick Bórquez, PhD^{2,**}

¹Interdisciplinary Research on Substance Use Joint Doctoral Program, San Diego State University and University of California, San Diego

²Department of Medicine, University of California, San Diego

³Department of Statistics, University of Oxford

⁴MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London

⁵Health Innovation Lab, Institute of Tropical Medicine “Alexander von Humboldt”, Universidad Peruana Cayetano Heredia, Lima, Peru

⁶Department of Family and Community Medicine, University of California, San Francisco

⁷Institute of Health Informatics, University College London

⁸Population Health Sciences, University of Bristol

Abstract

*Corresponding Author: cmarks@sdsu.edu.

**Senior Author

Contributors

CM contributed to conceptualization, data curation, data verification, formal analysis, methodology, software, validation, visualization, and writing -- original draft. DA contributed to conceptualization, formal analysis, methodology, and writing -- review & editing. CAD contributed to conceptualization, methodology, and writing -- review & editing. DC contributed to resources and writing -- review & editing. GCE contributed to methodology, data curation, data verification, formal analysis, methodology, and writing -- review & editing. RCH contributed to methodology, data curation, data validation, formal analysis, methodology, and writing -- review & editing. AGI contributed to methodology and writing -- review & editing. NK contributed to methodology and writing -- review & editing. SAS contributed to writing -- review & editing. DMS contributed to writing -- review & editing. AB contributed to conceptualization, data curation, data validation, funding acquisition, methodology, resources, supervision, writing -- original draft & review & editing. As noted, CM, GCE, RCH, and AB had access to and verified data used in this study.

Data Sharing Statement

This study uses restricted secondary data and, therefore, we cannot provide the dataset required to recreate our study. However, since the predictor data was all publicly available, we have made a dataset available with variables generated from the restricted mortality records censored. This data is available at DOI: [10.17632/t9wbtt3mt2.1](https://doi.org/10.17632/t9wbtt3mt2.1). The R code used to analyze data is included within the Supplement (Page 12) to this manuscript and is available along with the censored dataset at the DOI listed.

Declaration of Interest

CM has no conflict of interest to report. CAD has no conflict of interest to report. AB has no conflict of interest to report.

DC reports personal fees from Celero Systems, Motley Rice LLP, Mallinckrodt Pharmaceuticals, and Nektar Therapeutics, outside the submitted work.

Background.—The emergence of fentanyl around 2013 represented a new, deadly stage in the US opioid epidemic. We developed a statistical regression approach to identify counties at the highest risk of high overdose mortality in the next year by predicting annual county-level overdose death rates across the contiguous US and validated it against observed overdose mortality data from 2013 to 2018.

Methods.—We fit mixed effects negative binomial regression models to predict next year’s county-level overdose death rates for the years 2013 to 2018. We used publicly available county-level data related to healthcare access, drug markets, socio-demographics, and the geographic spread of opioid overdose as model predictors. The crude number of county-level overdose deaths was extracted from restricted Centers for Disease Control and Prevention mortality records. To predict county-level overdose rates for the year 201X: 1) a model was trained on county-level predictor data for the years 2010–201(X-2) paired with county-level overdose deaths for the year 2011–201(X-1); 2) county-level predictor data for the year 201(X-1) was then fed into the model to predict the 201(X) county-level crude number of overdose deaths; and 3) the latter was converted to a population-adjusted rate. For comparison, we generated a benchmark set of predictions by applying the observed slope of change in overdose death rates in the previous year to 201(X-1) rates. To assess the predictive performance of the model, we compared predicted values (of both the model and benchmark) to observed values by 1) calculating the mean average error, root mean squared error, and Spearman’s correlation coefficient and 2) assessing the proportion of counties in the top decile (10%) of overdose death rates that were correctly predicted as such. Finally, in a post-hoc analysis, we sought to identify variables with greatest predictive utility.

Findings.—Across the entire US and through time, our modeling approach outperformed the benchmark strategy across all metrics. The average county-level overdose death rate rose from 11.8/100,000 to 15.4 in 2017 before falling to 14.8 in 2018. Our modeling approach similarly identified an increasing trend, predicting an average 11.8 deaths/100,000 in 2013 up to 15.1 in 2017 and still increasing to 16.4 in 2018. The benchmark model over-predicted average death rates each year, ranging from 13.0/100,000 in 2013 to 18.3 in 2018. Our modeling approach successfully ranked counties by overdose death rate identifying between 41.6% and 56.8% of counties in the top decile of overdose mortality (compared to 28.7% and 42.6% using the benchmark) each year and identified 194 of the 808 with emergent overdose outbreaks (i.e., newly entered the top decile) across the study period, versus 31 using the benchmark. In the post-hoc analysis, we identified geospatial proximity of overdose in nearby counties, opioid prescription rate, presence of an urgent care facility, and several economic indicators as the variables with the greatest predictive utility.

Interpretation.—Our study demonstrates that a regression approach can effectively predict county-level overdose death rates and serve as a risk assessment tool to identify future high mortality counties throughout an emerging drug use epidemic.

Funding.—National Institute on Drug Abuse

Introduction

The opioid epidemic in the United States (U.S.) caused over 400,000 documented opioid overdose deaths from 1999 to 2018, with 46,000 deaths occurring in 2018 alone.^{1,2} In

particular, fentanyl, a synthetic opioid approximately fifty times more potent than heroin, emerged in the eastern U.S. illicit drug market in 2013 as an adulterant of, or substitute for, heroin.^{3–5} In 2012, synthetic opioid overdose resulted in fewer than 1 death per 100,000 individuals.⁶ By 2018, synthetic opioids were responsible for nearly 10 deaths per 100,000 – over 31,000 deaths, accounting for 65% of all opioid overdose deaths for that year.⁶

Rather than a uniform increase in opioid-related mortality, the opioid overdose crisis is the latest in a decades long series of escalating geographically concentrated, time- and drug-specific overdose outbreaks, dating back to at least 1979.^{7,8} The current crisis has been described as a ‘triple wave’ of overdoses due to opioid pills, followed by escalating heroin-related overdose and most recently by a crescendo in synthetic opioid deaths.⁵ In the second and third waves, while the national opioid overdose death rate rose steadily over the past decade, mortality has been concentrated within specific regions – primarily the Midwest, Appalachia, and New England.^{5,6} However, increasingly greater synthetic opioid overdose deaths rates are being reported in the West, now likely exacerbated due to socio-economic, healthcare and drug markets disruptions associated with the Covid-19 pandemic.⁹ This illustrates the need for the rapid development of tools to predict potential overdose outbreaks, particularly in localities that have yet to experience fentanyl-related overdose outbreaks.

Overall, the primary aim of this study was to validate the application of a statistical modeling approach for identifying counties at highest risk of a drug overdose outbreak in the next year, throughout the fentanyl epidemic, by predicting county-level overdose deaths rates. Unfortunately, inconsistent and poor reporting of drug-specific overdose mortality¹⁰ across counties inhibits us from modeling fentanyl-specific overdose outbreaks. We developed a series of regression models to predict next year’s county-level overdose death rates in the US from 2013 to 2018. We validated our predictions against existing data on overdose death rates for each year to show how such a predictive tool could have been used throughout the course of the epidemic.

Methods

Data Preparation

Annual, county-level data for both outcomes and predictors were aggregated for all contiguous U.S. state counties (i.e. excluding Alaska and Hawaii) for the years 2010 through 2018.

The primary outcome for this study was county-level crude overdose death rate for the next year (i.e., predictors from year n are paired with overdose death rate from year $n+1$). This outcome was extracted from the Centers for Disease Control and Prevention (CDC) Wonder restricted database (using UCD codes X40–44, X60–64, X85, Y10–14). Due to statistical disclosure control, the CDC does not publicly report the number of overdose deaths for a given county in a given year if the absolute total was less than 10 in order to protect individual privacy. Following the request protocol from the CDC, we were given access to the full dataset with all counties overdose death rates reported. To be consistent with CDC human subjects’ protections, we will not report or reflect on an individual county’s

overdose death rate in this report. As this study relied on the use of secondary de-identified county-level data, the Institutional Review Board (IRB) of the University of California San Diego determined that IRB review was not required (according to the Code of Federal Regulations, Title 45, part 46).

Predictors included as fixed effects in our modeling approach (see Table 1) were obtained from publicly available databases reporting county-level estimates throughout the study period. They were chosen to be consistent with prior literature modeling risk of overdose in the US, including indicators of healthcare access, drug markets, socio-economic indicators, and the geographic spread of the epidemic.^{11–14} To estimate the county-level availability of opioid use disorder treatment, we included the total number of buprenorphine physician waivers approved by the Substance Abuse and Mental Health Services Administration active each year. As well, to operationalize access to emergency health care, we included a binary variable measuring the presence of an urgent care facility within the given county. We included the county-level opioid prescription rate per 100 people and the state-level count of substances identified as including fentanyl in local-, state-, and federal-level forensic labs. Additionally, we included the log of the jail population. We included socio-economic indicators, such as unemployment rate, extracted from the Census American Community Survey. Consistent with health-related machine learning recommendations,¹⁵ we do not hypothesize there exists a relationship between race and overdose that is not mediated or confounded by latent structural racism (such as disparate opioid prescription patterns by race),¹⁶ thus we do not include race as a predictor (see Supplement Page 2 for further discussion). Finally, to account for the geographic spread of overdose death, we included a categorical measure of county urbanicity and a continuous gravity variable accounting for the overdose death rates of nearby counties. We provide full detail on the variables' selection in the Supplement Page 3, including our assessment of predictor collinearity.

Statistical Modeling Approach

The modeling approach was applied to predict overdose death rates for each year from 2013 to 2018. When predicting a given year (e.g., 201X), the model is trained on paired predictor-death rate data from years 2010 through two years prior to the prediction year (201X-2). Predictor data for a given year is paired with the crude number of overdoses that occurred in the subsequent year as the model outcome. Then, predictors from the year prior to the prediction year (201X-1) are fed into the model (which specifies coefficients relating each predictor to the outcome) to predict 201X county-level crude number of overdose deaths, which is then converted into a population rate (per 100,000). For example, as shown in Figure 1, in order to predict 2013 overdose death rates: 1) a model is trained using longitudinal predictor data from 2010 to 2011 (paired with outcomes for 2011 and 2012, respectively); 2) predictors from 2012 are then fed into the model to predict 2013 overdose death counts; 3) the predicted death counts are converted into overdose death rates (i.e. deaths per 100,000); and, finally, 4) the predicted overdose death rates for 2013 are compared to the actual overdose death rates to evaluate predictive accuracy.

For predicting each year's overdose death rates, we applied mixed effects negative binomial regression (as detailed in Figure 1 – see Supplement Page 1 for a detailed discussion

justifying our chosen modeling approach). A random intercept for each county was incorporated with a random slope for year. This model specification accounts for two hypothesized relationships within the data: 1) overdose death observations from the same county are correlated (justifying the random intercept for each county); and 2) the rate of change in overdose deaths will be dependent on the epidemic stage in a given county (justifying the application of random slopes for year). In addition, we included an offset term for the log of the population “carrying capacity”, similar to Sumetsky et al.¹⁷ We hypothesized that as more overdose deaths occur in a location, the population of susceptible individuals would diminish. Thus, we defined carrying capacity as 5% of the 2010 county population minus the number of overdoses in the county the prior three years (or the prior available years in the data for years 2011 and 2012), setting 50 as the minimum possible carrying capacity (see Supplement Page 1 for further discussion of carrying capacity). The outcome of the model was the number of overdose deaths the subsequent year in each county. We included each variable in Table 1 as a fixed effect. Given that our goal was to simulate real-time prediction and that we cannot know the accuracy of model performance *a priori*, it would be unrealistic to choose a set of optimally performing fixed effects. In *Post-Hoc Analysis*, we describe additional steps taken to determine which fixed effects best informed model prediction.

All analyses were conducted in R using the *lme4* package.^{18,19} Further details and code for running the analyses are available in the Supplement Pages 1 & 12.

Prediction Evaluation Approach.

We consider five primary metrics for assessing model performance. The first three, mean average error (MAE), root mean square error (RMSE), and Spearman’s ρ , measure the accuracy of outcome predictions. The MAE is the average magnitude of the difference between the predicted and observed overdose death rate for each county. The RMSE is the square root of the average magnitude of the difference squared – similar to MAE but penalizes prediction errors with greater magnitude. More accurate predictions will result in smaller MAE and RMSE. Spearman’s ρ compares the predicted ranking of counties by overdose death rate compared with the actual observed rankings – results closer to 1 indicate that the model was more effective at rank-ordering counties based on overdose death rate. The final two metrics seek to assess how well the model identified counties at highest risk of an overdose outbreak in the subsequent year (defined by an overdose death rate in the top decile relative to other counties). To do so, we first disaggregated the predicted and observed overdose death rates into deciles (10th, 20th, [...], 100th centile) and categorized all counties into their corresponding decile for both predicted and observed overdose rates. The first metric is the proportion of counties observed in the top decile (i.e. top 10% of observed overdose deaths rates) that were rightly predicted to be in the top decile. Then, to characterize model performance identifying counties with emergent overdose outbreaks, we defined such an emergent outbreak as a county being outside of the top decile in the year 201(X-1) and then entering the top decile in year 201(X). The second metric is the proportion of all observed emergent outbreak counties which the modeling approach accurately predicted as newly being in the top decile in 201X.

To contextualize the results, we generated benchmark predictions for comparison. This benchmark strategy assumed the change in overdose death rate between years 201X-2 and 201X-1 would remain the same between the years 201X-1 and 201X. We calculated the slope for the change in overdose death rate from year 201X-2 to 201X-1 and added it to the 201X-1 overdose death rate to predict the 201X rate. If the value predicted for a county for a given year was below 0, we rounded it up to 0. This heuristic approach provides a simple, yet intuitive, way to predict future overdose death rates – the utility of our modeling approach can be understood in comparison to the performance of this benchmark approach.

Data Exploration Application

To address the challenges in presenting county-level data for the full US, we provide a web application that can be used to explore the data in various ways at <http://overdosepredictiondashboard.emergens-project.com/>. We provide this dashboard as an aid to this manuscript and to display how such findings may be readily disseminated to appropriate stakeholders. In accordance with CDC data protections, we have censored data which are not available in the unrestricted CDC mortality records.

Post-Hoc Analyses

It is of interest to understand the contribution of fixed effects to the predictive accuracy of the model. When making predictions, it is also uncertain what the best set of fixed effects will be, given that the model cannot be evaluated until after the predicted events occur. We employ a bootstrapped forward variable selection strategy similar to that described by Beyene et al to identify the fixed effects with the greatest predictive utility (see Supplement Page 5 for full description).²⁰ We focus only on predicting overdose death rate for the year 2018 and the metric we are seeking to optimize is the proportion of counties correctly predicted in the top decile.

In total, we ran 100 bootstrap iterations. We display, as the result, the proportion of the time each variable was included in the final model. Fixed effects that are chosen more frequently are considered to have greater predictive value than fixed effects chosen less frequently.

We also implemented model diagnostics and a sensitivity analysis applying the model in the eastern and western regions of the U.S. to confirm its results are robust to changes in the model training process (see Supplement Pages 6 – 9). To execute the sensitivity analysis, the analytic approach described was run separately for counties east and west of the Mississippi River, respectively. Results were then evaluated to determine if the model still performed adequately when trained on smaller, distinct regions of the country.

Role of Funding Source

The funding source had no role in data collection, analysis, interpretation, writing of the manuscript, nor the decision to submit.

Results

Among the 3,106 counties included in the study, from 2013 to 2018, observed mean county-level overdose death rates increased from 11.8 deaths per 100,000 in 2013 to 15.4 deaths per 100,000 in 2017, before falling to 14.6 deaths per 100,000 in 2018 (Table 2). The benchmark prediction strategy over-predicted the mean county-level overdose death rate each year, increasing from 13.0 deaths per 100,000 in 2013 to 18.3 deaths per 100,000 in 2018. The negative binomial approach predicted a mean 11.8 deaths per 100,000 in 2013 and followed by a steady increase from 11.5 deaths per 100,000 in 2014 to 16.4 deaths per 100,000 in 2018.

The negative binomial approach outperformed the benchmark prediction strategy each year, according to MAE, RMSE, and Spearman's ρ (see Table 3). The benchmark MAE increased from 10.70 in 2013 to 12.37 in 2018, whereas the MAE of the negative binomial approach ranged from 6.58 to 7.73. The RMSE of the benchmark approach ranged from 18.38 to 20.67, whereas the negative binomial approach RMSE ranged from 10.04 to 11.55. The benchmark Spearman's ρ increased from 0.35 in 2013 to 0.45 in 2018, whereas the negative binomial model Spearman ρ was generally 0.2 greater, increasing from 0.57 in 2013 to 0.65 in 2018.

We then divided counties into deciles based on observed and predicted overdose death rates (i.e. top decile were the 10% of counties with the highest overdose death rate, second decile the next 10%, and so on), to identify if the counties predicted to have the highest overdose death rates indeed experienced them. The benchmark prediction strategy correctly predicted between 89 and 132 of the 310 counties observed to be in the top decile for each year (see Table 4). The negative binomial approach generally improved over time, identifying 129 of the 310 counties in the top decile in 2013 and 171 in 2018. This improvement may indicate that model performance improves in this regard given more training data.

The number of counties that newly entered the top decile fell from 175 counties from 2013 to 2014 to 149 counties from 2017 to 2018. The benchmark strategy, at its best in 2018, correctly predicted only 7 of 149 counties newly entering the top decile, whereas the negative binomial approach correctly predicted at least 33 (and up to 46) of the counties newly entering the top decile. While these results indicate further room for improvement, they display that the negative binomial approach employed represents a meaningful predictive improvement over our benchmark heuristic of predicting based on annual overdose death rate trends.

Finally, we sought to characterize the predictive value of each fixed effect in the model via forward selection bootstrapping approach (Table 5). The overdose gravity variable was included in 81% of simulations, indicating that the geospatial dimension of overdose is highly predictive of subsequent year overdose death rate. The opioid prescription rate was included 66% of simulations and the presence of an urgent care facility in the county was included 53% of the time, indicating that such health care indicators are of predictive value – though we note that the number of buprenorphine provider waivers in the county was only chosen 11% of the time. As well, several economic indicators including changes in county

payroll, median household income, and changes in employee were all chosen around half of the time. Diagnostic analyses indicated the model tended to underpredict high overdose death rates, but this improved over time. Separately implementing the model in the eastern and western regions, resulted in very similar (although marginally better) performance (see Supplement Page 8).

Discussion

This study demonstrated how a statistical modeling approach can be employed to identify counties at risk of experiencing overdose death outbreaks. Our model predicted counties' overdose death rates from 2013 to 2018 with substantially greater accuracy than an intuitive benchmark heuristic. Most importantly, it displayed far greater capacity than the benchmark for predicting counties experiencing emergent drug overdose outbreaks by identifying counties newly entering the top mortality bracket. As such, this model should be considered when attempting to identify which counties are in need of resources to respond to potential overdose outbreaks, including counties yet to experience them. We note that further research aimed at improving model performance and timely access to data are needed to ensure efficacious application. Our post-hoc analysis indicates our fixed effects capturing the geo-spatial spread of overdose, opioid prescribing patterns, and several economic indicators provided the most predictive utility.

While similar models have been used to inform funding allocation, such as the CDC's drug-related HIV outbreak risk assessment model,^{11,21} these have not been validated against data and have not been designed to provide yearly predictions (with a recent study from Sumetsky et al. as an exception).¹⁷ Model validation is key to both ensuring that the tools used for policy guidance are providing accurate information that will lead to an effective allocation of resources, as well as to improving our understanding of the epidemic processes. Given the changing nature of drug use epidemics, tools that capture risk over time are needed.

The study has limitations. First, the model performance is not optimal. However, predicting overdose outbreaks at national level is challenging and such improvements over a heuristic benchmark can prevent much harm by directing attention towards counties that would have otherwise not been considered at risk. While it was not possible to do so in this study, comparing the performance of our model with that of other models introduced in the literature (such as that by Sumetsky et al, Campo et al, and Cooper et al)^{17,22,23} may advance the broader effort to develop better performing models. Further, given that this is a nascent line of research, we highlight the importance of evaluating the performance of a variety of modeling strategies. To our knowledge, this study is the first to apply a mixed effects negative binomial regression strategy to predict overdose deaths. Recent works have applied Bayesian spatial-temporal models, polynomial functions, and a variation of the random forest algorithm.^{17,22,23} Future research should aim to replicate and compare these methods in order to identify strengths of each approach, which can inform future model development.

Second, longitudinal predictive studies require the consistent and timely dissemination of data. Thus, the outcome and predictors need to be available for the same localities (i.e. counties), same time periods, and same time steps (i.e. years, months) in order to be utilized. Such requirements limit the pool of available variables to include as predictors. For example, while we incorporated estimates of opioid prescription rates per county and fentanyl seizures by state to capture changes in drug markets, these indicators provide only partial information as they do not tell us about either drug volume or potency. In addition, having county-level seizure data would likely improve model performance. Similarly, we included active buprenorphine providers per year by county as a measure of drug treatment coverage. However, there is high variation in the number of patients seen by each provider and regulations on the limit of patients per provider have been relaxed over time.²⁴

Third, we took a simple approach for identifying the “susceptible” population in each county. Most people in each county are not at risk of experiencing an overdose. Sumetsky et al. provide an example of a more computationally intensive calculation of county carrying capacity.¹⁷ Future research should seek to design and validate approaches aimed at quantifying this county-level “susceptible” population.

As well, the timeliness of data availability shapes the utility of the method. As of January 2021, the restricted overdose death data from the CDC was available through 2018. This means that future applications of this or other predictive modeling approaches require more rapid dissemination of data to ensure the timely access of evidence-based guidance among relevant stakeholders. Increasingly, individual states and counties’ public health departments are implementing web-portals, such as California’s, the Rhode Island and the Michigan Opioid Overdose Surveillance Dashboards,^{25–27} where preliminary data are made publicly available on a quarterly, biannual and near-real time basis, respectively. States with more rapid data dissemination may apply this method for their specific locality. Analytic approaches can be modified to make predictions several years into the future but given the rapidly changing nature of drug use epidemics, the timely availability of data promises to provide greater predictive benefit. This is particularly true in the context of the Covid-19 pandemic, which has affected and will continue to shape drug use related behaviors and harms.^{28,29}

Based on these findings, we provide directions for future research and endeavors which can improve the utility of this modeling approach. First, as highlighted in the limitations section, better and more timely data of both drug use patterns and drug markets are needed to enable rigorous analyses of drug use epidemics and prediction analyses. This could be achieved through more timely and granular accessibility to NFLIS data and through establishing free and accessible drug testing programs in collaboration with harm reduction organizations.⁵ Publicly available data on prescription drugs is also key to evaluating risk in a population. Local data on both the size and socio-demographic characteristics of people who use drugs could be systematically collected and linked through coordinated collaboration with primary and emergency medical services, law enforcement institutions and harm reduction organizations. A recent study from Campo et al. displayed that concurrent Google search trends may be an effective strategy for making real-time, dynamic predictions of county-level overdose death rates, given the immediate availability of this data.²³

Second, to use prediction to mitigate the harms of the opioid crisis, it is crucial that we be able to swiftly communicate predictions to appropriate stakeholders – this is especially important considering how rapidly US drug markets are understood to change.³⁰ The development of dashboards can facilitate the application of these peer-reviewed methods in a way that allows for the rapid dissemination of results. We provide a dashboard (<http://overdosepredictiondashboard.emergens-project.com/>) where the results of this study can be explored – this dashboard represents a model for how this method may be applied to inform relevant stakeholders in making decisions about overdose prevention measures. Through such platforms, stakeholders can access prediction results and use the findings to inform resource allocation and overdose response initiatives. While this study focuses on the accuracy and validity of the approach employed, we expect to extend it to produce future predictions, conditional on data availability.

Conclusion

Our statistical model effectively rank-orders counties based on the predicted overdose death rates for the subsequent year and is able to predict counties that will experience emergent increases in overdose mortality. This study provides the first rigorously validated tool to inform policy planning in the context of overdose epidemics driven by emerging drugs and sets a new standard for the development of a data driven response to drug use epidemics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding information: This study was supported by a NIDA Avenir grant DP2DA049295. CAD acknowledges funding from the MRC Centre for Global Infectious Disease Analysis (reference MR/R015600/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO), under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union.

References

1. Hedegaard H, Minino A, Warner M. Drug Overdose Deaths in the United States, 1999–2018.; 2020. <https://www.cdc.gov/nchs/data/databriefs/db356-h.pdf>.
2. 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(11):290–297. doi:10.15585/mmwr.mm6911a4 [PubMed: 32191688]
3. Ciccarone D. Fentanyl in the US heroin supply: A rapidly changing risk environment. *Int J Drug Policy.* 2017;46:107–111. doi:10.1016/j.drugpo.2017.06.010 [PubMed: 28735776]
4. Ciccarone D, Ondocsin J, Mars SG. Heroin uncertainties: Exploring users’ perceptions of fentanyl-adulterated and -substituted “heroin”. *Int J Drug Policy.* 2017;46:146–155. doi:10.1016/j.drugpo.2017.06.004 [PubMed: 28735775]
5. Ciccarone D. The triple wave epidemic: Supply and demand drivers of the US opioid overdose crisis. *Int J Drug Policy.* 2019;71:183–188. doi:10.1016/j.drugpo.2019.01.010 [PubMed: 30718120]
6. Centers for Disease Control and Prevention National Center for Health Statistics. Multiple Cause of Death 1999–2018 on CDC WONDER Online Database, released in 2020. Data are from the Multiple Cause of Death Files, 1999–2018, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative. 2020. <http://wonder.cdc.gov/mcd-icd10.html>.

7. 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-). 2018;361(6408). doi:10.1126/science.aau1184
8. Kiang M V, Basu S, Chen J, Alexander MJ. Assessment of Changes in the Geographical Distribution of Opioid-Related Mortality Across the United States by Opioid Type, 1999–2016. *JAMA Netw open*. 2019;2(2):e190040. doi:10.1001/jamanetworkopen.2019.0040
9. Shover CL, Falasinnu TO, Dwyer CL, et al. Steep increases in fentanyl-related mortality west of the Mississippi River: Recent evidence from county and state surveillance. *Drug Alcohol Depend*. 2020;216:108314. doi:10.1016/j.drugalcdep.2020.108314
10. Slavova S, O'Brien DB, Creppage K, et al. Drug Overdose Deaths: Let's Get Specific. *Public Health Rep*. 2015;130(4):339–342. doi:10.1177/003335491513000411 [PubMed: 26345488]
11. Van Handel MM, Rose CE, Hallisey EJ, et al. County-Level Vulnerability Assessment for Rapid Dissemination of HIV or HCV Infections Among Persons Who Inject Drugs, United States. *JAIDS J Acquir Immune Defic Syndr*. 2016;73(3):323–331. doi:10.1097/QAI.0000000000001098 [PubMed: 27763996]
12. Monnat SM, Peters DJ, Berg MT, Hochstetler A. Using Census Data to Understand County-Level Differences in Overall Drug Mortality and Opioid-Related Mortality by Opioid Type. *Am J Public Health*. 2019;109(8):1084–1091. doi:10.2105/AJPH.2019.305136 [PubMed: 31219718]
13. Rossen LM, Khan D, Warner M. Trends and Geographic Patterns in Drug-Poisoning Death Rates in the U.S., 1999–2009. *Am J Prev Med*. 2013;45(6):e19–e25. doi:10.1016/j.amepre.2013.07.012 [PubMed: 24237925]
14. Haffajee RL, Lin LA, Bohnert ASB, Goldstick JE. Characteristics of US Counties With High Opioid Overdose Mortality and Low Capacity to Deliver Medications for Opioid Use Disorder. *JAMA Netw Open*. 2019;2(6):e196373. doi:10.1001/jamanetworkopen.2019.6373
15. Robinson WR, Renson A, Naimi AI. Teaching yourself about structural racism will improve your machine learning. *Biostatistics*. 11 2019. doi:10.1093/biostatistics/kxz040
16. Om A. The opioid crisis in black and white: the role of race in our nation's recent drug epidemic. *J Public Health (Bangkok)*. 2018;40(4):e614–e615. doi:10.1093/pubmed/fdy103
17. Sumetsky N, Mair C, Wheeler-Martin K, et al. Predicting the Future Course of Opioid Overdose Mortality. *Epidemiology*. 2020;Publish Ah. doi:10.1097/EDE.0000000000001264
18. R Core Team. R: A language and environment for statistical computing. 2019. <https://www.r-project.org/>.
19. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw*. 2015;67(1). doi:10.18637/jss.v067.i01
20. Beyene J, Atenafu EG, Hamid JS, To T, Sung L. Determining relative importance of variables in developing and validating predictive models. *BMC Med Res Methodol*. 2009;9(1):64. doi:10.1186/1471-2288-9-64 [PubMed: 19751506]
21. Rickles M, Rebeiro PF, Sizemore L, et al. Tennessee's In-state Vulnerability Assessment for a "Rapid Dissemination of Human Immunodeficiency Virus or Hepatitis C Virus Infection" Event Utilizing Data About the Opioid Epidemic. *Clin Infect Dis*. 2018;66(11):1722–1732. doi:10.1093/cid/cix1079 [PubMed: 29228122]
22. Lyle Cooper R, Thompson J, Edgerton R, et al. Modeling dynamics of fatal opioid overdose by state and across time. *Prev Med Reports*. 2020;20:101184. doi:10.1016/j.pmedr.2020.101184
23. Campo DS, Gussler JW, Sue A, Skums P, Khudyakov Y. Accurate spatiotemporal mapping of drug overdose deaths by machine learning of drug-related web-searches. Blackard J, ed. *PLoS One*. 2020;15(12):e0243622. doi:10.1371/journal.pone.0243622
24. Duncan A, Anderman J, Deseran T, Reynolds I, Stein BD. Monthly Patient Volumes of Buprenorphine-Waivered Clinicians in the US. *JAMA Netw Open*. 2020;3(8):e2014045. doi:10.1001/jamanetworkopen.2020.14045
25. California Department of Public Health. California Opioid Overdose Surveillance Dashboard. <https://skylab.cdph.ca.gov/ODdash/>. Published 2020.
26. Rhode Island Department of Health. Prevent Overdose Rhode Island. <https://preventoverdoseri.org/overdose-deaths/>. Published 2020.

27. University of Michigan. Michigan System for Opioid Overdose Surveillance. <https://systemforoverdosesurveillance.com/>. Published 2019.
28. Volkow ND. Collision of the COVID-19 and Addiction Epidemics. *Ann Intern Med.* 2020;173(1):61–62. doi:10.7326/M20-1212 [PubMed: 32240293]
29. Becker WC, Fiellin DA. When Epidemics Collide: Coronavirus Disease 2019 (COVID-19) and the Opioid Crisis. *Ann Intern Med.* 2020;173(1):59–60. doi:10.7326/M20-1210 [PubMed: 32240291]
30. 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. doi:10.1016/j.drugalcdep.2019.107779

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Research in context

Evidence before this study

The rapid diversification in synthetic opioids of increased potency, the expansion of drug markets through the dark web and the observed increases in polydrug use associated with higher risk of health harms are all contributing to the emergence of increasingly rapid and harmful fatal overdose outbreaks in the United States (U.S.). To mitigate future harms, it is crucial to predict where and when opioid use-related outbreaks will occur and plan for a preemptive response. We reviewed the literature to identify quantitative studies aimed at predicting public health outbreaks associated with opioid use epidemics in the U.S. by searching for the following terms in PubMed (updated on January 20 2021): (“Substance-Related Disorders”[Mesh] OR drug use[tiab] OR opioid[tiab]) AND (outbreak[tiab] OR “Epidemics”[Mesh] OR overdose[tiab]) AND (“Statistics as Topic” [Mesh] OR “Regression Analysis”[Mesh] OR statistic*[tiab] OR predictive[tiab] OR model[tiab]) AND (“United States”). While the search retrieved over 1,000 studies, a minority were directly relevant to our research question as most employed an explanatory framework and few extended it for predictive purposes. An influential CDC study by Van Handel et al. aimed to assess the risk of injection drug use HIV and HCV associated outbreaks across U.S. counties. However, their methods have not been validated and did not include fatal overdose as an outcome. A recent study led by Sumetsky et al. addressed the urgent need for overdose outbreak prediction models and 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. While their findings are promising, they have not yet been evaluated across the entire country, which is important given the high geographical heterogeneity in overdose outcomes in the U.S. Another recent study by Cooper et al. used three-degree polynomial models to investigate fatal overdose dynamics from 2012 to 2016 by state, disaggregating rates by heroin, semi-synthetic and synthetic opioids. They identified states with highest elasticity (i.e. rate of change over time) for each of the opioid sub-epidemics. These findings are useful in terms of improving our understanding of different opioid sub-epidemics dynamics; however, there is no assessment of the model’s predictive performance and how outputs may be operationalized to inform policy. Finally, a 2020 study by Campo et al. applied a variation of the random forest algorithm to predict state and county-level overdose deaths rates, by using concurrent Google search trends as model predictors. Predictive performance was high, but they used publicly available overdose death data, hiding much of the heterogeneity across smaller counties. There remains a need to further develop the nascent field of overdose epidemic prediction through the design and validation of analytic methods that provide actionable information to guide the response at national and local levels in the context of emerging drug use epidemics.

Added value of this study

In this study, using publicly available predictor data, we implemented and validated a mixed effects negative binomial regression method for predicting county-level overdose death rates in the next year from the emergence of fentanyl in 2013 to 2018, across

the contiguous U.S. We compared our yearly overdose mortality predictions to observed data and to a simple predictive benchmark to further characterize our model’s predictive value. To produce meaningful results to guide policy, we identified counties in the top mortality decile as well as those newly entering that category (corresponding to counties with emerging outbreaks). We displayed that, if our method had been implemented in real-time, we would have had an improved capacity to identify counties at the highest risk of experiencing overdose outbreaks throughout the fentanyl wave of the opioid crisis.

Implications of all the available evidence

Taken together, to address the harms of the U.S. opioid crisis, it is crucial that available analytic approaches be employed to identify localities at the highest risk of experiencing an overdose outbreak in the near future. Our study contributes to ongoing efforts to strengthen our epidemiologic toolset to inform the opioid response, and the development of further quantitative methods, including geospatial, machine learning and dynamic modeling approaches should be encouraged. Importantly, timely and geographically representative data on drug use and associated outcomes, as well as drug markets, are crucial to increasing predictive power of these tools. A stronger drug market surveillance infrastructure is needed. Further, it is important that strategies to disseminate findings to relevant stakeholders be implemented. Here, we display our findings through an interactive a dashboard to illustrate how this can aid in the transparent dissemination of predictions. By improving our ability to make such predictions and relay this information to appropriate stakeholders, we will improve our ability to swiftly and precisely allocate resources and instantiate responses to effectively mitigate potential overdose harms.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

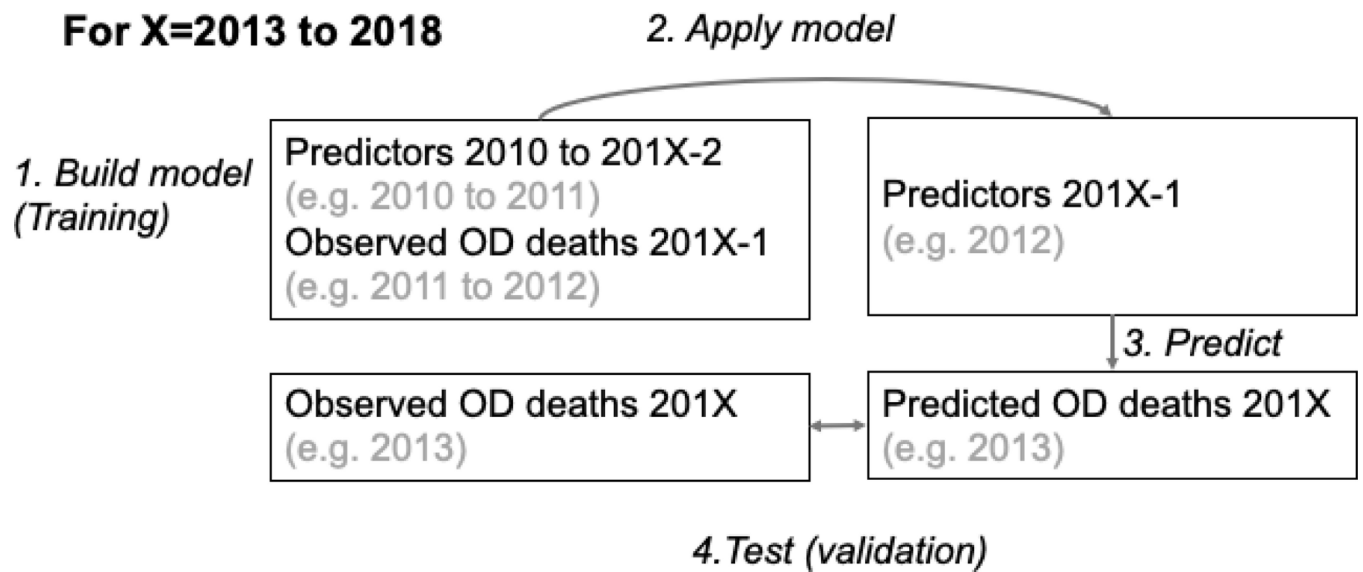


Figure 1. Workflow of modeling approach.

Predictors from the years 2010 to 201X-2, paired with respective overdose outcomes from 2011 to 201X-1 were used to train each model. Then predictors from the year 201X-1 were fed into the model to predict the overdose death rate in 201X. Finally, these predicted rates were compared to the observed rates for 201X. OD = overdose.

Table 1.

Predictors included as fixed effects in the final modeling approach.

| Predictor | Description | Source* |
|--|--|---------------------|
| Healthcare access | | |
| Buprenorphine waived physicians | Crude number of active buprenorphine waived physicians each year | SAMHSA |
| Urgent Care Presence | Presence of an Urgent Care facility within county | HSIP Gold |
| Drug markets | | |
| Opioid Prescription Rate | Opioid prescribing rate per 100 people each year | CDC (IQVIA Xponent) |
| Log Fentanyl Seizure Data | State-level count of fentanyl tested in local-, state-, and federal-level forensic labs each year | NFLIS |
| Log Jail Population Size | The log of the jail population size | VERA |
| Socio-economic indicators | | |
| High School Graduation Rate | Proportion of people living in the county estimated to have graduated from high school or received an equivalent certification. | ACS |
| Poverty Rate | Proportion of households in the county estimated to be living at or below the poverty line. | ACS |
| Unemployment Rate | Proportion of people able to work in the county estimated to be unemployed. | ACS |
| Employee capacity Difference | Difference in the employment capacity (measured as number of staff employed) of all companies across industries between current and past year in the county. | CBP |
| Payroll Difference | Difference in payroll (measured in US dollars) of all companies across industries between current and past year in the county. | CBP |
| Log Median Household Income | The logarithm of the estimated median household income in the county. | ACS |
| Proportion of Homeowner Households That Spend At Least 35% of Income on Mortgage | The proportion of homeowner households in the county where it estimated that the household spends at least 35% of their income on their mortgage. | ACS |
| Proportion of Renter Household That Spend At Least 35% of Income on Rent | The proportion of renter households in the county where it estimated that the household spends at least 35% of their income on their rent. | ACS |
| Geographic Spread of Epidemic | | |
| Log Overdose Gravity | Continuous variable generated to operationalize overdose death rates in neighboring counties. To derive the gravity variable for a given county x in year t , we first identified the set of all counties Y within 200 miles of county x . Distances were measured from central, internal points in each county and were extracted from a dataset created by the National Bureau of Economic Research. Second, for each county y in Y , we divided the overdose death rate for county y in the year t by the distance between counties x and y , squared. Third, we summed the values calculated in the previous step for each county y in Y . Finally, we took the natural logarithm of this summed value to get the final value. | NBER |
| Urbanicity | Six category variable based on US Office of Management and Budget 2013 determination of metropolitan statistical areas, coded on a spectrum from most urban (1) to most rural (6). | NCHS |

* Detailed source information for each variable is provided in the Supplement Page 3.

ACS: Census American Community Survey; NFLIS: National Forensic Laboratory Information System; CBP: County Business Patterns; NCHS: National Center for Health Statistics; SAMHSA: Substance Use and Mental Health Services Administration; HSIP: Homeland Security Infrastructure Program; VERA: VERA Institute of Justice; NBER: National Bureau of Economic Research

Table 2.

Mean observed, benchmark prediction, and model prediction of county-level overdose death rates per 100,000 each year from 2013 to 2018.

| | Observed Mean Overdose Death Rate | Benchmark Prediction Mean Overdose Death Rate | Negative Binomial Prediction Mean Overdose Death Rate |
|-------------|--|--|--|
| 2013 | 11.8 | 13.0 | 11.8 |
| 2014 | 12.6 | 14.1 | 11.5 |
| 2015 | 13.1 | 14.7 | 12.3 |
| 2016 | 14.6 | 15.8 | 13.3 |
| 2017 | 15.4 | 18.0 | 15.1 |
| 2018 | 14.6 | 18.3 | 16.4 |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Mean average error, root mean squared error and Spearman's ρ for the both the benchmark and negative binomial predictions for each year from 2013 to 2018.

| | Benchmark | | | Negative Binomial – With Fixed Effects | | |
|-------------|-----------|-------|-------------------|--|-------|-------------------|
| | MAE | RMSE | Spearman's ρ | MAE | RMSE | Spearman's ρ |
| 2013 | 10.70 | 18.38 | 0.35 | 6.58 | 10.04 | 0.57 |
| 2014 | 10.92 | 18.09 | 0.36 | 6.70 | 10.42 | 0.58 |
| 2015 | 11.18 | 19.32 | 0.40 | 6.74 | 10.34 | 0.62 |
| 2016 | 11.72 | 20.20 | 0.41 | 7.66 | 11.55 | 0.64 |
| 2017 | 12.34 | 20.67 | 0.45 | 7.52 | 11.22 | 0.67 |
| 2018 | 12.37 | 20.67 | 0.45 | 7.73 | 10.95 | 0.65 |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4.
Number of total and new counties in the top decile of overdose death rates correctly predicted by the benchmark and model each year.

In the top portion of the table (**Top Decile**), the number of counties in the top decile ($n = 310$) that were accurately predicted (i.e., true positives) as such are presented for each year and each approach. The number of false positives (i.e., counties incorrectly predicted to be in the top decile) can be calculated by subtracting the number of true positives from 310. In the bottom half, the number of counties that newly entered the top decile (i.e., were not in the top decile the year before) that were accurately predicted as such are presented for each year and each approach.

| | Benchmark | Negative Binomial |
|-------------|----------------------------|----------------------------|
| | Top Decile | Top Decile |
| 2013 | 102/310 (32.9%) | 129/310 (41.6%) |
| 2014 | 89/310 (28.7%) | 145/310 (46.8%) |
| 2015 | 104/310 (33.5%) | 158/310 (51.0%) |
| 2016 | 111/310 (35.8%) | 154/310 (49.7%) |
| 2017 | 132/310 (42.6%) | 176/310 (56.8%) |
| 2018 | 122/310 (39.4%) | 171/310 (55.2%) |
| | Newly in Top Decile | Newly in Top Decile |
| 2014 | 8/175 (4.6%) | 46/175 (26.3%) |
| 2015 | 6/170 (3.5%) | 40/170 (23.5%) |
| 2016 | 6/165 (3.6%) | 37/165 (22.4%) |
| 2017 | 4/149 (2.7%) | 38/149 (25.5%) |
| 2018 | 7/149 (4.7%) | 33/149 (22.1%) |

Table 5.

Percentage of bootstrap simulations in which each fixed effect was selected, ordered from most frequently included to least

| Variable | % of Times Chosen |
|--|-------------------|
| Log Overdose Gravity | 81% |
| Opioid Prescriptions Per 100 | 66% |
| Payroll Difference | 54% |
| Urgent Care Presence | 53% |
| Median Household Income | 48% |
| Employee Difference | 43% |
| Urbanicity | 40% |
| Percent of Renters Spend 35+% of Income on Rent | 39% |
| Log NFLIS | 35% |
| Percent of Homeowners Spend 35+% of Income on Mortgage | 32% |
| Poverty Rate | 25% |
| High School Graduation Rate | 20% |
| Log Jail Population | 17% |
| Buprenorphine Provider Waivers | 11% |
| Unemployment Rate | 10% |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript