

## Practice of Epidemiology

# Identifying Predictors of Opioid Overdose Death at a Neighborhood Level With Machine Learning

**Robert C. Schell\***, Bennett Allen, William C. Goedel, Benjamin D. Hallowell, Rachel Scagos, Yu Li, Maxwell S. Krieger, Daniel B. Neill, Brandon D. L. Marshall, Magdalena Cerda, and Jennifer Ahern

\*Correspondence to Robert C. Schell, Division of Health Policy and Management, School of Public Health, University of California, Berkeley, Berkeley Way West, 2121 Berkeley Way, 5th Floor, Berkeley, CA 94704 (e-mail: schell\_bobby@berkeley.edu).

Initially submitted April 22, 2021; accepted for publication November 16, 2021.

Predictors of opioid overdose death in neighborhoods are important to identify, both to understand characteristics of high-risk areas and to prioritize limited prevention and intervention resources. Machine learning methods could serve as a valuable tool for identifying neighborhood-level predictors. We examined statewide data on opioid overdose death from Rhode Island (log-transformed rates for 2016–2019) and 203 covariates from the American Community Survey for 742 US Census block groups. The analysis included a least absolute shrinkage and selection operator (LASSO) algorithm followed by variable importance rankings from a random forest algorithm. We employed double cross-validation, with 10 folds in the inner loop to train the model and 4 outer folds to assess predictive performance. The ranked variables included a range of dimensions of socioeconomic status, including education, income and wealth, residential stability, race/ethnicity, social isolation, and occupational status. The  $R^2$  value of the model on testing data was 0.17. While many predictors of overdose death were in established domains (education, income, occupation), we also identified novel domains (residential stability, racial/ethnic distribution, and social isolation). Predictive modeling with machine learning can identify new neighborhood-level predictors of overdose in the continually evolving opioid epidemic and anticipate the neighborhoods at high risk of overdose mortality.

addiction; epidemiologic methods; machine learning; neighborhoods; opioids; overdose; prediction

Abbreviations: CBG, census block group; LASSO, least absolute shrinkage and selection operator; OUD, opioid use disorder.

Despite substantial public health investment, the opioid overdose crisis continues unabated in the United States (1). Opioid-involved overdose claimed the lives of 90 Americans per day in 2015, rising to a staggering 137 per day in 2019 (1, 2). Prescriptions for opioid analgesics have decreased precipitously since their peak in 2012, and greater proportions of opioid use and opioid-involved mortality are now attributable to heroin, illicitly manufactured fentanyl, and other synthetic opioids (3). Synthetic opioids increasingly adulterate the illicit drug supply, greatly increasing overdose risk (4). In some jurisdictions, including New England, fentanyl has largely supplanted heroin in the drug supply (5).

Due to the difficulty of monitoring the continually evolving illicit drug supply and rapidly changing overdose risk environment, proactive interventions focused at the commu-

nity level, including naloxone distribution, street outreach, expansion of opioid use disorder (OUD) treatment services, and provision of mobile and low-threshold OUD treatment, have become more important than ever (6, 7). However, at present, community-level increases in opioid-involved overdose deaths are detected in hindsight, with data often derived from autopsy reports, and substantial time lags limit their utility (8). Syndromic surveillance of nonfatal opioid overdoses (e.g., using emergency department or emergency medical services data sources) is more timely, but case definitions are inconsistently applied across jurisdictions (9). To improve understanding of specific characteristics that place communities most at risk for opioid overdose and other drug-related harms, and to inform where best to prioritize limited public health and harm reduction resources, researchers and

practitioners must better understand the predictors of opioid overdose death at the community level. Because of their ability to detect patterns in data and account for complex variable interactions, machine learning methods could serve as a valuable tool in identifying these community-level predictors (10).

Most studies to date that have utilized machine learning to identify predictors of OUD and opioid-involved overdose mortality have focused on individual-level data in clinical settings (11–16). More recent predictive modeling efforts have recognized the importance of social factors in driving opioid overdose and have incorporated data from criminal justice records and human services providers and self-reported data on socioeconomic status and illicit drug use (17–19). These individual-level predictive models consistently find that inclusion of social factors achieves meaningful improvements in model performance. Researchers have also applied machine learning methods to predict the future spatial distribution of overdose deaths and to detect emerging geographic and demographic patterns of overdoses (20).

However, to our knowledge, none of these predictive modeling efforts have attempted to identify which specific community-level predictors could help in understanding the geographic distribution of opioid overdose mortality risk (20, 21). Prior research on neighborhood characteristics associated with opioid overdose mortality risk has identified potential domains of interest, including occupational factors, income and wealth, and racial/ethnic segregation (22–24). While these analyses are illuminating, they neglect the relative influence, complex interactions, and high degree of correlation between these community-level predictors and so provide an incomplete picture of the factors most predictive of overdose death (25–27). A significant gap in the literature exists in analyzing, synthesizing, and prioritizing the rapidly changing community-level predictors of the opioid overdose crisis.

A machine learning approach to examining community-level predictors of overdose facilitates both the ranking of existing domains and the identification of new domains that predict recent trends in overdose death rates while accounting for the high degree of correlation and interaction between different domains. Herein, we employ machine learning techniques involving 206 neighborhood-level demographic variables derived from the US Census Bureau's American Community Survey from 2016–2019. The predictors encompass domains such as educational attainment, income, disability, employment, and other measures of the built environment (e.g., age of housing). We use these variables to predict the neighborhood-level distribution of opioid overdose deaths in Rhode Island, the state with the 11th highest rate of opioid overdose mortality nationally as of 2018 (28). We leverage double cross-validation to select a model without overfitting. For the outer loop, the data are divided into 4 folds. Each fold is used to evaluate the out-of-sample performance of a predictive model trained using the other 3 folds, and performance is averaged over the 4 folds. For the inner loop, 10-fold cross validation (within the training data) is used to tune the hyperparameters of the least absolute shrinkage and selection operator (LASSO) and random forest algorithms. As an additional sensitivity analysis, we substitute

elastic net regression for the LASSO; results are reported in Web Table 1 (available at <https://doi.org/10.1093/aje/kwab279>). This study follows the Transparent Reporting of a Multivariable Prediction Model (TRIPOD) checklist, available in Web Table 2.

## METHODS

### Opioid overdose death rates

Our unit of analysis was the census block group (CBG), the smallest geographic unit for which the US Census Bureau publishes sample data (29). We chose CBGs as the level of analysis because they consist of small areas (600–3,000 residents) and have been found to be good proxies for neighborhoods in previous research (29, 30). We excluded special land-use CBGs, which are areas with minimal population used to denote important land features (29). Of the 815 CBGs present in Rhode Island, we excluded 7 because of excessive missingness and 66 special land-use CBGs, for a final analytical sample of 742 CBGs.

The outcome—the rate of opioid overdose death per 100,000 population at the CBG level—was calculated from data provided by the Rhode Island Department of Health using Rhode Island's State Unintentional Drug Overdose Reporting System (31). Information for this Centers for Disease Control and Prevention–funded program is abstracted from multiple data sources, including medical examiners' records, death certificates, law enforcement records, and toxicological results; a complete breakdown of case definitions is available online (32). As part of data collection, professional abstractors identify the precise location of injury, defined as the address closest to the location where the overdose occurred. For this analysis, staff at the Rhode Island Department of Health geocoded all cases to the CBG using a census geocoder.

We included all accidental opioid-involved overdose deaths that occurred in Rhode Island between July 1, 2016, and June 30, 2019. Where injury location was missing, we used residence location ( $n = 56$ ; 6.48%), and we excluded from analysis cases that were missing data on both location of death and residence ( $n = 7$ ; 0.81%). We calculated the annual rate of opioid overdose death per 100,000 residents by summing the number of opioid overdose deaths, dividing by the total population of the CBG according to the American Community Survey and number of years, and multiplying by 100,000 (33). Because of the existence of a small number of extreme outliers, we used  $\log(1 + x)$ , or the log-transformed opioid overdose death rate plus 1, as the primary outcome to account for zeros in the data; this practice has been shown to perform well in right-skewed, strictly positive data in previous studies (34, 35).

### Covariates

We selected a total of 206 variables from the Census Bureau's annual American Community Survey that represented a variety of domains, based on prior literature examining the neighborhood-level determinants of overdose risk;

all variables and definitions are available in Web Table 3 (22–27). We excluded covariates with over 5% missingness as a conservative method to reduce the dimensions of the data set. As a result, from the 206 variables in the original data set, we used 203 variables without excessive missingness for the final analysis. We used an average of the covariates' values, all continuous variables, from 2016 to 2019 as the predictors because these demographic characteristics change little over the course of 3 years.

### Statistical analysis

We developed a prediction model through double cross-validation by training community-level predictors in the 10 inner folds within each of the 4 outer folds with log-transformed annual opioid death rates averaged over the study period. We evaluated the performance of our model, averaged over the 4 outer folds of the double cross-validation design. We first implemented a LASSO algorithm to reduce variability in our estimates and multicollinearity caused by the existence of many strongly correlated covariates that explain little of the variation in the opioid death rate by CBG. The LASSO algorithm addresses the possibility of overfitting by setting covariates that do not meaningfully improve predictive performance to 0. This is achieved by minimizing the sum of the mean squared error (as in ordinary least squares linear regression) and a penalty term equal to some constant  $\lambda$  times the sum of the coefficients' absolute values. To select the optimal value of the penalty parameter  $\lambda$ , we performed an inner 10-fold cross-validation (within the training data) for each outer fold of the double cross-validated design, which resulted in an average optimal value across the 4 folds of 0.0998. The coefficients and performance derived from the LASSO regression are available in Web Table 4. After discarding the covariates set to 0 by the LASSO algorithm, 40 total covariates remained across the 4 folds. Descriptive statistics for these 40 covariates in the overall data set are available in Web Table 5 (36, 37).

We next used the random forest algorithm for its ability to handle high-dimensional data with complex interactions and to rank covariates by variable importance. While other tree-based methods like gradient boosting can perform variable importance ranking, we chose random forest because it performs comparatively well with this data and for its interpretability by a non-machine learning specialist audience. We averaged the permutation variable importance across folds of each of the remaining covariates and ranked them accordingly. We computed permutation variable importance by 1) taking the sum of the decrease in mean squared error from splitting on a covariate for the out-of-sample data; 2) permuting each variable to understand the mean squared error increase from excluding the variable; and 3) averaging and normalizing by the standard error and then scaling to produce a range from 0 to 100 (38). There are probably complex interactions between the variables, some that are intuitive (e.g., between income and education) and some that a researcher cannot anticipate. Thus, even with a modest number of interactions, the problem quickly becomes intractable for parametric modeling, where specifying more

parameters than available observations leads to overfitting (39, 40).

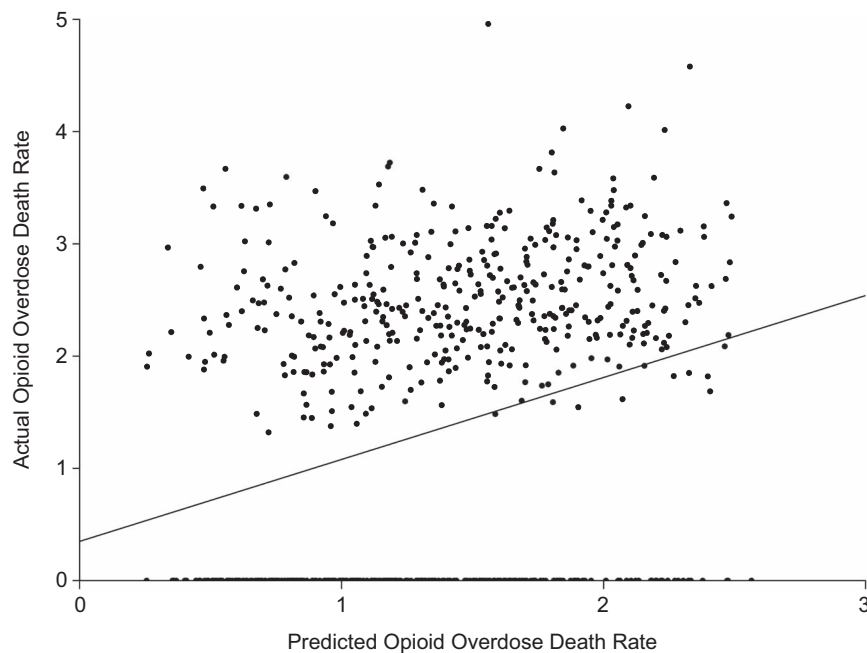
The mechanism behind the random forest algorithm is an ensemble, or collection, of nonparametric regression trees. In a simple regression tree, observations are split into 2 different nodes, or groups of observations based on different values of the covariate that cause the largest reduction of mean squared error in the outcome within groups after the split. After the first split, this process continues, and the data are further split into successive groups according to other important covariates that create increasingly homogenous groups.

While a nonparametric regression tree creates an intuitive prediction of opioid overdose death rates based on mean values in covariate subgroups, a tree's form is extremely sensitive to the variables the algorithm splits on. To understand this intuitively, imagine that the opioid overdose death rate is strongly predicted by 3 variables: low educational attainment, low household income, and median age. Low educational attainment is strongly correlated with low household income. As a result, conditional on having split by low household income, low education explains little further regarding the differences between groups, and vice versa. Thus, while 2 regression trees might explain similar levels of variation in the opioid overdose death rate, 1 tree might split on low household income first and the other on low education, after which the collinearity of the two variables would make the other seem unimportant. The random forest algorithm addresses this issue of high variability in the structure of individual trees by drawing bootstrapped samples of the training data and averaging the results of different regression trees conducted on each of these samples. Random forest reduces the correlation between regression trees by randomly drawing only a subset of covariates to split on at each node. To understand whether each covariate has a protective or harmful association with opioid overdose death, we used the sign of the coefficient from the LASSO algorithm to determine the direction of the relationship.

To train the random forest model, we performed hyperparameter tuning on the same inner folds as for the LASSO regression. Through cross-validation, we tuned the parameters by setting the number of trees at 1,000 or more for each model, and we found the optimal number of covariates randomly sampled at each split to be 21, on average, across the folds. The chosen set of model parameters was then used to fit the predictive model on the outer folds, at which point we evaluated out-of-sample prediction performance. We relied on mean squared error as the measure by which to gauge prediction accuracy. All analyses were performed in R 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria), and the analysis code is available online (41). The study was approved by the Brown University and Rhode Island Department of Health institutional review boards.

### RESULTS

A total of 863 overdose deaths attributable to opioids occurred in Rhode Island from January 2016 to June 2019, of which 741 (85.9%) were eligible for inclusion in this



**Figure 1.** Calibration of predicted and actual log rates of opioid overdose death (number of deaths per 100,000 population) at the census block group level, American Community Survey, 2016–2019. The calibration plot has a slope of 0.73 and an intercept of 0.35.

analysis. Over the 42-month study period, a median of 1 opioid overdose death occurred per CBG, with an interquartile range of 0–1 and a range of 0–12. Opioid overdose deaths occurred at higher rates in the urban areas of Rhode Island (31 deaths/100,000 population), which consist of the Providence metropolitan area and much of the eastern border of the state, than in less urban areas (22 deaths/100,000 population).

The variables selected by LASSO and ordered by the random forest pertained to a wide range of the dimensions of socioeconomic status, including 9 covariates related to education, 10 related to income and wealth, 6 related to residential stability, 2 related to race/ethnicity, 3 related to occupational status, 4 related to social isolation, 3 related to age, and 2 related to sex, as defined by the US Census Bureau (descriptive statistics are provided in Web Table 5, and a full list of variables considered in the LASSO algorithm is provided in Web Table 3). The remaining variable, percentage of residents who did not speak English, did not fall into any of these categories. Web Table 4 shows the coefficients from this regression. Web Figure 1 shows the covariates with the highest variable importance metrics (grouped by domain and direction of coefficient), and Web Figure 2 shows the same results for the elastic net sensitivity analysis. The  $R^2$  and root mean squared error of the model on testing data were 0.17 and 16.2, respectively, and Figure 1 shows the calibration plot. A perfectly calibrated model would produce a slope of 1, whereas our model's slope was 0.73, as the model tended to overestimate overdose deaths in CBGs with 0 cases and underestimate overdose deaths in CBGs at the upper end of the distribution.

## DISCUSSION

While existing predictive models that focus on overdose risk at an individual level can assist in clinical decision-making, our focus on neighborhood-level predictors is better suited to optimizing the allocation of public health resources and targeted community-focused interventions. In this analysis of neighborhood-level predictors of fatal opioid overdose burden in Rhode Island, we identified a set of 40 predictors that together explained approximately 17% of the variance in fatal overdose rates between CBGs. While many of the predictors fell into domains established from the existing literature, like education, income/wealth, and occupational status (22, 23, 25–27), they also included important new domains, including residential stability, racial/ethnic distribution, and social isolation. Importantly, these factors predicted overdose death burden at the *neighborhood* level using injury location and may not necessarily reflect traits or risks experienced by individual decedents.

The high variable importance and positive LASSO regression coefficient suggest that one of the most important covariates identified—percentage of men with only a high school education—is associated with higher rates of opioid overdose death. Educational attainment serves as an important proxy measure of social status, and recent research suggests that men without a college education and/or residing in areas with little economic opportunity face markedly higher rates of substance use and related disorders (42–44). The combination of the variable importance ranking and the protective relationship between increased educational attainment and opioid overdose death at the CBG



level implies that fatal overdoses concentrate in areas with lower socioeconomic status among men, a trend replicated by the importance of income and occupation as predictors (described below). Consistent with this pattern, the percentage of the population with some college education had a high variable importance and a negative association with opioid overdose death rates.

Income and wealth, captured by measures of household income, median home values, poverty, and car ownership, have also been identified in past research on the opioid crisis that identified deaths at the county and state levels as occurring most frequently in areas of lower income and high unemployment (22, 25). In this analysis, areas that had a larger percentage of households with incomes under \$50,000 per year, lower median housing values, and lower car ownership tended also to have higher rates of opioid overdose death. Consistent with this pattern, a higher proportion of households earning at least \$100,000 per year was associated with fewer opioid overdose deaths. Several hypotheses exist regarding the mechanism for this relationship, including drug use to manage chronic stress due to economic hardship, previously identified by studies of other types of drug use (45). Two of the income variables identified denoted households living at or below the poverty line, and both predicted higher opioid death rates; this is consistent with research that found severe economic stress had a strong association with heroin-involved deaths in urban zip codes in the United States (25).

Occupational characteristics were represented in 3 of the covariates selected, including percentage of men in production occupations and percentages of men and people in management occupations. A greater number of men in production occupations was positively associated with higher overdose death rates, while greater numbers of men and people in management occupations were protective. This coheres with recent work on overdose deaths by occupation in Massachusetts and Rhode Island, which found that workers in occupations with limited sick leave and insurance coverage, especially construction, production, and service workers, were far more likely to suffer an opioid-related overdose death than their higher-earning peers at less risk of injury (26, 32). This relationship likely exists because work injuries in these sectors are often treated with opioids, which results in neighborhoods with both more prescription opioid use and a greater supply of prescription opioids in circulation (32).

While not previously established as a driver of opioid overdose deaths, variables that capture information on residential instability served as important predictors of overdose mortality in our analysis. Residential instability has an independent and synergistic effect relative to neighborhood and individual-level poverty on the health of individuals in a neighborhood (46, 47). In social disorganization theory, residential instability is hypothesized to have a negative effect on health through a lack of institutional strength, limited social network interaction, and lesser feelings of neighborhood attachment (46). Furthermore, an established body of literature shows that people experiencing homelessness face risks of OUD and opioid-involved mortality far in excess of those with stable housing (47).

In combination, these factors could increase the population-level burden of OUD and opioid-related death. The covariates denoting residential stability, including percentage of renter-occupied households, percentage of owner-occupied households, percentage of housing units that experienced turnover in 2010 and in the 1990s, and percentage of houses for sale that are vacant, probably correlate with both relative economic hardship and residential instability, which could each have an independent association with opioid overdose deaths in a CBG. This hypothesis is supported by the positive associations between opioid overdose death rate and a higher percentage of renter-occupied housing, a higher percentage of houses for sale that are vacant, and higher rates of recent housing unit turnover. Consistent with these results, the percentage of owner-occupied households and housing turnover in the 1990s were both associated with fewer opioid overdose deaths.

From the late 1990s to approximately 2010, rates of OUD were higher among Whites; however, overdose death rates among Blacks and Hispanics have increased considerably in recent years (27). This model detected a significant positive association between the percentage of neighborhood residents who are Black and opioid overdose death rates. The inclusion of these variables in the variable importance graphs reflects a potential change in the demographic profile of individuals at risk for opioid overdose and underscores contributions of systemic racism as a significant driver of drug overdose risk in communities of color (48). This demonstrates machine learning's ability to detect both previously theorized associations and emergent trends that surpass current theoretical knowledge.

Lastly, measures of social isolation, long understood to play an important role in OUD, also have a strong association with increased opioid overdose death rates at the neighborhood level (49). Specifically, our model suggests that communities with more people living alone or unmarried have higher opioid overdose death rates. These are the same covariates as those identified in previous cross-sectional analyses of the risk of opioid overdose (49). Due in part to the multidimensional nature of social isolation, its exact role in community opioid overdose mortality remains underexplored; additional research is needed to elucidate the causal mechanisms.

There are several limitations to the current study. First, this analysis focused on CBG-level covariates as predictors of neighborhood-level opioid overdose mortality rates. While the  $R^2$  value of our predictive model was low relative to those presented in individual-level studies with clinical data, the focus on prediction at an aggregate level using only socioeconomic determinants made highly accurate prediction far more challenging (50). Still, in the absence of comparable efforts focused at the neighborhood level, this relatively low  $R^2$  could affect the current model's utility. While the predictive power of the approach remains low, a model incorporating other data sources could pave the way for focused community-level interventions in neighborhoods at high risk of overdose outbreaks. This type of predictive modeling may help policy-makers anticipate types of communities that could have a higher rate of overdose mortality, although it would be important to examine

these relationships in settings beyond Rhode Island and prospectively as well. Additionally, the model would have to consider the fairness of resource targeting, a persistent concern in predictive modeling (10).

Second, the data were collected at the CBG level, a level that is more fine-grained than a census tract but still represents communities defined by the Census Bureau and not necessarily neighborhood divisions. Third, opioid overdose mortality as an outcome is less common and more variable than nonfatal opioid overdose or OUD more generally. This makes it both less stable and more challenging to predict. Fourth, the burden of opioid-related mortality may not precisely align with the burdens of opioid use and OUD in a particular area; however, from the perspective of policy-makers trying to minimize opioid-related harms, identification of predictors of opioid overdose death can provide crucial information on where the epidemic might have the most severe impacts. Lastly, this analysis relied on the location of the overdose (otherwise known as the location of injury) and not the location of residence, so it may not have targeted the neighborhoods in which victims actually resided. However, because area of residence and area of overdose often concord in Rhode Island—with 73.7% of events in the data set occurring in the victim's residential CBG—and most community-level interventions focus on intervening at the point of drug use, this is not a significant concern.

In summary, our machine learning approach identified many of the covariates implicated in the existing literature, albeit at a neighborhood level, and revealed potential new domains associated with opioid overdose death rates. The use of public data and the large number of standard covariates from the American Community Survey allow researchers to further explore the relationships identified in this analysis, as well as the covariate domains associated with opioid overdose death rates, in their own contexts of interest. The model's ability to identify important predictors of overdose death suggests that after incorporating other important ecological data, a neighborhood-level predictive model could become a valuable and interpretable asset for public health departments both in determining where to target resources for community-focused overdose prevention intervention and in understanding the characteristics of communities most heavily impacted by this ongoing crisis in their own jurisdictions. Rhode Island provides an instructive case-study setting for identifying predictors of overdose mortality, as a state with below-average opioid prescription rates but an opioid overdose death rate over twice the national average (51). While Rhode Island is only 1 state, the covariates identified in this analysis bear remarkable similarity to those identified in other places across the country (51, 52). Broader investigations should consider these relationships in diverse settings.

## ACKNOWLEDGMENTS

Author affiliations: Division of Health Policy and Management, School of Public Health, University of

California, Berkeley, Berkeley, California, United States (Robert C. Schell); Center for Opioid Epidemiology and Policy, Department of Population Health, Grossman School of Medicine, New York University, New York, New York, United States (Bennett Allen, Magdalena Cerda); Department of Epidemiology, School of Public Health, Brown University, Providence, Rhode Island, United States (William C. Goedel, Yu Li, Maxwell S. Krieger, Brandon D. L. Marshall); Center for Health Data and Analysis, Rhode Island Department of Health, Providence, Rhode Island, United States (Benjamin D. Hollowell, Rachel Scagos); Center for Urban Science and Progress, New York University, New York, New York, United States (Daniel B. Neill); Department of Computer Science, Courant Institute of Mathematical Sciences, New York University, New York, New York, United States (Daniel B. Neill); Robert F. Wagner Graduate School of Public Service, New York University, New York, New York, United States (Daniel B. Neill); and Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California, United States (Jennifer Ahern).

This work was supported by the National Institutes of Health (grants R01DA046620, T32-AG000246, and T32-LM-012417).

The American Community Survey is publicly available at <https://www.census.gov/programs-surveys/acs/>. Overdose death data were obtained from July 2016 to June 2019 through an approved request to the Rhode Island Department of Health (RIDOH) and cannot be shared to protect decedent privacy and confidentiality.

We thank Jesse Yedinak and Claire Pratty for their research and administrative assistance, as well as RIDOH Director Dr. Alexander Scott and many other RIDOH staff, for their insights and contributions to this project.

This work was presented at the 2021 annual meeting of the Society for Epidemiologic Research, held virtually on June 22–25, 2021.

The RIDOH is not responsible for the authors' analysis, opinions, or conclusions.

Conflict of interest: none declared.

## REFERENCES

1. Bonnie RJ, Kesselheim AS, Clark DJ. Both urgency and balance needed in addressing opioid epidemic: a report from the National Academies of Sciences, Engineering, and Medicine. *JAMA*. 2017;318(5):423–424.
2. Centers for Disease Control and Prevention. U.S. drug overdose deaths continue to rise; increase fueled by synthetic opioids. (Press release). <https://www.cdc.gov/media/releases/2018/p0329-drug-overdose-deaths.html>. Published March 29, 2018. Accessed November 11, 2020.
3. Kerr T. Public health responses to the opioid crisis in North America. *J Epidemiol Community Health*. 2019;73(5):377–378.
4. Ciccarone D. The triple wave epidemic: supply and demand drivers of the US opioid overdose crisis. *Int J Drug Policy*. 2019;71:183–188.

5. Ciccarone D. Fentanyl in the US heroin supply: a rapidly changing risk environment. *Int J Drug Policy*. 2017;46:107–111.
6. Walsh SL, El-Bassel N, Jackson RD, et al. The HEALing (Helping to End Addiction Long-term<sup>SM</sup>) Communities Study: protocol for a cluster randomized trial at the community level to reduce opioid overdose deaths through implementation of an integrated set of evidence-based practices. *Drug Alcohol Depend*. 2020;217:108335.
7. Green TC, Bratberg J, Dauria EF, et al. Responding to opioid overdose in Rhode Island: where the medical community has gone and where we need to go. *RI Med J (2013)*. 2014; 97(10):29–33.
8. Romano B. Tracking fentanyl in the drug supply. <https://heller.brandeis.edu/news/items/releases/2020/summer-magazine-green-fentanyl.html>. Published June 3, 2020. Accessed November 11, 2020.
9. Ising A, Proescholdbell S, Harmon KJ, et al. Use of syndromic surveillance data to monitor poisonings and drug overdoses in state and local public health agencies. *Inj Prev*. 2016;22(suppl 1):i43–i49.
10. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA*. 2018;319(13):1317–1318.
11. Hylan TR, Von Korff M, Saunders K, et al. Automated prediction of risk for problem opioid use in a primary care setting. *J Pain*. 2015;16(4):380–387.
12. Lo-Ciganic WH, Huang JL, Zhang HH, et al. Evaluation of machine-learning algorithms for predicting opioid overdose risk among Medicare beneficiaries with opioid prescriptions. *JAMA Netw Open*. 2019;2(3):e190968.
13. Ellis RJ, Wang Z, Genes N, et al. Predicting opioid dependence from electronic health records with machine learning. *BioData Min*. 2019;12(1):3.
14. Hasan MM, Noor-E-Alam M, Patel MR, et al. A big data analytics framework to predict the risk of opioid use disorder [preprint]. *arXiv*. 2019. (<https://arxiv.org/abs/1904.03524v3>). Accessed December 16, 2020.
15. Reps JM, Soledad Cepeda M, Ryan PB. Wisdom of the CROUD: development and validation of a patient-level prediction model for opioid use disorder using population-level claims data. *PLoS One*. 2020;15(2):e0228632.
16. Sun JW, Franklin JM, Rough K, et al. Predicting overdose among individuals prescribed opioids using routinely collected healthcare utilization data. *PLoS One*. 2020; 15(10):e0241083.
17. Saloner B, Chang HY, Krawczyk N, et al. Predictive modeling of opioid overdose using linked statewide medical and criminal justice data. *JAMA Psychiatry*. 2020;77(11): 1155–1162.
18. Han DH, Lee S, Seo DC. Using machine learning to predict opioid misuse among U.S. adolescents. *Prev Med*. 2020; 130:105886.
19. Lo-Ciganic WH, Donohue JM, Hulsey EG, et al. Integrating human services and criminal justice data with claims data to predict risk of opioid overdose among Medicaid beneficiaries: a machine-learning approach. *PLoS One*. 2021; 16(3):e0248360.
20. Neill DB, Herlands W. Machine learning for drug overdose surveillance. *J Technol Hum Serv*. 2018;36(1):8–14.
21. Ertugrul AM, Lin YR, Taskaya-Temizel T. CASTNet: Community-Attentive Spatio-Temporal Networks for Opioid Overdose Forecasting. In: Brefeld U, Fromont E, Hotho A, et al., eds. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part III*. (Lecture Notes in Computer Science). New York, NY: Springer Publishing Company; 2020:432–448.
22. Rudolph KE, Kinnard EN, Aguirre AR, et al. The relative economy and drug overdose deaths. *Epidemiology*. 2020; 31(4):551–558.
23. Cerdá M, Ransome Y, Keyes KM, et al. Revisiting the role of the urban environment in substance use: the case of analgesic overdose fatalities. *Am J Public Health*. 2013;103(12): 2252–2260.
24. Cerdá M, Gaidus A, Keyes KM, et al. Prescription opioid poisoning across urban and rural areas: identifying vulnerable groups and geographic areas. *Addiction*. 2017;112(1): 103–112.
25. Pear VA, Ponicki WR, Gaidus A, et al. Urban-rural variation in the socioeconomic determinants of opioid overdose. *Drug Alcohol Depend*. 2019;195:66–73.
26. Hawkins D, Roelofs C, Laing J, et al. Opioid-related overdose deaths by industry and occupation—Massachusetts, 2011–2015. *Am J Ind Med*. 2019;62(10):815–825.
27. Allen B, Nolan ML, Kunins HV, et al. Racial differences in opioid overdose deaths in New York City, 2017. *JAMA Intern Med*. 2019;179(4):576–578.
28. National Center for Health Statistics. Drug overdose mortality by state. [https://www.cdc.gov/nchs/pressroom/sosmap/drug\\_poisoning\\_mortality/drug\\_poisoning.htm](https://www.cdc.gov/nchs/pressroom/sosmap/drug_poisoning_mortality/drug_poisoning.htm). Accessed December 21, 2020.
29. Bureau of the Census, US Department of Commerce. Block groups for the 2020 census—final criteria. Federal Register 83, no. 219 (November 13, 2018):56293–56298. <https://www.federalregister.gov/documents/2018/11/13/2018-24570/block-groups-for-the-2020-census-final-criteria>. Accessed November 11, 2020.
30. Roux AVD, Merkin SS, Arnett D, et al. Neighborhood of residence and incidence of coronary heart disease. *N Engl J Med*. 2001;345(2):99–106.
31. Jiang Y, McDonald JV, Goldschmidt A, et al. State unintentional drug overdose reporting surveillance: opioid overdose deaths and characteristics in Rhode Island. *RI Med J*. 2018;101(7):25–30.
32. Scagos R, Lasher L, Viner-Brown S. Accidental or undetermined opioid-involved drug overdose deaths in Rhode Island and usual occupation—higher rates observed in natural resources, construction, and maintenance occupations. *RI Med J (2013)*. 2019;67(33):925–930.
33. Bureau of the Census, US Department of Commerce. American Community Survey data. <https://www.census.gov/programs-surveys/acs/data.html>. Published 2012. Accessed January 9, 2021.
34. Zhu L, Gorman DM, Horel S. Alcohol outlet density and violence: a geospatial analysis. *Alcohol Alcohol*. 2004;39(4): 369–375.
35. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ*. 2001;20(4): 461–494.
36. Kim H, Lee SH, Lee SE, et al. Depression prediction by using ecological momentary assessment, Actiwatch data, and machine learning: observational study on older adults living alone. *JMIR Mhealth Uhealth*. 2019;7(10):e14149.
37. Filzmoser P, Liebmann B, Varmuza K. Repeated double cross validation. *J Chemom*. 2009;23(4):160–171.
38. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods*. 2009;14(4):323–348.

39. Breiman L. Random forests. *Mach Learn*. 2001;45(1): 5–32.
40. Grömping U. Variable importance assessment in regression: linear regression versus random forest. *Am Stat*. 2009;63(4): 308–319.
41. Schell RC. AJE-Identifying-Predictors-of-Opioid-Overdose-Death-at-a-Neighborhood-Level-with-Machine-Learning. 2021. <https://github.com/BobbySchell>. Published November 7, 2021. Accessed November 7, 2021.
42. Dasgupta N, Beletsky L, Ciccarone D. Opioid crisis: no easy fix to its social and economic determinants. *Am J Public Health*. 2018;108(2):182–186.
43. Gilson TP, Shannon H, Freiburger J. The evolution of the opiate/opioid crisis in Cuyahoga County. *Acad Forensic Pathol*. 2017;7(1):41–49.
44. Rembert M, Betz M, Feng B, et al. *Taking Measure of Ohio's Opioid Crisis*. (C. William Swank Program in Rural-Urban Policy—October 2017). Columbus, OH: The Ohio State University; 2017.
45. Boardman JD, Finch BK, Ellison CG, et al. Neighborhood disadvantage, stress, and drug use among adults. *J Health Soc Behav*. 2001;42(2):151–165.
46. Browning CR, Cagney K. Moving beyond poverty: neighborhood structure, social processes, and health. *J Health Soc Behav*. 2003;44(4):552–571.
47. Yamamoto A, Needleman J, Gelberg L, et al. Association between homelessness and opioid overdose and opioid-related hospital admissions/emergency department visits. *Soc Sci Med*. 2019;242:112585.
48. James K, Jordan A. The opioid crisis in Black communities. *J Law Med Ethics*. 2018;46(2):404–421.
49. Day BF, Rosenthal GL. Social isolation proxy variables and prescription opioid and benzodiazepine misuse among older adults in the U.S.: a cross-sectional analysis of data from the National Survey on Drug Use and Health, 2015–2017. *Drug Alcohol Depend*. 2019;204:107518.
50. Bhavsar NA, Gao A, Phelan M, et al. Value of neighborhood socioeconomic status in predicting risk of outcomes in studies that use electronic health record data. *JAMA Netw Open*. 2018;1(5):e182716.
51. National Institute on Drug Abuse. Rhode Island: opioid-involved deaths and related harms. <https://www.drugabuse.gov/drug-topics/opioids/opioid-summaries-by-state/rhode-island-opioid-involved-deaths-related-harms>. Published April 3, 2020. Accessed November 11, 2020.
52. Altekruze SF, Cosgrove CM, Altekruze WC, et al. Socioeconomic risk factors for fatal opioid overdoses in the United States: findings from the Mortality Disparities in American Communities Study (MDAC). *PLoS One*. 2020; 15(1):e0227966.