Preventive Medicine 194 (2025) 108276

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Evaluating the predictive performance of different data sources to forecast overdose deaths at the neighborhood level with machine learning in Rhode Island

John C. Halifax [a,*], Bennett Allen [b], Claire Pratty [c], Victoria Jent [b], Alexandra Skinner [c], Magdalena Cerdá [b], Brandon D.L. Marshall [c], Daniel B. Neill [d,e,f], Jennifer Ahern [a]

[a] *Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, USA*
[b] *Center for Opioid Epidemiology and Policy, Department of Population Health, Grossman School of Medicine, New York University, New York, NY, USA*
[c] *Department of Epidemiology, School of Public Health, Brown University, Providence, RI, USA*
[d] *Center for Urban Science and Progress, New York University, New York, NY, USA*
[e] *Department of Computer Science, Courant Institute for Mathematical Sciences, New York University, New York, NY, USA*
[f] *Robert F. Wagner Graduate School of Public Service, New York University, New York, NY, USA*

ARTICLE INFO

Keywords:
Overdose
Forecast
Machine learning
Community health
Substance use

ABSTRACT

Objectives: To evaluate the predictive performance of different data sources to forecast fatal overdose in Rhode Island neighborhoods, with the goal of providing a template for other jurisdictions interested in predictive analytics to direct overdose prevention resources.
Methods: We evaluated seven combinations of data from six administrative data sources (American Community Survey (ACS) five-year estimates, built environment, emergency medical services non-fatal overdose response, prescription drug monitoring program, carceral release, and historical fatal overdose data). Fatal overdoses in Rhode Island census block groups (CBGs) were predicted using two machine learning approaches: linear regressions and random forests embedded in a nested cross-validation design. We evaluated performance using mean squared error and the percentage of statewide overdoses captured by CBGs forecast to be in top percentiles from 2019 to 2021.
Results: Linear models trained on ACS data combined with one other data source performed well, and comparably to models trained on all available data. Those including emergency medical service, prescription drug monitoring program, or carceral release data with ACS data achieved a priori goals for percentage of statewide overdoses captured by CBGs prioritized by models on average.
Conclusions: Prioritizing neighborhoods for overdose prevention with forecasting is feasible using a simple-to-implement model trained on publicly available ACS data combined with only one other administrative data source in Rhode Island, offering a starting point for other jurisdictions.

1. Introduction

The United States continues to experience an unprecedented overdose crisis. Between 2002 and 2022, the drug overdose mortality rates in the United States increased four-fold, with drug overdose deaths surpassing 100,000 in both 2021 and 2022 (Spencer et al., 2024). The lack of accessible, stable, and preventative care for people who use drugs (PWUD) in the United States too often results in healthcare that is reactive, unnecessarily expensive, and inadequate (Buresh et al., 2021; Adeniran et al., 2023; Vohra et al., 2022). In response, community-based overdose prevention services to meet PWUD where they are is a practical and cost-effective care delivery model for wound care, infectious disease management, substance use disorder treatment access, and other overdose prevention services (e.g. naloxone distribution) (Hill et al., 2023; Springer, 2023; Robinowitz et al., 2014; Chan et al., 2021; Rosecrans et al., 2022; Pepin et al., 2023).

While funding for overdose prevention and response has increased in the United States, public health resources are still subject to finite limits. Practitioners need strategies to ensure that they deploy available resources efficiently to maximally reduce overdose mortality. Data science

* Corresponding author at: 2121 Berkeley Way, Fifth Floor, Berkeley, CA 94704, USA.
E-mail address: john_halifax@berkeley.edu (J.C. Halifax).

https://doi.org/10.1016/j.ypmed.2025.108276
Received 22 October 2024; Received in revised form 26 March 2025; Accepted 28 March 2025
Available online 29 March 2025

methods can be used to prioritize communities at highest risk of future overdose deaths, thereby improving the efficiency of care delivery for PWUD, particularly in the context of limited resources (Saloner et al., 2018; Wilt et al., 2019). Machine learning techniques have advanced accurate prediction of public health outcomes at both the individual and neighborhood level, with performance typically enhanced by utilizing a variety of data sources (Mhasawade et al., 2021). Prior scholarship has established the capacity of multiple machine learning approaches to forecast overdoses spatially, and the use of such forecasts to improve public health intervention deployment is currently under study statewide in Rhode Island and elsewhere (Neill and Herlands, 2018; Schell et al., 2022; Marshall et al., 2022; Allen et al., 2023; Allen et al., 2024).

The use of neighborhood-level predictive modelling to guide overdose crisis response has potential for applications beyond Rhode Island, but portability considerations have not been well explored. Rhode Island is an excellent environment to trial predictive modelling approaches due to its centralized, statewide public health infrastructure. Public health data and funding at all levels in the state are overseen by a single health authority that coordinates action in partnership with a network of community-based harm reduction organizations, which in turn have strong ties to technical experts at local academic institutions (Yedinak et al., 2024).

However, such “ideal” conditions are not necessarily generalizable to other jurisdictions. Where public health activities are governed by multiple state and local agencies, access to multiple data sources may be infeasible. Overdose prediction efforts by public health departments may also be hindered by a lack of technical expertise. A previous machine learning effort in Rhode Island did not prioritize usability or accessibility of the model for non-technical stakeholders, solely focusing on maximizing predictive performance (Allen et al., 2024). We consider this model, coded in two languages and resource-intensive to train, inaccessible as a template for non-technical public health stakeholders, highlighting the need for an accessible template (Allen et al., 2024). To inform the feasibility of using predictive analytics to guide overdose prevention intervention by local authorities across the United States and beyond, this study used Rhode Island data from several commonly used overdose and public health surveillance data sources to A) assess which commonly collected sociodemographic, public health, and overdose monitoring data sources have the best predictive performance in the Rhode Island context; and B) apply simple-to-implement machine learning and univariate techniques to assess performance and to provide a template (including statistical code in R) for application in new settings. We hypothesized that different combinations of data sources would have varying degrees of predictive performance in forecasting overdose deaths.

2. Methods

2.1. Data sources and preparation

This study used data from Rhode Island from January 1, 2019 to December 31, 2021. The census block group (CBG) was used as the unit of analysis, as it is the smallest geographical unit for which both Rhode Island overdose mortality and United States census data are available. Further, CBGs were considered a realistic geographic unit for Rhode Island practitioners to locally target public health interventions in communities, as they consist of approximately 600–3000 residents and are established proxies for neighborhoods (Roux et al., 2001). Boundaries used for CBGs were from the 2010 United States Census; data from 2020 and 2021 were geo-coded or cross-walked to 2010 census boundaries. Data for CBGs were aggregated at six-month intervals (with period one defined as January 1st-June 30th, period two defined as July 1st-December 31st) (Manson et al., 2021). A period of six months was identified by Rhode Island practitioners as a realistic time frame to implement interventions in prioritized neighborhoods (Marshall et al., 2022). Rhode Island has 815 total CBGs; six of these were excluded from the analysis due to zero population size or special land use (e.g., airport) status.

The outcome in all models was the count of unintentional drug overdose deaths for each CBG in the subsequent six-month prediction period, obtained from Rhode Island’s State Unintentional Drug Overdose Reporting Surveillance (SUDORS) (Jiang et al., 2018). SUDORS is a national system operating in 49 states and District of Columbia in the United States that includes detailed data on the location and other characteristics of overdose deaths. Wherever possible, we used the location of injury to geocode overdose deaths, with the address of decedents’ residence used if this information was missing. CBG-level covariates came from five other sources (Table 1): the five-year American Community Survey (ACS) data (Bureau, 2025); Built Environment data (collated from Rhode Island Geographic Information System (Rhode Island Geographic Information System, 2025), Substance Use and Mental Health Services Administration treatment center data (Substance Abuse and Mental Health Services Administration, 2025), and Rhode Island state business and healthcare state license data (Rhode Island Department of Business Regulation, 2025; Rhode Island Department of Health, 2025)); Emergency Medical Service (EMS) overdose data for responses to suspected, nonfatal, opioid-involved overdoses (Hallowell et al., 2021); Prescription Drug Monitoring Program (PDMP) data (Rhode Island Department of Health, 2022); and Rhode Island Department of Corrections (RIDOC) inmate release data (Rhode Island Department of Corrections, 2023). SUDORS data were used as a sixth source to provide a single covariate: current period all drug overdose death count. Data sources were limited to high-quality administrative sources that were comprehensive for the state of Rhode Island to avoid geographic discrepancies in data quality to avoid biased predictions. Descriptions of these data sources, rationales for their predictive utility, and their availability beyond Rhode Island are included in Table 1.

To evaluate predictive performance consistency over time, two evaluation windows were utilized. A schematic of the evaluation windows can be found in Fig. 1. Window one utilized training data that consisted of covariates from 2019 period one, 2019 period two, and 2020 period one to predict respective next period fatal overdoses, with covariates from 2020 period two held out as a corresponding test set. Window two utilized training data that consisted of covariates from 2019 period two, 2020 period one, and 2020 period two to predict the respective next period fatal overdoses, with covariates from 2021 period one held out as a corresponding test set. Covariates with greater than 5 % missingness across all periods from 2019 to 2021 were excluded. Data were split into the training and test sets and next period all drug overdose death counts prepared for each evaluation window separately. Missing data in the training and test sets were imputed separately with the CBG-specific mean value for the variable.

2.2. Data subset preparation and predictive approaches

Due to its accessibility and national ubiquity in the United States, ACS data were used as the starting set of covariate data in all analyses (outside the United States analogous national socio-economic census data could be used). Data from the five other sources (Built Environment, EMS, PDMP, RIDOC, and SUDORS datasets) were appended to the ACS data one at a time to create five other datasets. A seventh dataset (hereby referred to as “All Data”) was created by merging data from all six sources.

Two predictive models were developed with the goal of being simple-to-implement for non-technical stakeholders. In brief, both models utilized a nested cross-validation design employing least absolute shrinkage and selective operator (LASSO) as a screening algorithm for variable selection. After variable selection, the first model used a linear model and the second a tuned random forest to make predictions. Technical details on model design are included in Supplementary Materials, with model schematics visualized in Supplementary Figs. 1 A and 1B. After model training, variable importance was analyzed by two

Table 1
Description of Data Sources Used as Training and Test Data for Predictive Models to Forecast Fatal Overdoses in Rhode Island 2019–2021.

Data source	Source description	Controlling agency	Rationale for use as predictors of fatal overdose	Collected ubiquitously in a standardized manner in the United States?	Free to access?
American community survey 5-year estimates (ACS)	Demographic, socio-economic, and housing structure census data	United states census Bureau	Fatal overdose has social and economic determinants, which the American community survey is uniquely positioned to capture	Yes	Yes
Built environment	Business license, healthcare license, and treatment center data	Rhode Island geographic information system, Rhode Island Department of Business Regulation, Substance Abuse and Mental Health Services Administration	Business and service locations, as well as substance use disorder treatment availability, may be associated with where people use drugs and are subsequently at risk for fatal overdose	No	No
Emergency medical service responses to nonfatal opioid-involved overdoses (EMS)	Emergency medical service responses to nonfatal overdose events involving opioids using a validated case definition (Hallowell et al., 2021), geocoded by the incident location.	Rhode island department of health	Nonfatal overdose is associated with future fatal overdose at the individual level, and areas with heightened response to nonfatal overdoses are likely to benefit from proactive care interventions	No, but emergency medical service data is uniformly collected nationally in the National Emergency Medical Services Information System database, to which a case definition for nonfatal overdose could be applied	No
Rhode Island prescription drug monitoring program (PDMP)	A statewide, centralized database collecting data for controlled substances prescriptions and prescription fills designed for clinician and pharmacist utilization	Rhode island department of health	Misuse of prescription opioids or other controlled medications, either in isolation or with other drug classes, can be one source of overdose deaths. The inclusion of buprenorphine prescription data also provides insight into areas with a burden of opioid use disorders.	No	No
Rhode island department of corrections Carceral releases (RIDOC)	Data for releases from incarceration, geocoded to where the person was released	Rhode island department of corrections	Recent incarceration is associated with future fatal overdose in Rhode Island at both the individual and neighborhood level	No	No
Rhode Island state unintentional drug overdose reporting system (SUDORS)	A Centers for Disease Control and Prevention-funded, state-based surveillance system to record and characterize fatal overdoses in Rhode Island, geocoded at the injury incident location	Rhode island department of health	Future fatal overdose was the outcome of interest and current fatal overdoses were thought likely to be predictive of future fatal overdoses.	Yes, it is available in 49 of 50 states and District of Columbia as of 2024	No

Notes: Listed as the controlling agencies are the United States or Rhode Island agencies used in this case study. For the Rhode Island or United States specific controlling agencies, there are likely analogous agencies in other jurisdictions.

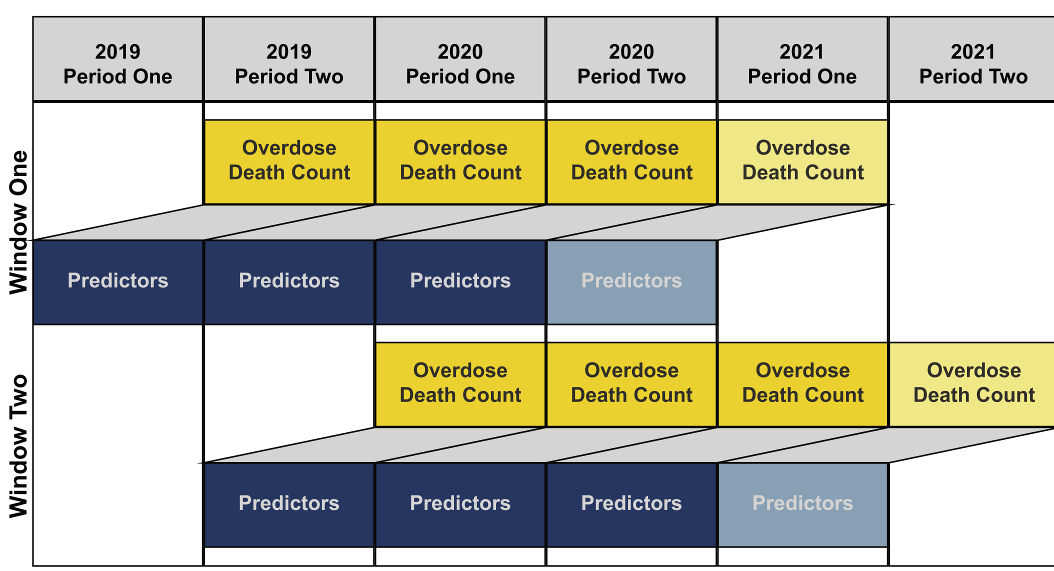

Fig. 1. Schematics of Evaluation Windows for Models Predicting Fatal Overdose for Census Block Groups in Rhode Island 2019–2021 *Notes:* Schematic of model evaluation windows used to train and test models. Training sets for each evaluation window are shown in solid fill, while test sets are shown as partially transparent. Periods One and Two refer to January 1st -June 30th and July 1st-December 31st of the given year respectively.

standard metrics specific to each model: absolute value of covariate coefficient *t* statistic for linear models and by increase in node purity for random forests (Grömping, 2015; Hastie et al., 2009).

To explore if comparable predictive performance could be achieved with traditional analysis, we trained univariate linear models to rank CBGs. One variable from each data source (e.g. EMS) per dataset (e.g. ACS + EMS) per model type was chosen based on maximum average importance value across the two evaluation windows. Univariate linear models were fit to training data, and predictions on respective test sets were used to rank CBGs. All analyses were performed in R 4.2.2 (R Core Team, 2021).

2.3. Model and univariate ranking evaluation

Each model made predictions using its respective test set of covariates. Models were evaluated by comparing these predictions to the actual test set fatal overdose counts. Model and univariate performance was assessed using two sets of metrics: first, we calculated mean squared error (MSE), a traditional model metric, for all approaches. Second, to contextualize the predictive performance of datasets and their associated models, we calculated the percentage of actual statewide fatal overdoses captured in the CBGs with predicted fatal overdose ranked greater than a desired percentile. This evaluation criterion is designed to evaluate model performance as it might be implemented in the applied public health setting (Allen et al., 2023). Resource-intensive interventions, such as street-based, low-threshold buprenorphine induction services, may only be feasible for a small percentage of neighborhoods (e.g. 5 %). Other, less resource-intensive interventions, such as naloxone distribution, could be scaled to more neighborhoods (e.g. 20 %). Understanding how different data sources and models perform to target different intervention scopes is useful to differentiate what datasets and variables to prioritize based on the scale of the planned intervention. Four variations of this metric, referred to as four "prioritization scenarios", were evaluated: percentages of total statewide fatal overdoses captured in the CBGs predicted to be greater than 95th, 90th, 85th, and 80th percentiles of predicted fatal overdoses (with 5 %–20 % of CBGs prioritized respectively). For each percentile, the a priori goal of percentage of fatal overdose captured is double the prioritized percentage of CBGs (i.e. if 20 % of CBGs are being prioritized, the model should be able to identify CBGs that capture 40 % or more of statewide fatal overdoses), as agreed upon with Rhode Island Department of Health and community partners (Marshall et al., 2022). Study procedures were approved by the Brown University Institutional Review Board.

3. Results

From January 1st, 2019 to December 31st, 2021, the period of analysis, Rhode Island experienced 1072 fatal overdoses, averaging 179 fatal overdoses per six-month period with fatal overdoses increasing moderately with time. Of the 352 covariates available when combining all six data sources, variable selection included 86 unique covariates across the 28 linear models and random forests. Supplementary Tables 1 and 2 rank these included covariates by variable importance across the linear models and random forests respectively. For linear models, the most influential variables were derived from across the Built Environment, RIDOC, ACS, PDMP, and EMS data sources. In contrast, the random forests relied more narrowly on variables from the PDMP, ACS, and RIDOC sources to predict future fatal overdose. Variable importance measures resulted in thirteen individual variables being selected for use in univariate rankings, with some variables being the most important variable across multiple dataset-model combinations.

Assessing overall model performance, MSE was lower (better) for linear models than random forest models trained on the same data sources and window in all cases. Shown in Fig. 2, the models trained on All Data had the lowest (i.e. best) MSE for both model classes.

Assessing approaches by the percentage of statewide fatal overdoses captured metrics, the models trained on All Data provided an upper benchmark of performance. Shown in Fig. 3, linear models slightly outperformed their random forest counterparts across the four prioritization scenarios on average across the two windows. Both predictive approaches performed well across all data combinations in the first three prioritization scenarios (top 5th-top 15th percentiles of CBGs), generally

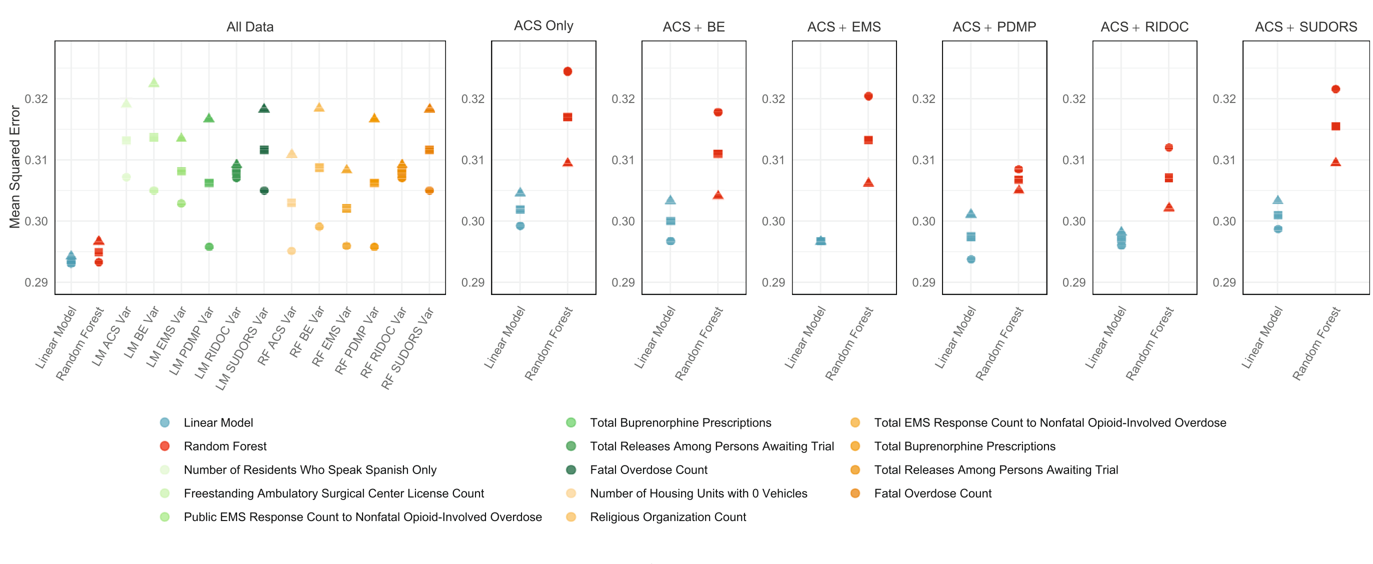

Fig. 2. Mean Squared Error by Data Source Combination for Models Predicting Fatal Overdose for Census Block Groups in Rhode Island Using Data from 2019 to 2021. *Notes:* Mean squared error for all models' predictions for CBG-level fatal overdoses in Rhode Island on test set data unused in model training in evaluation windows one and two. Univariate rankings by variables selected as most important from their data source from linear models and random forests trained on All Data are shaded green and orange respectively. See Supplemental Fig. 2 for full univariate results. CBG, Census block group. ACS, American Community Survey data. BE, Built Environment. EMS, Emergency medical service runs for opioid-involved nonfatal overdose data. LM, Linear Model. PDMP, Prescription drug monitoring program data. RF, Random Forest. RIDOC, Rhode Island Department of Corrections prisoner release data. SUDORS, State Unintentional Drug Overdose Reporting System prior period fatal overdose count data. Var, Variable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

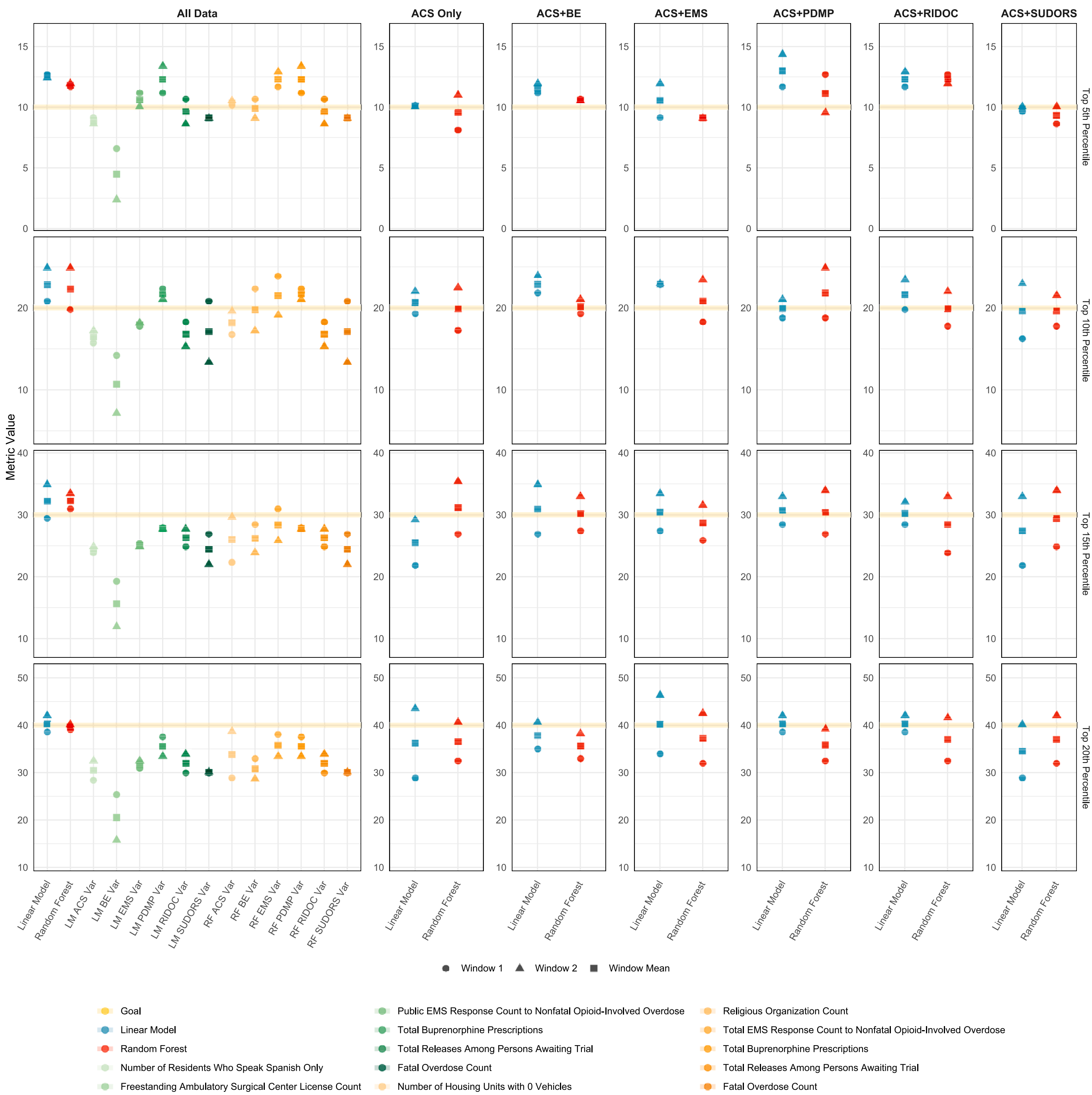

Fig. 3. Percent of Statewide Overdoses Captured by Census Block Groups Prioritized by Models in Rhode Island Using Data from 2019 to 2021. *Notes:* Percentage of statewide fatal overdoses in Rhode Island captured in the top 5th, 10th, 15th, and 20th percentiles of CBGs by predicted fatal overdose from the respective test sets in evaluation windows 1 and 2. Univariate rankings by variables selected as most important from their data source from linear models and random forests trained on All Data are shaded green and orang respectively. See Supplemental Fig. 3 for full univariate results. CBG, Census block group. ACS, American Community Survey data. BE, Built Environment. EMS, Emergency medical service runs for opioid-involved nonfatal overdose data. LM, Linear Model. PDMP, Prescription drug monitoring program data. RF, Random Forest. RIDOC, Rhode Island Department of Corrections prisoner release data. SUDORS, State Unintentional Drug Overdose Reporting System prior period fatal overdose count data. Var, Variable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

achieving prioritization goals on average. Scaling efficient CBG prioritization to the top 20 % proved more challenging. Four linear models trained on All Data, ACS + EMS, ACS + PDMP, and ACS + RIDOC were the only models to achieve the a priori goal on average, with no random forest achieving the a priori goal and only the random forest trained on All Data coming close. Performance of both approaches was generally better in window two, with the three linear models trained on All Data, ACS + PDMP, and ACS + RIDOC most consistent across the two windows for all four prioritization scenarios. Despite meeting the a priori goal on average in the top 20 % scenario, the linear model trained on ACS + EMS data did not do so in window one.

In terms of proportion of overdoses captured averaged over the two windows, univariate rankings generally underperformed compared to models, but often the performances were quite similar. For clarity, only univariate results for variables selected from the All Data models are shown in Figs. 2 and 3, with full results in Supplementary Figs. 2 and 3. In limited instances, univariate rankings in a single window outperformed one or both models trained on the corresponding dataset, with best performing variables derived from the ACS, EMS, and PDMP sources. These instances were generally in window one, where both model classes generally performed worse than window two. Better performance by the multivariable models is most evident in the top 15th and 20th percentile prioritization scenarios.

4. Discussion

This analysis establishes a template for utilizing and evaluating simple-to-implement univariate or machine learning methods to predict area-level fatal overdose, generating six-month forecasts that can guide overdose prevention efforts even when only limited data sources are available. In the Rhode Island context, linear regression approaches trained on only the ACS + EMS, ACS + PDMP, and ACS + RIDOC data sources on average performed well, and comparably to both the linear and random forest models trained on All Data. This finding suggests that, in the context of limited data availability, practitioners might consider acquisition of at least one of emergency medical service responses for nonfatal opioid-involved overdose, prescription drug monitoring program, or carceral release data in conjunction with socio-economic census data to train simple prediction models to forecast fatal overdoses at the neighborhood level. This recommendation relies on the assumption that both the processes that drive overdose and the predictor data available in new jurisdictions are similar to those in Rhode Island. Future work in new jurisdictions should be mindful of the limits of this assumption and consider the differences between Rhode Island and other jurisdictions to adapt to new contexts. Wherever there are a variety of predictors readily available, it would be advisable to incorporate all of them into a predictive model that includes a variable selection step.

Results indicate that on average in Rhode Island the linear regression approach outperformed the more complex random forest algorithm and the simpler univariate rankings, providing a method to prioritize neighborhoods for overdose prevention efforts that is quick to train and accessible to interpret for non-technical stakeholders. Due to limitations in generalizing these specific findings, we recommend that future work in new jurisdictions still evaluate multiple prediction approaches ranging in complexity. The code provided in Supplementary Materials offers a starting place for such evaluation.

The standout predictive performances of emergency medical service responses for suspected, nonfatal, opioid-involved overdoses, prescription drug monitoring programs, and carceral release data align well with established knowledge. For example, recent release from incarceration is strongly associated with increased risk of fatal overdose in Rhode Island, at both the individual and neighborhood levels (Brinkley-Rubinstein et al., 2018; Cartus et al., 2023). However, unlike many other jurisdictions, carceral release data in Rhode Island comes from one agency as the state has a single, combined jail and prison system and no federal prisons. While carceral release data may improve predictive ability efficiently (i.e., the RIDOC dataset consisted of only three variables), barriers to comparable data access from carceral institutions may inhibit the portability of this approach beyond Rhode Island.

The first wave of the North American overdose crisis was characterized by prescription opioid involved deaths, with PDMPs established in response to prescription opioid misuse (Smith et al., 2023). Recent trends in overdose mortality have been driven by illicitly manufactured synthetic opioids, primarily fentanyl, and design and utilization heterogeneity among PDMPs may hinder PDMP data portability beyond Rhode Island (Ciccarone, 2021). Prescription opioid related variables were less included in and less important for models considering PDMP data compared to buprenorphine related variables. Of the 39 times PDMP variables were included in models in this analysis, 22 of these instances involved a buprenorphine-related variable, which in linear models were always positively associated with future fatal overdose. An over-reliance on PDMP data may skew models to prioritize areas with existing treatment services, such as buprenorphine access.

Despite less stable window to window performance, emergency medical services data may be more natural to integrate into overdose forecasting efforts by local stakeholders. Emergency medical services data for responses to suspected overdoses have been used previously to identify geographic units to target overdose prevention resources in eastern United States (Dworkis et al., 2018; Pesarsick et al., 2019), and non-fatal overdose is a known predictor of fatal overdose at the individual level (Guo et al., 2021). A prediction approach that incorporates EMS responses for nonfatal opioid-involved overdose data likely benefits dually from improved prediction performance and more direct prioritization of areas to target preventative care where the wider health system is already responding. While in the window one top 15th and 20th percentiles scenarios ACS + EMS models did not achieve prioritization goals, the univariate ranking by Total EMS Responses still performed well, indicating the utility of EMS data in the face of poor individual model performance. A nationally standardized case definition for an EMS response for non-fatal overdose might improve the generalizability of these data regarding their predictive performance in other settings (Hallowell et al., 2021). The case definition for nonfatal opioid-involved overdose used in this analysis could form such a basis, or could be broadened to capture nonfatal stimulant-involved overdoses (Hallowell et al., 2021).

The stronger predictive performance of the linear regression models than the more complex random forest models was surprising given the size and assumed complexity of possible predictor interactions. However, traditional regressions outperforming machine learning methods is not without precedent (Christodoulou et al., 2019). The random forests may have suffered from their inability to extrapolate as statewide fatal overdose counts increased longitudinally from training to test set periods, and predictor relationships may lack significant complexity (Hastie et al., 2009). The strong performance of the univariate rankings (though typically underperforming the multivariable models) supports this lack of significant complexity, and the linear models may have found a sweet spot in terms of complexity.

While previous work has explored the utility of using predictive analytics to address the overdose crisis, the majority of these modelling efforts have focused on targeting clinical interventions based on individual-level overdose risk (Bharat et al., 2021). Efforts to predict fatal overdoses at the neighborhood level have used a variety of sophisticated methods (Schell et al., 2022; Allen et al., 2024; Lo-Ciganic et al., 2019; Yedinak et al., 2021; Bozorgi et al., 2021), but often without specific regard for accessibility and interpretability for non-technical stakeholders in whose hands the models may make the most difference. Machine learning approaches to guide community response to the overdose crisis rely on engagement with public health professionals and community stakeholders for effective implementation (Yedinak et al., 2021). Our aim is to establish simple-to-implement methods evaluated across different data sources as a template for the uptake of predictive analytics to inform preventative outreach by overdose crisis stakeholders who may have less technical expertise or have access to data from fewer domains.

This analysis is subject to several limitations. The high quality and standardized nature of data used in this analysis, due to diligent management by Rhode Island's centralized state health department, may not be reproducible in other jurisdictions with more fragmented public health authorities. All data used to train and evaluate this study's models either precede or are concurrent with the COVID-19 pandemic, limiting generalizability to future time periods. Data sources' predictive performances may not be generalizable to other jurisdictions as associations making certain variables predictive for overdose in Rhode Island may be different in other regions. Practitioners working in new jurisdictions or outside the United States should consider incorporation of further data sources that fit conceptually as predictive of fatal overdose. The application of these methods in other jurisdictions is a crucial next step in our research to evaluate this approach. Using only fatal overdose as the outcome variable may not capture the greater burden of nonfatal overdoses and substance use disorders, which preventative programs should also target. Some model predictors may reflect areas already with services (e.g. count of freestanding ambulatory clinics or total buprenorphine prescriptions), introducing a bias in model predictions. Mitigation of the overdose crisis at the neighborhood level should be complemented by similar endeavors at the individual level. All predictive models that inform the delivery of healthcare will only ever be as effective as the care they facilitate. Further work is required to address the upstream causes of overdose mortality (Dasgupta et al., 2018).

5. Conclusions

This analysis presents a template for straightforward modelling approaches to forecast fatal overdoses at the neighborhood level with data from administrative sources commonly used as part of overdose surveillance. A screened linear regression model trained on only either ACS and emergency medical service response to non-fatal overdoses, ACS and carceral release data, or ACS and prescription drug monitoring program data performed comparably on average to the same model trained on six diverse administrative data sources. This template has potential to enable jurisdictions in the United States and beyond to predict overdose at the neighborhood level and coordinate deployment of overdose prevention services to where they are needed most.

CRediT authorship contribution statement

John C. Halifax: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Bennett Allen:** Writing – review & editing, Supervision, Conceptualization. **Claire Pratty:** Writing – review & editing, Data curation. **Victoria Jent:** Writing – review & editing, Validation. **Alexandra Skinner:** Writing – review & editing. **Magdalena Cerdá:** Writing – review & editing, Funding acquisition. **Brandon D.L. Marshall:** Writing – review & editing, Funding acquisition. **Daniel B. Neill:** Writing – review & editing, Supervision. **Jennifer Ahern:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing interests that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Dr. Benjamin Hallowell, Maxwell Krieger, and Dr. William Goedel for their help in this endeavor. We would like to thank Dr. Robert Schell for his help early in this endeavor. This study is funded by the National Institute on Drug Abuse (NIDA; project reference R01-DA046620) and by the Centers for Disease Control and Prevention (CDC; project reference K01CE003586).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ypmed.2025.108276.

Data availability

The authors do not have permission to share the original data used in the full analysis, but a full code template with simulated data representative of the data used is included in Supp Materials

References

Adeniran, E., Quinn, M., Wallace, R., et al., 2023. A scoping review of barriers and facilitators to the integration of substance use treatment services into US mainstream health care. Drug Alcohol Depend Rep. 7, 100152. https://doi.org/10.1016/j.dadr.2023.100152.

Allen, B., Neill, D.B., Schell, R.C., et al., 2023. Translating predictive analytics for public health practice: a case study of overdose prevention in Rhode Island. Am J Epidemiol., kwad119 https://doi.org/10.1093/aje/kwad119.

Allen, B., Schell, R.C., Jent, V.A., et al., 2024. PROVIDENT: development and validation of a machine learning model to predict neighborhood-level overdose risk in Rhode Island. Epidemiology 2. https://doi.org/10.1097/EDE.0000000000001695.

Bharat, C., Hickman, M., Barbieri, S., Degenhardt, L., 2021. Big data and predictive modelling for the opioid crisis: existing research and future potential. Lancet Digit Health. 3 (6), e397–e407. https://doi.org/10.1016/S2589-7500(21)00058-3.

Bozorgi, P., Porter, D.E., Eberth, J.M., Eidson, J.P., Karami, A., 2021. The leading neighborhood-level predictors of drug overdose: a mixed machine learning and spatial approach. Drug Alcohol Depend. 229, 109143. https://doi.org/10.1016/j.drugalcdep.2021.109143.

Brinkley-Rubinstein, L., Macmadu, A., Marshall, B.D.L., et al., 2018. Risk of fentanyl-involved overdose among those with past year incarceration: findings from a recent outbreak in 2014 and 2015. Drug Alcohol Depend. 185, 189–191. https://doi.org/10.1016/j.drugalcdep.2017.12.014.

Bureau, U.C., 2025. American Community Survey (ACS). Census.gov. Accessed September 20, 2023. https://www.census.gov/programs-surveys/acs.

Buresh, M., Stern, R., Rastegar, D., 2021. Treatment of opioid use disorder in primary care. BMJ 373, n784. https://doi.org/10.1136/bmj.n784.

Cartus, A.R., Goedel, W.C., Jent, V.A., et al., 2023. Neighborhood-level association between release from incarceration and fatal overdose, Rhode Island, 2016–2020. Drug Alcohol Depend. 247, 109867. https://doi.org/10.1016/j.drugalcdep.2023.109867.

Chan, B., Hoffman, K.A., Bougatsos, C., Grusing, S., Chou, R., McCarty, D., 2021. Mobile methadone medication units: a brief history, scoping review and research opportunity. J. Subst. Abuse Treat. 129, 108483. https://doi.org/10.1016/j.jsat.2021.108483.

Christodoulou, E., Ma, J., Collins, G.S., Steyerberg, E.W., Verbakel, J.Y., Van Calster, B., 2019. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J. Clin. Epidemiol. 110, 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004.

Ciccarone, D., 2021. The rise of illicit fentanyls, stimulants and the fourth wave of the opioid overdose crisis. Curr. Opin. Psychiatry 34 (4), 344–350. https://doi.org/10.1097/YCO.0000000000000717.

Dasgupta, N., Beletsky, L., Ciccarone, D., 2018. Opioid crisis: no easy fix to its social and economic determinants. Am. J. Public Health 108 (2), 182–186. https://doi.org/10.2105/AJPH.2017.304187.

Dworkis, D., Weiner, S., Liao, V., Rabickow, D., Goldberg, S., 2018. Geospatial clustering of opioid-related emergency medical services runs for public deployment of naloxone. West. J. Emerg. Med. 19 (4), 641–648. https://doi.org/10.5811/westjem.2018.4.37054.

Grömping, U., 2015. Variable importance in regression models. WIREs Comput Stat. 7 (2), 137–152. https://doi.org/10.1002/wics.1346.

Guo, J., Lo-Ciganic, W.H., Yang, Q., et al., 2021. Predicting mortality risk after a hospital or emergency department visit for nonfatal opioid overdose. J. Gen. Intern. Med. 36 (4), 908–915. https://doi.org/10.1007/s11606-020-06405-w.

Hallowell, B.D., Chambers, L.C., Rhodes, J., Basta, M., Viner-Brown, S., Lasher, L., 2021. Using emergency medical services data to monitor nonfatal opioid overdoses in real time. Public Health Rep. 136 (1 Suppl), 40S–46S. https://doi.org/10.1177/00333549211018989.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer, New York. https://doi.org/10.1007/978-0-387-84858-7.

Hill, K., Kuo, I., Shenoi, S.V., Desruisseaux, M.S., Springer, S.A., 2023. Integrated care models: HIV and substance use. Curr HIV/AIDS Rep. 12. https://doi.org/10.1007/s11904-023-00667-9.

Jiang, Y, McDonald, J, Goldschmidt, A, et al., 2018. State Unintentional Drug Overdose Reporting Surveillance: Opioid Overdose Deaths and Characteristics in Rhode Island. R I Med J. 101 (7), 25–30.

Lo-Ciganic, W.H., Huang, J.L., Zhang, H.H., et al., 2019. Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. JAMA Netw. Open 2 (3), e190968. https://doi.org/10.1001/jamanetworkopen.2019.0968.

Manson, S., Schroeder, J., Van Riper, D., Kugler, T., Ruggles, S., 2021. National Historical Geographic Information System: Version 16.0. https://doi.org/10.18128/D050.V16.0.

Marshall, B.D.L., Alexander-Scott, N., Yedinak, J.L., et al., 2022. Preventing overdose using information and data from the environment (PROVIDENT): protocol for a randomized, population-based, community intervention trial. Addiction 117 (4), 1152–1162. https://doi.org/10.1111/add.15731.

Mhasawade, V., Zhao, Y., Chunara, R., 2021. Machine learning and algorithmic fairness in public and population health. Nat Mach Intell. 3 (8), 659–666. https://doi.org/10.1038/s42256-021-00373-4.

Neill, D.B., Herlands, W., 2018. Machine learning for drug overdose surveillance. J. Technol. Hum. Serv. 36 (1), 8–14. https://doi.org/10.1080/15228835.2017.1416511.

Pepin, M.D., Joseph, J.K., Chapman, B.P., et al., 2023. A mobile addiction service for community-based overdose prevention. Front Public Health 11. Accessed September 29, 2023. https://doi.org/10.3389/fpubh.2023.1154813.

Pesarsick, J., Gwilliam, M., Adeniran, O., Rudisill, T., Smith, G., Hendricks, B., 2019. Identifying high-risk areas for nonfatal opioid overdose: a spatial case-control study using EMS run data. Ann. Epidemiol. 36, 20–25. https://doi.org/10.1016/j.annepidem.2019.07.001.

R Core Team, 2021. R: The R Project for Statistical Computing. Accessed March 14, 2025. https://www.r-project.org/.

Rhode Island Department of Business Regulation, 2025. Commercial Licensing | Dept. of Business Regulation. Accessed September 20, 2023. https://dbr.ri.gov/real-estate-and-commercial-licensing.

Rhode Island Department of Corrections. Department of Corrections. 2023. https://doc.ri.gov/.

Rhode Island Department of Health, 2022. Prescription Drug Monitoring Program: Department of Health. Accessed December 8, 2023. https://health.ri.gov/healthcare/medicine/about/prescriptiondrugmonitoringprogram/.

Rhode Island Department of Health, 2025. Healthcare Licensing : Department of Health. Accessed September 20, 2023. https://health.ri.gov/licenses/.

Rhode Island Geographic Information System, 2025. Rhode Island Geographic Information System (RIGIS). Accessed September 20, 2023. https://www.rigis.org/.

Robinowitz, N., Smith, M.E., Serio-Chapman, C., Chaulk, P., Johnson, K.E., 2014. Wounds on wheels: implementing a specialized wound clinic within an established syringe exchange program in Baltimore, Maryland. Am J Public Health. 104 (11), 2057–2059. https://doi.org/10.2105/AJPH.2014.302111.

Rosecrans, A., Harris, R., Saxton, R.E., et al., 2022. Mobile low-threshold buprenorphine integrated with infectious disease services. J. Subst. Abuse Treat. 133, 108553. https://doi.org/10.1016/j.jsat.2021.108553.

Roux, A.V.D., Merkin, S.S., Arnett, D., et al., 2001. Neighborhood of residence and incidence of coronary heart disease. N. Engl. J. Med. 345 (2), 99–106. https://doi.org/10.1056/NEJM200107123450205.

Saloner, B., McGinty, E.E., Beletsky, L., et al., 2018. A public health strategy for the opioid crisis. Public Health Rep 133 (1S), 24S–34S.

Schell, R.C., Allen, B., Goedel, W.C., et al., 2022. Identifying predictors of opioid overdose death at a neighborhood level with machine learning. Am. J. Epidemiol. 191 (3), 526–533. https://doi.org/10.1093/aje/kwab279.

Smith, N.Z.Y., Thornton, J.D., Fenton, S.H., Simmons, D., Champagne-Langabeer, T., 2023. Helpful, unnecessary, or harmful: a systematic review of the effects of prescription drug monitoring program use on opioid prescriptions. Pharmacoepidemiology 2 (4), 350–365. https://doi.org/10.3390/pharma2040030.

Spencer, M.R., Garnett, M.F., Miniño, A.M., 2024. Drug Overdose Deaths in the United States, 2002–2022. NCHS Data Brief No 491 Natl Cent Health Stat, p. 491.

Springer, S.A., 2023. Ending the HIV epidemic for persons who use drugs: the practical challenges of meeting people where they are. J. Gen. Intern. Med. 38 (12), 2816–2818. https://doi.org/10.1007/s11606-023-08142-2.

Substance Abuse and Mental Health Services Administration, 2025. Treatment Locator Map. FindTreatment.gov. Accessed September 20, 2023. https://findtreatment.gov/.

Vohra, N., Paudyal, V., Price, M.J., 2022. Homelessness and the use of emergency department as a source of healthcare: a systematic review. Int. J. Emerg. Med. 15 (1), 32. https://doi.org/10.1186/s12245-022-00435-3.

Wilt, G.E., Lewis, B.E., Adams, E.E., 2019. A spatial exploration of changes in drug overdose mortality in the United States, 2000–2016. Prev. Chronic Dis. 16, E33. https://doi.org/10.5888/pcd16.180405.

Yedinak, J.L., Li, Y., Krieger, M.S., et al., 2021. Machine learning takes a village: assessing neighbourhood-level vulnerability for an overdose and infectious disease outbreak. Int. J. Drug Policy 96, 103395. https://doi.org/10.1016/j.drugpo.2021.103395.

Yedinak, J., Krieger, M.S., Joseph, R., et al., 2024. Public health dashboards in overdose prevention: the Rhode Island approach to public health data literacy, partnerships, and action. J. Med. Internet Res. 26, e51671. https://doi.org/10.2196/51671.