A Spatio-temporal Cluster-aware Supervised Learning Framework for Predicting County-level Drug Overdose Deaths

Zixuan Feng[1]*, Qing Ye[2]*, Weijun Xie[2], Qiushi Chen[1]

[1] Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University
[2] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology
zixuan.feng@psu.edu, qye40@gatech.edu, wxie@gatech.edu, q.chen@psu.edu

Abstract

The soaring drug overdose crisis in the United States has claimed more than half a million lives in the past decade and remains a major public health threat. The ability to predict drug overdose deaths at the county level can help local communities develop action plans in response to emerging changes. Applying off-the-shelf machine learning algorithms for prediction can be challenging due to the heterogeneous risk profiles of the counties and suppressed data in common publicly available data sources. To fill these gaps, we develop a cluster-aware supervised learning (CASL) framework to enhance the prediction of county-level drug overdose deaths. This CASL model simultaneously clusters counties into groups based on geographical and socioeconomic characteristics and minimizes the loss function that accounts for suppressed values and cluster-specific regularization. Our computational study uses real-world data from 2010 to 2021, focusing on the ten states most severely impacted by the drug overdose crisis. The results demonstrate that our proposed CASL framework significantly outperforms state-of-the-art methods by achieving a superior balance in prediction accuracy for both unsuppressed and suppressed observations. The proposed model also identifies different clusters of counties, capturing heterogeneous patterns of overdose mortality among counties of diverse characteristics.

Code and technical appendix —
https://github.com/zixuan-feng/CASL-DrugOverdoseDeathsPrediction-2024

Introduction

The drug overdose crisis continues to be a critical public health issue in the United States (US) (Duff et al. 2022; Centers for Disease Control and Prevention 2024b). More than 100,000 Americans died from drug overdoses in 2021, a 30% increase from the previous year (Centers for Disease Control and Prevention 2024b). This crisis has evolved through multiple waves, initially driven by prescription opioids, followed by heroin, then synthetic opioids such as fentanyl (Ciccarone 2019), and now a fourth wave characterized by polysubstance use involving opioids and stimulants such as methamphetamine and cocaine (Ciccarone 2021). Beyond the devastating mortality burden on society, the drug overdose crisis has also notoriously impacted healthcare, criminal justice, and productivity (Florence, Luo, and Rice 2021). A congressional report estimated that the total economic toll of addiction and drug overdose crisis reached nearly $1.5 trillion in 2020 alone and is expected to increase (Aboulenein 2022).

The ability to predict drug overdose deaths can inform public health policymakers and communities of emerging threats as early warning signals to help prepare for response strategies (Blanco et al. 2020; Jalali et al. 2020b). By tracking the historical trend and predicting overdose deaths in the near future, policymakers can make *data-driven* decisions to adjust existing intervention strategies or reallocate resources to priority areas to more effectively mitigate the harms of the overdose crisis. In the public health domain, predictive analytics has been widely used to support responses to emerging epidemics. For example, previous research has used prediction models to predict the trend and understand the periodicity of the COVID-19 pandemic (Wang et al. 2022; Aslam and Biswas 2023). Santangelo et al. (2023) provides a systematic review of machine learning prediction models for a wide range of infectious diseases (e.g. influenza and malaria).

Previous studies on predicting drug overdose deaths have utilized various analytical approaches. Simple regression-based methods, including generalized linear models (e.g., logistic or negative binomial regression) (Cuomo et al. 2023; Marks et al. 2021) and lasso regression (Sumner et al. 2022), have been widely employed to predict the number of overdose deaths based on demographic and socioeconomic covariates at county or state levels. Recent studies have utilized more advanced machine learning models such as random forest (Schell et al. 2022) and extreme gradient boosting (Tatar et al. 2023) for drug overdose death predictions to capture complex, nonlinear interactions between county or neighborhood features. Several studies have applied machine learning models to internet search trend data (e.g., Google Trends) of drug-related search terms for spatio-temporal mapping of drug overdose deaths, near real-time prediction, and timely detection of emerging hotspots (Campo et al. 2020; Ghosh et al. 2022; Mukherjee et al. 2020). In addition to the population-level prediction, there

*Zixuan Feng and Qing Ye contributed equally to this work and share first authorship.

is a stream of research on predicting overdose mortality outcomes at the individual level using clinical and administrative data (Bharat et al. 2021; Lo-Ciganic et al. 2019; Saloner et al. 2020).

However, there are several unique challenges in predicting drug overdose deaths that have not been adequately addressed by previous studies. First and foremost, it is crucial to recognize the substantial heterogeneity in the overdose outcomes between geographical areas (e.g., counties) driven by diverse local contexts. This makes conventional prediction models—that typically treat the training data as being drawn from a homogeneous population—potentially inadequate and unreliable (Jalali et al. 2020a). Moreover, the data are not necessarily independent between geographical areas. They can be geographically correlated due to spatial proximity and shared resources, or socially correlated due to similar demographic and socioeconomic characteristics. In other words, the underlying structure among the data points can potentially contribute to additional insights for prediction, which remains underexplored and thus warrants novel methods to directly capture such correlations.

Another challenge lies in the use of publicly available data. Despite their desirability due to greater accessibility by communities and greater transparency for sharing analysis results, publicly available data related to drug overdoses, such as overdose deaths, may suffer from data suppression, which means that any value below a certain threshold is "masked" to ensure non-identifiability of the data. Data suppression is not uncommon, especially when the outcome of interest has low counts in small geographic areas. The annual number of drug overdose deaths obtained from the CDC's Wide-ranging ONline Data for Epidemiologic Research (WONDER) database, is suppressed in more than half of US counties. Simply excluding suppressed data or treating them as missing values with imputation as a data preprocessing step will not be an effective approach. Restricted datasets can provide full details, but for practitioners from local communities, acquiring access to such data is not straightforward. Integrating multiple siloed datasets across agencies, even within the same state, can be challenging (Chen et al. 2024). The practical barriers to more detailed data with restricted access underscore the need for methods that can more effectively handle suppressed data to make better use of publicly available sources.

To tackle these challenges, we develop an interpretable cluster-aware supervised learning (CASL) framework to predict annual drug overdose deaths at the county level. By simultaneously clustering data into subgroups and making predictions, the CASL framework balances both prediction errors and clustering quality to capture underlying heterogeneity in the data, which is more desirable than other strategies of combining regression and clustering either as two separate steps sequentially (Wang, Ning, and Kong 2019; Tran et al. 2018) or as one integrated step but focusing on minimizing the regression error only (Zhu, Li, and Kong 2012; Devijver 2017). Our CASL framework extends the previous work of Chen and Xie (2022) for only *cross-sectional* data to handle *time-series* data while maintaining time consistency of clusters and to consider *suppressed values* in the data. In this study, we apply the CASL model to real-world data for 2010-2021 collected from 10 states that are most impacted by the overdose crisis in the US and show that it outperforms the random forest and lasso regression, the commonly used models for similar public health application settings.

Model Formulation

In this section, we introduce the time series model for the overdose death prediction problem, define the loss function that accounts for suppressed observations in the data, and describe the proposed cluster-aware supervised learning (CASL) model that simultaneously divides counties into groups and takes advantage of the cluster structures of counties to predict overdose deaths.

Autoregressive Model for Time Series Overdose Death Data

We use an autoregressive model as the statistical model to predict drug overdose deaths at the county level. Consider a set of I counties indexed by $i \in [I]$. We denote $[n]$ as the set $\{1, 2, \cdots, n\}$ for any given integer $n \in \mathbb{N}$. Let $y_{i,t}$ represent the total number of drug overdose deaths in county i at time period t as the primary outcome variable for prediction. In our application setting, we consider one year as one time period given that most of the data we collect are available at the annual level; the time period can be changed to shorter intervals if more frequent data were available. In addition, we define $\boldsymbol{u}_{i,t}$ as a vector of time-varying predictors for county i at time t (such as the number of drug-related crime incidents, interaction terms involving at least one time-varying predictor) and $\boldsymbol{z}_i$ as a vector of time-invariant predictors for county i (such as a county's socioeconomic factors that typically remain stable over time). To predict the number of overdose deaths of the next time period $t+1$, we assume the following autoregressive model with lagged terms of at most $L-1$ time periods:

$$y_{i,t+1} = \gamma + \sum_{\ell \in [L]} \left(\alpha_\ell y_{i,t-\ell+1} + \boldsymbol{\theta}_\ell^\top \boldsymbol{u}_{i,t-\ell+1}\right) + \boldsymbol{\eta}^\top \boldsymbol{z}_i + \epsilon_{i,t}, \tag{1}$$

where the residual term $\epsilon_{i,t}$ captures the random error in each observation. In model (1), the coefficients $\boldsymbol{\beta} = (\gamma; \boldsymbol{\alpha}; \boldsymbol{\theta}_1; \cdots; \boldsymbol{\theta}_L; \boldsymbol{\eta})$ are the model parameters that capture the predictive relationships between predictors and outcomes and to be learned from the data. Specifically, the intercept γ represents the baseline level of predicted overdose deaths, vector $\boldsymbol{\alpha}$ represents the effect of historical overdose deaths in the current and past $L-1$ periods to capture the autocorrelations for predicting future values, vectors $\{\boldsymbol{\theta}_\ell\}_{\ell \in [L]}$ represent the effects of additional drug-related factors other than the overdose death outcomes in the current and past $L-1$ periods, and vector $\boldsymbol{\eta}$ represents the influence associated with the time-invariant characteristics of the counties. For notational convenience, we let $\boldsymbol{X}_{i,t} = (1; \boldsymbol{y}_{i,t:(t-L+1)}; \boldsymbol{u}_{i,t}; \cdots; \boldsymbol{u}_{i,t-L+1}; \boldsymbol{z}_i)$ be all predictor variables for predicting the overdose death in county i at period $t+1$. Then the autoregressive model (1) can be

represented in the following compact form:

$$y_{i,t+1} = \boldsymbol{\beta}^\top \boldsymbol{X}_{i,t} + \epsilon_{i,t}. \tag{2}$$

Loss Function Considering Data Suppression

Data suppression is common in public health data reporting for protecting privacy and preventing data identifiability in publicly available data sources. Typically, a count below a certain threshold is marked as suppressed, and this threshold is known in advance. For example, for all queries of the number of overdose deaths from the Centers for Disease Control and Prevention (CDC) Wide-ranging ONline Data for Epidemiologic Research (WONDER) database (Centers for Disease Control and Prevention 2024a), any cell value $\leq$9 is marked as suppressed. Such data suppression poses challenges to applying off-shelf machine learning models for predictions because the exact true values that are needed for training the model are unknown. In fact, data suppression is not a trivial issue and cannot be overlooked in the context of predicting drug overdose deaths, as we have observed that more than half of the US counties have suppressed values in the data. Moreover, treating suppressed data as missing values may not be effective because data suppression is more informative than completely missing the data in the way that it provides an upper bound for the actual value and is spatial-temporal correlated.

To capture this, we define a new loss function to account for both unsuppressed and suppressed observations. In particular, we let constant c denote the prespecified threshold for data suppression and $\mathcal{S}$ (or $\bar{\mathcal{S}}$) represent the suppressed (or unsuppressed) dataset, i.e., the set of (i,t) pairs with $y_{i,t} \leq c$ being suppressed (or with $y_{i,t} > c$ being unsuppressed). Then we define the loss function for predicting $y_{i,t+1}$ with given parameters $\boldsymbol{\beta}$ of model (2) as follows:

$$L(\boldsymbol{\beta}; (y_{i,t+1}, \boldsymbol{X}_{i,t})) = \begin{cases} \frac{1}{|\bar{\mathcal{S}}|} \cdot \left| y_{i,t+1} - \boldsymbol{\beta}^\top \boldsymbol{X}_{i,t} \right|, & \text{if } (i,t) \in \bar{\mathcal{S}}, \\ \kappa \cdot \frac{1}{|\mathcal{S}|} \cdot \left(\boldsymbol{\beta}^\top \boldsymbol{X}_{t,i} - c \right)_+, & \text{if } (i,t) \in \mathcal{S}. \end{cases}$$

The loss function $L(\boldsymbol{\beta}; (y_{i,t+1}, \boldsymbol{X}_{i,t}))$ has two components: a mean absolute error for unsuppressed data ($\bar{\mathcal{S}}$) and a hinge loss function (i.e., $(\cdot)_+ = \max(\cdot, 0)$) for suppressed data ($\mathcal{S}$) scaled by parameter κ. We choose to use the more outlier-robust mean absolute error, instead of squared error, for unsuppressed data to maintain consistency with the piecewise linearity of the hinge loss function for suppressed data and avoid overpenalizing outliers. This way, our loss function penalizes both errors of overpredicting the actual suppressed data to be above the suppression threshold c and the distance between the predicted value and the ground truth for unsuppressed data. We assume all counties follow the same suppression rule with a single threshold c, since our data for $\{\boldsymbol{y}_i\}_{i\in[I]}$ used in this study are from the same source. If the data are collected from multiple sources with different suppression rules across states or counties, it is straightforward to modify the above loss function by specifying the suppression threshold as county-specific values of c_i.

Cluster-aware Supervised Learning (CASL) Framework

One unique challenge for our overdose death prediction problem is the heterogeneity and correlation of observations at the county level. These counties have very diverse characteristics in terms of demographic, social, and economic factors (Langabeer et al. 2022). Therefore, the dataset used for model training should not simply be viewed as identically distributed observations from a homogeneous "population." On the other hand, the observations from the various counties are not completely independent and can be intercorrelated. For example, counties may share similar risks or burdens from the opioid crisis due to geographical proximity, similar characteristics of social and economic determinants, or shared healthcare and public health infrastructures.

The complex interplay of heterogeneity and correlations in the observations from the counties limits the direct applicability of most existing machine learning models, calling for a new approach to explicitly capture the *underlying structures* in the observed data. One learning strategy is to identify the underlying subgroups of counties that exhibit similar patterns of overdose mortality and associated driving factors, while simultaneously providing predictions. The work (Chen and Xie 2022) presents a learning framework that integrates clustering with supervised learning, which offers promises for our overdose prediction problem. However, their model is developed for *static* prediction problems without explicitly considering temporal dynamics and time series data, which, in contrast, are critical to our problem setting. This can complicate the prediction solution, as the level of clustering may not align with the granularity of the data.

To address this gap, we extend the framework in Chen and Xie (2022) to time series data at the county-year level and formulate the clustering at the county level to maintain the time consistency of the clusters. For time series data, the layout of the training data is illustrated in Figure 1.

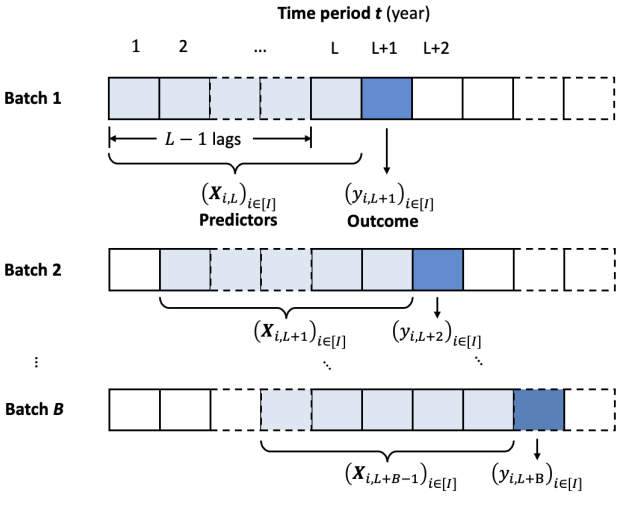

Figure 1: Schematics of the data layout for training data that consist of B batches.

As shown in Figure 1, the training dataset will include B batches of data, namely, $(\boldsymbol{X}_{i,L}, y_{i,L+1})_{i\in[I]}$, $\cdots$, $(\boldsymbol{X}_{i,L+B-1}, y_{i,L+B})_{i\in[I]}$, among which each batch includes predictors based on the data in the current period and also the historical data with lags up to $(L-1)$ periods. With such a data layout, one county can have B observations in the training data.

We consider county-level clusters of the data points for ease of interpretation; in contrast, a county-year-level cluster may consist of counties from different years, making such clusters difficult to interpret for their policy implications. In other words, all observations of different years from the same county are more meaningful if they belong to the same cluster. This model setup assumes that the variations within a county over multiple years are less significant than those between counties. Our numerical experiments validate this assumption.

Next, we formulate the cluster-aware supervised learning (CASL) model as the following optimization problem. Consider K clusters in total, where K can be tuned through cross-validation. We define the clustering decision variable $\delta_{k,i} \in \{0,1\}$, representing whether the data points $(y_{i,t}, \boldsymbol{X}_{i,t})$ of county i for all different time t belong to the cluster k. We extend the autoregressive model (2) by having cluster-specific model parameters $\boldsymbol{\beta}_k$ for $k \in [K]$. We use $\boldsymbol{m}_k$ to denote the centroid of the cluster k (defined in the space of all predictors). Then, the CASL model is given as:

$$\min_{\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{m}} R(\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{m}) :=$$

$$\sum_{k\in[K]}\left(\sum_{b\in[B]}\sum_{i\in[I]}\delta_{k,i}L(\boldsymbol{\beta}_k;\boldsymbol{X}_{i,L+b-1})+\lambda_k\|\boldsymbol{\beta}_k\|_1\right)$$

$$+\rho\sum_{k\in[K]}\sum_{b\in[B]}\sum_{i\in[I]}\delta_{k,i}D\left(\boldsymbol{m}_k;\boldsymbol{X}_{i,L+b-1}\right), \tag{3a}$$

$$\text{s.t.} \quad \sum_{k\in[K]}\delta_{k,i}=1, \quad \forall i\in[I], \tag{3b}$$

$$\sum_{i\in[I]}\delta_{k,i}\geq\mu I, \quad \forall k\in[K], \tag{3c}$$

$$\delta_{k,i}\in\{0,1\}, \quad \forall i\in[I], k\in[K]. \tag{3d}$$

In this optimization model formulation, the objective function (3a) is to minimize both the total losses with regularization penalties from all clusters for the supervised learning and the total dissimilarity between the data points of selected counties and the "virtual" centroid of the cluster for the clustering, modulated by a nonnegative weight ρ for the clustering performance metric. The dissimilarity function $D(m_k;\boldsymbol{X}_{i,t})$ is defined by the squared Euclidean distance between the data point $\boldsymbol{X}_{i,t}$ and the cluster centroid $\boldsymbol{m}_k$. That is, we assume that the same set of factors $\boldsymbol{X}_{i,t}$ for predicting overdose deaths are also used for measuring county similarity and developing the clusters. Constraint (3b) ensures that each county i is assigned to exactly one of the K clusters. The restriction (3c) ensures that each group contains at least a certain percentage, $\mu \in (0, 1/K]$, of the total number of counties, thus preventing the formation of groups that are too small or too large.

To solve the non-convex optimization problem of CASL, we employ the regularized alternating minimization (RAM) algorithm, which has been proven to converge to a stationary point within a finite number of iterations (Chen and Xie 2022). The RAM algorithm works as an iterative process between learning the cluster-specific parameter $\boldsymbol{\beta}_k$ in the prediction model (i.e., the parameter update step) and improving the assignment of counties to clusters (i.e., the cluster reassignment step). Specifically, initialized with any given cluster assignment $\boldsymbol{\delta}^0$ and iteration $\tau = 1$, the RAM algorithm begins iterating through the following steps:

- **Step 1 (Parameter update)**: At iteration τ, with the given cluster assignment $\boldsymbol{\delta} = \boldsymbol{\delta}^{\tau-1}$, update the centroid $\boldsymbol{m}_k^\tau$ and optimize the model parameter $\boldsymbol{\beta}_k^\tau$ for each cluster k. This step is supervised learning within each cluster $k \in [K]$ to optimize $\boldsymbol{\beta}_k$ by minimizing the regularized loss function.
- **Step 2 (Cluster reassignment)**: With the given cluster-specific model parameters $\boldsymbol{\beta}_k = \boldsymbol{\beta}_k^\tau$ and current cluster centroids $\boldsymbol{m}_k = \boldsymbol{m}_k^\tau$, this step reoptimizes cluster assignment decisions $\boldsymbol{\delta}^\tau$. To reallocate data points from each county to K clusters, we minimize $R(\boldsymbol{\beta},\boldsymbol{\delta},\boldsymbol{m}) + \sigma\|\boldsymbol{\delta} - \boldsymbol{\delta}^{\tau-1}\|_{1,1}$ subject to constraints (3b)-(3d). The regularization term, $\sigma\|\boldsymbol{\delta} - \boldsymbol{\delta}^{\tau-1}\|_{1,1}$, is added to the objective function to ensure a smoother transition from the previous iteration and prevent drastic changes in the clustering structure. Moreover, this regularization term can be reformulated into an affine function of $\boldsymbol{\delta}$ as $\sigma\sum_{k\in[K]}\sum_{i\in[I]}[\boldsymbol{\delta}_{k,i}^{\tau-1} + (1-2\boldsymbol{\delta}_{k,i}^{\tau-1})\boldsymbol{\delta_{k,i}}]$ to further facilitate the computation (Chen and Xie 2022).
- **Step 3 (Stopping criteria)**: If there is no change in the clustering decisions from the previous iteration, i.e., $\boldsymbol{\delta}^\tau = \boldsymbol{\delta}^{\tau-1}$, stop the algorithm and return $(\boldsymbol{\beta}^\tau, \boldsymbol{m}^\tau)$; otherwise, set $\tau \leftarrow \tau + 1$ and return to **Step 1**.

Model Selection and Cross-validation

The proposed CASL model has several hyperparameters that must be pre-specified prior to model training, including the scaling parameter κ, the number of clusters K, the cluster-specific regularization $\boldsymbol{\lambda}$, and the weight ρ for the clustering performance metric in the optimization objective function. To select these model hyperparameters, we conduct cross-validations to compare the prediction performance under different hyperparameter settings. To define the performance metric, we combine the prediction performance for unsuppressed and suppressed data. For unsuppressed data, we calculate the mean absolute error (MAE); for suppressed data, we calculate the false positive rate (FPR), the percentage of overpredicting a suppressed observation as unsuppressed, i.e., above the suppression threshold. That is, $\text{FPR} = \frac{1}{|\mathcal{S}|}\sum_{(i,t)\in\mathcal{S}}\mathbb{1}\{(\sum_k\delta_{i,k}\boldsymbol{\beta}_k)^\top\boldsymbol{X}_{i,t} > c\}$. To combine these two metrics, we define a mixed-error score (averaged by county),

$$\text{Mixed-error score} = \frac{1}{|\bar{\mathcal{S}}|+|\mathcal{S}|}\left(|\bar{\mathcal{S}}|\cdot\text{MAE} + \omega\cdot|\mathcal{S}|\cdot\text{FPR}\right),$$

where the mixing parameter ω moderates the relative importance between accurately predicting high overdose

deaths (typically in large counties) versus correctly identifying counties with limited overdose deaths (typically in small counties), which can be determined by decision-makers. Our Computational Study Section will present the results for a range of mixing parameter values.

For cross-validation, we create a k split of the time series data to train and validate the model in rolling time windows. For example, the first split takes batches 1 through B for training and batch $B+1$ for validation; the second split shifts all data by one time period forward, i.e., taking Batches 2 through $B+1$ for training and the batch $B+2$ for validation, etc., until the validation data in the k-th split reach the end of the time series data (see the illustration later in Figure 2 with $B = 2$ for the computational study). The average of the mixed-error scores across k splits is used to assess overall model performance and guide model selection.

Computational Study

In this section, we conduct a computational study using real-world data on drug overdose deaths in the US to demonstrate the performance of CASL in predicting county-level drug overdose deaths. We compare prediction performance with off-the-shelf machine learning algorithms and draw policy implications from analyzing the underlying structures of the counties uncovered by the proposed CASL framework.

Data Sources and Preparation

For a broader applicability of the proposed CASL framework, we focus on publicly available data sources and include the variables that are commonly available across states, instead of ad hoc datasets for which the availability varies substantially by state. Considering both the high drug overdose burden and sufficient data coverage throughout the study period 2010-2021, our final dataset includes 414 counties from 10 states. In the following, we briefly describe the variables included in our computational study and refer to more details in Appendix Table A1.

Prediction outcome: The prediction outcome of interest is the annual number of drug overdose deaths in each county. Following the common approach in the literature to extract overdose death estimates from the CDC WONDER database (Nam et al. 2017; Mattson et al. 2021), we identify drug overdose deaths by International Classification of Diseases Codes 10th Revision (ICD-10 codes) X40-44, X60-64, X85, and Y10-14 for the underlying cause of death and ICD-10 codes T36-50 for multiple causes of death. Any cell value $\leq$9 is suppressed. In our dataset, 54% (=2,658/4,968) of all annual overdose death numbers in all 414 counties and 12 years are suppressed, and the suppression rate ranges between 45%-70% in each state by each year across 2010-2021.

Predictor variables: A broad range of county-level predictors have been considered in the literature, while many of these factors have shown relatively mixed evidence (Cano et al. 2023; Kariisa et al. 2022). Following a similar categorization of these factors, we include the predictor variables from the following dimensions:

- **Supply-side factors:** As previous research has shown the high correlation between drug-related crime data and overdose outcomes (Chen et al. 2022; Cano et al. 2024), we include the annual number of drug-related crime incidences obtained from the National Incidence-Based Reporting System (NIBRS) as a proxy for activities in local drug markets. We also consider the percentage of fentanyl among drug seizures in each year by state based on public reports from the National Forensic Laboratory Information System (NFLIS) to capture the shifting patterns of the opioid crisis toward synthetic opioids (Rawson, Erath, and Clark 2023; Ciccarone 2021).
- **Health factors:** Among a wide range of health factors and health outcomes that are used to assess the overall county health performance by the County Health Rankings' model (University of Wisconsin Population Health Institute 2024), we select several measures that are relevant to our problem context, including adult smoking rate and adult obesity rate as the indicators for the health behaviors, annual poor mental health days and annual poor physical health days to measure the overall quality of life and the health conditions of the population, and the number of primary care physicians as an important measure for healthcare accessibility in general.
- **Socioeconomic factors:** County population, household median income, Gini index (for measuring the disparities of income distribution), high school graduation rate, and Rural-Urban Commuting Area (RUCA) codes are included to represent the counties' baseline characteristics.
- **Geospatial factors:** To capture spatial patterns, in addition to using longitudinal and latitudinal coordinates (of the centroid of the county) as a common approach for geospatial analysis (Li et al. 2011; Banerjee et al. 2022), we include indicators of whether the county is at the state border and whether a highway crosses the county, as previous studies have shown that proximity to state borders and the presence of highways are associated with increased drug trafficking and distribution (Lin et al. 2023; Thurston and Freisthler 2020), which therefore can potentially influence drug overdose death rates (Srinivasan et al. 2024). We also include state indicators as fixed effects to capture the cross-state variations for prediction.

In addition, we create several derived features to further enrich the dataset for prediction. First, we add the log-transformed overdose death numbers (assumed to be 0 if suppressed) to account for the skewed distribution of the overdose deaths and also to capture the non-linear predictive relationship between past and future overdose deaths. Second, we include the "crime gravity" measure (Marks et al. 2021), which calculates the total drug-related crime rates in all counties within a radius of 50 miles of the given county, weighted by the inverse of the squared driving distance from these counties, i.e., the gravity for county i is calculated as $\log(\sum_{j\in\mathcal{J}}$ (drug-related crime rates of county i)/(driving distance between i and $j)^2)$, where set $\mathcal{J}$ represents all counties within 50 miles of county i. Driving distances between counties are extracted via Google Map distance ma-

trix API (Google 2024). Lastly, we added time-varying interaction terms for overdose deaths (including the original and logarithmic transformation values) by the same-year drug-related crime rates and fentanyl percentages, providing the model (2) of a linear structure by design with more flexibility to capture complex dynamics.

All predictor variables, except for $\{y_{i,t+1-\ell}\}_{\ell\in[L]}$, $\{\log(y_{i,t+1-\ell})\}_{\ell\in[L]}$, longitudinal and latitudinal coordinates, RUCA code, and binary indicators, are first winsorized at the 5^{th} and 95^{th} percentiles to bound outliers and then normalized to the range of $[0, 1]$.

Experiment Settings

With our raw data collected between 2010 and 2021, we configure our autoregressive model with $L = 3$ and the training data with the batch $B = 2$. We hold out the data of the two most recent years (2020 and 2021) as our prediction years for the final evaluation of the CASL model's performance and train the model for each prediction year separately. That is, for the prediction year 2020 (2021), we first use the data of the years 2010-2019 (2011-2020) for cross-validation and selecting hyperparameters, then retrain the model with selected hyperparameters on 2015-2019 (2016-2020) using 2 batches of data, and finally evaluate the performance of predicting outcomes in 2020 (2021) (see Figure 2 for an illustration). For the CASL model, we set $c = 9$ to be consistent with the suppression rule for the overdose death data queried from the CDC WONDER database (Centers for Disease Control and Prevention 2024a), and $\mu = 10\% \times \frac{1}{K}$ to ensure each cluster is no smaller than 10% of the average size of the cluster if uniformly distributed. We consider the combination of a wide range of hyperparameters $w_{\bar{\mathcal{S}}}, K, \lambda_{1:K}$, and ρ (see the detailed list in Appendix Table A2) and select the hyperparameter setting with the highest mixed-error score under various mixing parameter $\omega \in \{0.1, 1, 10, 100\}$, respectively. We compare the performance of our proposed CASL framework with random forest regression and lasso regression. The hyperparameters for these methods (e.g. maximum number and depths of trees, L_1 regularization terms) are selected following the same cross-validation procedure as the CASL model.

Experimental Results

Prediction Performance of CASL We illustrate the performance of the CASL model using a baseline case with $K = 4$ clusters. As shown in Table 1, our CASL model achieved good prediction accuracy while maintaining consistent performance between cross-validation and testing results. For both prediction years 2020 and 2021, an increase in the mixing parameter ω leads to a higher mean absolute error (MAE) between unsuppressed counties and a lower false positive rate (FPR) among suppressed counties. This behavior is in line with the design of the mixing parameter, which is intended to emphasize the prediction accuracy in suppressed counties in the model selection. For example, in 2021 data, as ω is increased from 0.1 to 100, the testing FPR gradually drops from 0.14 to 0.03 and MAE is raised from 11.89 to 14.60.

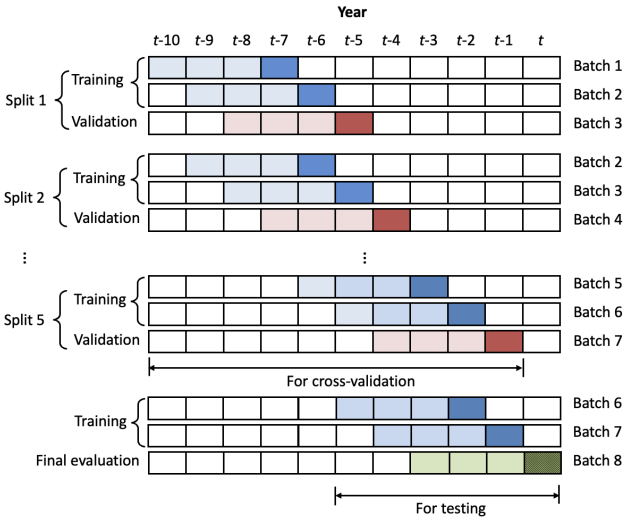

Figure 2: Schematics of the data layout for model cross-validation and testing for prediction year t (=2020 or 2021) in our computational study.

Figure 3 presents the prediction results for each county in 2020 and 2021 when $\omega = 100$. For predictions in 2020, the predicted values for most unsuppressed counties across all clusters are slightly lower than the observed values, reflecting the model's conservative nature, especially when prioritizing the prediction accuracy in suppressed counties. For prediction in 2021, the MAE for unsuppressed counties is lower than that in 2020. However, there are some counties where the predicted values for the suppressed counties exceed the suppression threshold, as shown in the figure. This leads to a slight increase in the FPR, highlighting a trade-off between improving overall prediction accuracy and maintaining strict control over predictions for suppressed data.

Effect of Number of Clusters Under different levels of the mixing parameter ω, the effect of the number of clusters K on model performance varies. As shown in Figure 4, when $\omega = 0.1$, the mixed-error score tends to increase as K increases, indicating that fewer clusters generally yield better performance. It is worth noting that setting $K = 1$ is an ablated version of CASL without clustering, which has the lowest MAE values when ω is small. This may be because the overall error is predominately driven by unsuppressed counties, while the misclassifications of suppressed counties become negligible. In this case, there is no incentive to have more clusters for differentiation. However, as ω increases to 10 and 100, the relationship becomes less straightforward. For $\omega = 10$, there is a notable drop in the mixed-error score at $K = 2$, suggesting a temporary improvement before the score increases again with more clusters. For $\omega = 100$, the trend shifts, with mixed-error scores generally decreasing as K increases, particularly in 2021, where $K = 5$ achieves the lowest score. This suggests that as the emphasis on suppressed data increases, the number of clusters may also increase to better capture the complex patterns in the data.

Prediction year	ω	cross-validation			Testing for the prediction year			Hyperparameters		
		MAE	FPR	Mixed-error score	MAE	FPR	Mixed-error score	λ	ρ	κ
2020	0.1	12.63	0.11	6.07	16.12	0.03	9.89	[0.2, 0.5, 5, 10]	0.1	1
	1	12.63	0.11	6.12	16.12	0.03	9.90	[0.2, 0.5, 5, 10]	0.1	1
	10	12.63	0.11	6.66	16.12	0.03	10.01	[0.2, 0.5, 5, 10]	0.1	1
	100	13.91	0.01	7.11	18.13	0.00	11.12	[0.5, 1, 5, 10]	0.5	10
2021	0.1	13.42	0.64	6.95	11.89	0.14	7.84	[0.1, 1, 5, 10]	0.5	0.1
	1	13.74	0.12	7.15	12.04	0.12	7.98	[0.1, 0.5, 5, 10]	0.5	1
	10	13.99	0.09	7.66	13.09	0.12	9.04	[0.1, 0.2, 5, 10]	0.2	1
	100	15.19	0.00	8.01	14.60	0.03	10.59	[0.1, 0.2, 1, 5]	1	10

Table 1: Prediction performance of CASL model with four clusters ($K = 4$).

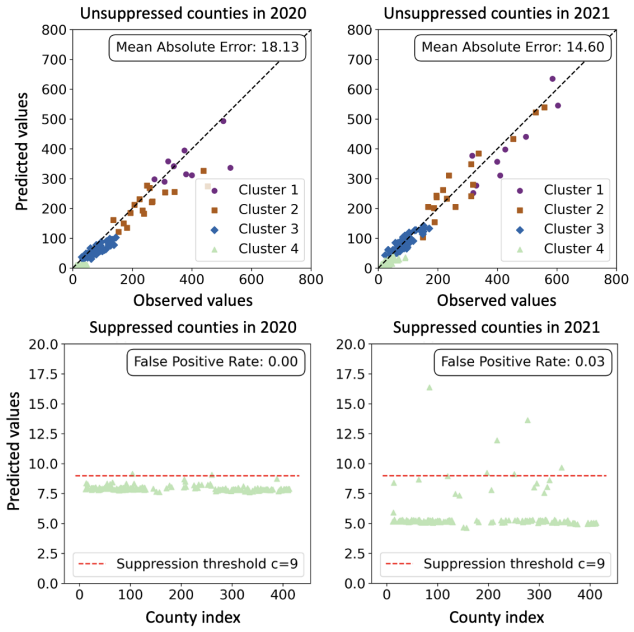

Figure 3: CASL model performance: predicted vs. observed drug overdose deaths for unsuppressed counties (top panel) and predicted outcomes compared with the suppression threshold for suppressed counties (bottom panel) in 2020 and 2021, respectively ($K = 4, \omega = 100$).

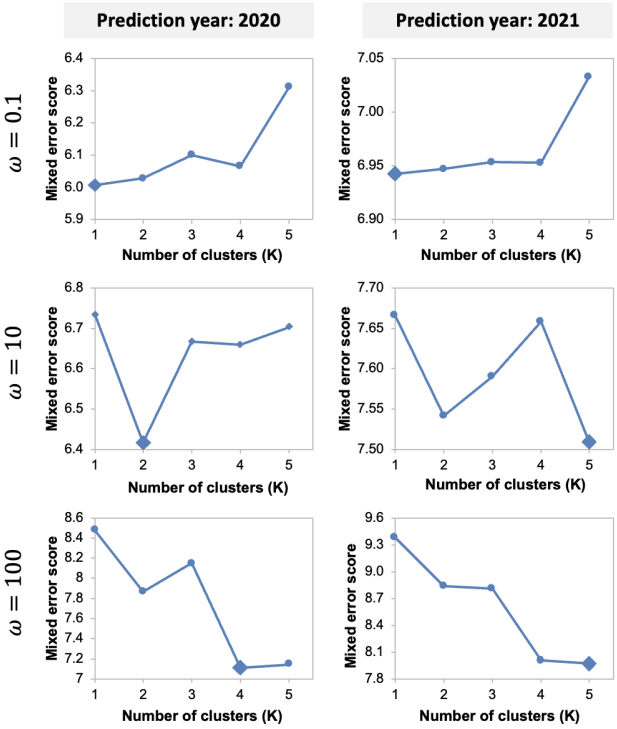

Figure 4: Comparison of CASL model performance from cross-validation with different numbers of clusters K under various values of mixing weight ω.

Comparison with Other Models We compare our CASL model with random forest and lasso regression using MAE as the loss function,[1] employing the same cross-validation strategy and feature space as our CASL model under various levels of $\omega \in \{0.1, 1, 10, 100\}$. For these comparisons, the suppressed values are imputed with 0 for training random forest and lasso regression. The hyperparameters tuned for both models can be found in Appendix Table A3. The CASL model generally outperforms or remains comparable to the random forest and lasso regression models across different levels of the mixing parameter ω, particularly when focusing on the testing results. Figure 5 presents the Pareto frontier plot for MAE and FPR for the prediction years 2020 and 2021. A more detailed summary of the comparisons can be found in Appendix Table A4. CASL models generally occupy positions closer to the lower left of the frontier, indicating a more favorable balance between MAE and FPR. For example, in 2020 data, CASL models with $K = 1$ and $\omega = 0.1$ or 1, as well as the CASL model with $K = 4$ and $\omega = 100$, capture 2 out of the 3 points on the frontier. Besides, although the CASL model with $K = 2$ and $\omega = 10$ does not appear directly on the frontier, it is comparable to the best-performing model (Lasso regression with $\omega = 100$), achieving the same level of FPR at 0.0125 with only a slightly higher MAE of 16.97 compared to 16.7.

[1] We have also tested supported vector regression, linear regression, and recurrent neural networks. Random forest and lasso regression performed best according to our numerical experience.

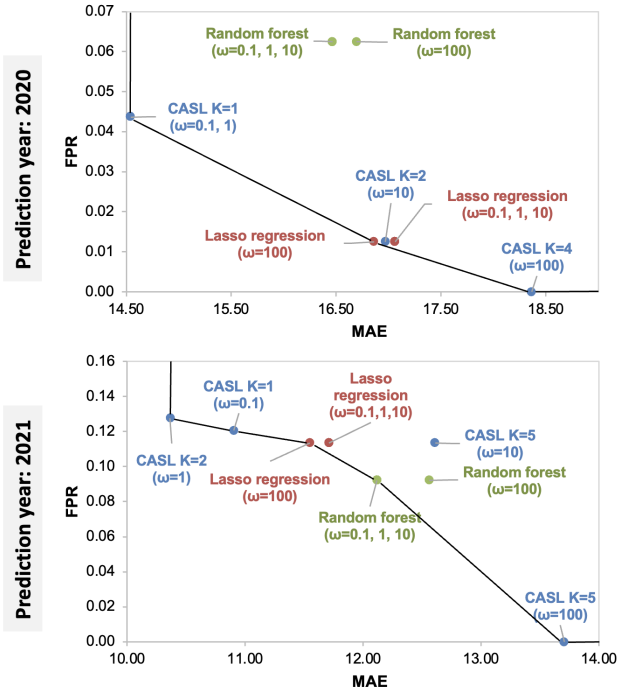

Figure 5: Comparison of mean absolute error (MAE) and false positive rate (FPR) of CASL with random forest and lasso regression under different mixing weight ω.

Characteristics of Clusters One of the unique strengths of the proposed CASL framework compared to other existing methods is to uncover the underlying cluster structures in the data, which can also help improve the interpretability of the prediction results. We present the four clusters identified in our base case ($K = 4$) in Figure 6 and examine how these clusters are different in terms of the characteristics of the counties within each cluster. The map in Figure 6 shows the geographic distribution of the four clusters. The clusters for predicting overdose deaths in the year 2020 are mostly similar to those for 2021 except for some minor differences, which is reasonable since the burden in a county may change as the crisis evolves over time. However, the main structure of the clusters remains relatively robust. Counties in Cluster 1 are primarily urban and more populated areas, and those in Cluster 4 are rural and less populated areas.

Moreover, different clusters also exhibit distinct patterns in various characteristics, as shown in Figure 6: Cluster 1 counties tend to show a higher burden of overdose death, higher incidences of drug-related crimes, higher income, more access to primary health care resources, while Cluster 4 counties tend to show the characteristics in the opposite direction. A comprehensive comparison of counties' characteristics by cluster is provided in Appendix Table A5. These results demonstrate that our proposed CASL model can distinguish counties by subgroups of distinct characteristics in a meaningful way to capture regional heterogeneity and tailor the prediction to the specific cluster.

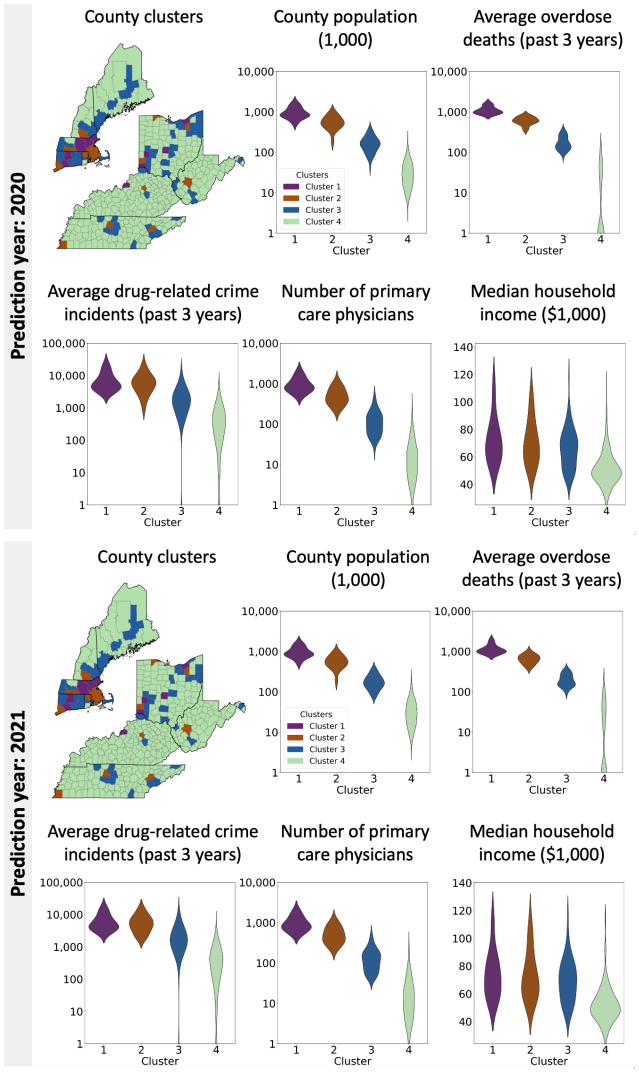

Figure 6: Geographic distribution and characteristics of four clusters identified by the CASL model ($K = 4, \omega = 100$) for the states in New England and Appalachian regions.

Ablation Study To validate the key designs of our CASL framework, we compare its performance with alternative strategies of handling suppressed data and cluster assignment: (1) imputing suppressed data with value of 0, (2) imputing suppressed data with most likely values using multivariate imputation with chained equations (MICE), (3) assigning counties to clusters randomly. Results are summarized in Appendix Table A6. We observe that imputing suppressed values using either method as a data preprocessing step results in higher mixed-error scores for predictions in both years. The degraded performance is largely driven by the increased FPR in suppressed counties, which dominates the marginally improved MAE in unsuppressed counties, highlighting the importance of tailoring the loss function to capture the data suppression within the learning framework. Imputation is not expected to be effective as suppressed data

are unlikely to be missing at random. Similarly, randomly assigning counties to clusters leads to significantly higher FPR in suppressed counties and the overall performance by the mixed-error score, demonstrating the necessity of learning the underlying cluster structures of heterogeneous counties and customizing the prediction based on clusters.

Discussion and Conclusion

This paper presents a cluster-aware supervised learning (CASL) framework for time series data, which is developed to address the unique challenges in the county-level drug overdose prediction problem. Our approach integrates supervised learning and clustering to uncover the underlying subgroup structures in the data and make predictions based on the subgroup, which is well suited to capture the heterogeneity and correlations of the counties. Our proposed approach also handles suppressed data that are common in publicly available data, making the method more broadly applicable.

Using real-world data collected from 10 states for 2010-2021, our numerical study demonstrates that the CASL model can accurately predict the number of overdose deaths for unsuppressed counties and correctly identify suppressed counties. It outperforms other commonly used methods, such as random forest and lasso regression, in most settings. More importantly, our method successfully identifies the clusters of counties that exhibit distinct patterns in overdose death outcomes and demo-socioeconomic characteristics, emphasizing the need for understanding the overdose crisis in counties with different contexts.

One limitation of our study is that our current data have limited time granularity (i.e., by year only) and have not accounted for other time-varying factors related to substance use in general, such as changes in intervention policies, evolving drug supply environment, and implemented treatment and harm reduction programs. Including more such contextual information at the county level may potentially help improve the prediction quality and its generalizability considering the data shift along with the evolving opioid crisis. On the other hand, it could be challenging to curate comprehensive public health surveillance data and data for different intervention programs in communities, given the practical administrative barriers of harmonizing siloed datasets from various agencies and organizations. It calls for more open data sharing to facilitate the translation of predictive modeling research to practical impact on mitigating the opioid crisis (Bharat et al. 2021). In addition, our current CASL framework uses the same feature space for prediction and clustering. Future work can extend this assumption to explore allowing different sets of factors to drive the prediction for different clusters.

Another important consideration that warrants future research is the evaluation of the potential bias in the predictions and its societal impact. For example, the prediction performance from the current model may differ across communities, such as by rural/urban status, health resource deprivation level, and social vulnerability. When applied in practice, the prediction outcomes generated from our model can potentially inform policymaker's decisions on reallocating resources and impact communities in different ways. Therefore, it is critical to understand whether the proposed method results in equitable prediction outcomes across diverse communities. One potential solution to address this issue is to incorporate fairness metrics in the machine learning framework to ensure equitable outcomes and social impact across communities with diverse socioeconomic profiles, and to mitigate model biases that can inadvertently and adversely affect disadvantaged communities.

To facilitate translating the current framework to practical applications for real-world decision-making, additional efforts are needed to enhance coordinated data collection by multiple organizations, to not only expand data with factors that reflect major policy changes and represent the landscape of intervention resources available at the community level, but also to increase data update frequency and time granularity. Beyond the mortality outcome, other drug overdose-related outcomes can also be considered for prediction, leveraging different county clustering structures. It will also be crucial to collaborate closely with stakeholders to refine the selection of meaningful predictors and interpret the machine learning results to draw implications that can drive changes in practice. One future work is to share and communicate the prediction outcomes and their implications with community stakeholders to assist them in developing proactive and effective actionable plans to mitigate the drug overdose crisis.

Acknowledgements

This work was supported by the National Science Foundation under grant agreements CMMI-2240408 and CMMI-2240409. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the US National Science Foundation.

References

Aboulenein, A. 2022. Opioid crisis cost U.S. nearly $1.5 trillion in 2020 - congressional report. https://www.reuters.com/world/us/opioid-crisis-cost-us-nearly-15-trillion-2020-congressional-report-2022-09-28/. Accessed: August 1, 2024.

Aslam, H.; and Biswas, S. 2023. Analysis of COVID-19 death cases using machine learning. *SN Comput. Sci.*, 4(4): 403.

Banerjee, K.; Bali, V.; Nawaz, N.; Bali, S.; Mathur, S.; Mishra, R. K.; and Rani, S. 2022. A machine-learning approach for prediction of water contamination using latitude, longitude, and elevation. *Water (Basel)*, 14(5): 728.

Bharat, C.; Hickman, M.; Barbieri, S.; and Degenhardt, L. 2021. Big data and predictive modelling for the opioid crisis: existing research and future potential. *Lancet Digit. Health*, 3(6): e397–e407.

Blanco, C.; Wiley, T. R. A.; Lloyd, J. J.; Lopez, M. F.; and Volkow, N. D. 2020. America's opioid crisis: the need for an integrated public health approach. *Transl. Psychiatry*, 10(1): 167.

Campo, D. S.; Gussler, J. W.; Sue, A.; Skums, P.; and Khudyakov, Y. 2020. Accurate spatiotemporal mapping of drug overdose deaths by machine learning of drug-related web-searches. *PLoS One*, 15(12): e0243622.

Cano, M.; Oh, S.; Osborn, P.; Olowolaju, S. A.; Sanchez, A.; Kim, Y.; and Moreno, A. C. 2023. County-level predictors of US drug overdose mortality: a systematic review. *Drug and alcohol dependence*, 242: 109714.

Cano, M.; Timmons, P.; Hooten, M.; Sweeney, K.; and Oh, S. 2024. A scoping review of law enforcement drug seizures and overdose mortality in the United States. *Int. J. Drug Policy*, 124(104321): 104321.

Centers for Disease Control and Prevention. 2024a. CDC wide-ranging online data for epidemiologic research (WONDER) multiple cause of death data. Accessed: August 1, 2024.

Centers for Disease Control and Prevention. 2024b. Understanding the opioid overdose epidemic. https://www.cdc.gov/drugoverdose/epidemic/index.html. Accessed: August 1, 2024.

Chen, Q.; Sterner, G.; Rhubart, D.; Newton, R.; Shaw, B.; and Scanlon, D. 2024. Creating a robust coordinated data and policy framework for addressing substance use issues in the United States. *International Journal of Drug Policy*, 134: 104629.

Chen, Q.; Sterner, G.; Segel, J.; and Feng, Z. 2022. Trends in opioid-related crime incidents and comparison with opioid overdose outcomes in the United States. *Int. J. Drug Policy*, 101(103555): 103555.

Chen, S.; and Xie, W. 2022. On cluster-aware supervised learning: frameworks, convergent algorithms, and applications. *INFORMS Journal on Computing*, 34(1): 481–502.

Ciccarone, D. 2019. The triple wave epidemic: supply and demand drivers of the US opioid overdose crisis. *International Journal of Drug Policy*, 71: 183–188.

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.

Cuomo, R.; Purushothaman, V.; Calac, A. J.; McMann, T.; Li, Z.; and Mackey, T. 2023. Estimating county-level overdose rates using opioid-related Twitter data: Interdisciplinary infodemiology study. *JMIR Form. Res.*, 7: e42162.

Devijver, E. 2017. Model-based regression clustering for high-dimensional data: application to functional data. *Adv. Data Anal. Classif.*, 11(2): 243–279.

Duff, J. H.; Shen, W. W.; Rosen, L. W.; and Lampe, J. R. 2022. The Opioid crisis in the United States: a brief history. Congressional Research Service.

Florence, C.; Luo, F.; and Rice, K. 2021. The economic burden of opioid use disorder and fatal opioid overdose in the United States, 2017. *Drug and alcohol dependence*, 218: 108350.

Ghosh, A.; Bisaga, A.; Kaur, S.; and Mahintamani, T. 2022. Google Trends data: A potential new tool for monitoring the opioid crisis. *Eur. Addict. Res.*, 28(1): 33–40.

Google. 2024. Google Distance Matrix API. https://developers.google.com/maps/documentation/distance-matrix/start. Accessed: 2024-08-13.

Jalali, M. S.; Botticelli, M.; Hwang, R. C.; Koh, H. K.; and McHugh, R. K. 2020a. The opioid crisis: a contextual, social-ecological framework. *Health Res. Policy Syst.*, 18(1): 87.

Jalali, M. S.; Botticelli, M.; Hwang, R. C.; Koh, H. K.; and McHugh, R. K. 2020b. The opioid crisis: need for systems science research. *Health Res. Policy Syst.*, 18(1): 88.

Kariisa, M.; Davis, N. L.; Kumar, S.; Seth, P.; Mattson, C. L.; Chowdhury, F.; and Jones, C. M. 2022. Vital signs: Drug overdose deaths, by selected sociodemographic and social determinants of health characteristics - 25 states and the District of Columbia, 2019-2020. *MMWR Morb. Mortal. Wkly. Rep.*, 71(29): 940–947.

Langabeer, J. R.; Chambers, K. A.; Cardenas-Turanzas, M.; and Champagne-Langabeer, T. 2022. County-Level factors underlying opioid mortality in the United States. *Substance Abuse*, 43(1): 76–82.

Li, J.; Heap, A. D.; Potter, A.; and Daniell, J. J. 2011. Application of machine learning methods to spatial interpolation of environmental variables. *Environ. Model. Softw.*, 26(12): 1647–1659.

Lin, Q.; Aguilera, J. A. R.; Williams, L. D.; Mackesy-Amiti, M. E.; Latkin, C.; Pineros, J.; Kolak, M.; and Boodram, B. 2023. Social-spatial network structures among young urban and suburban persons who inject drugs in a large metropolitan area. *International Journal of Drug Policy*, 122: 104217.

Lo-Ciganic, W.-H.; Huang, J. L.; Zhang, H. H.; Weiss, J. C.; Wu, Y.; Kwoh, C. K.; Donohue, J. M.; Cochran, G.; Gordon, A. J.; Malone, D. C.; Kuza, C. C.; and Gellad, W. F. 2019. Evaluation of machine-learning algorithms for predicting opioid overdose risk among Medicare beneficiaries with opioid prescriptions. *JAMA Netw. Open*, 2(3): e190968.

Marks, C.; Abramovitz, D.; Donnelly, C. A.; Carrasco-Escobar, G.; Carrasco-Hernández, R.; Ciccarone, D.; González-Izquierdo, A.; Martin, N. K.; Strathdee, S. A.; Smith, D. M.; and Bórquez, A. 2021. Identifying counties at risk of high overdose mortality burden during the emerging fentanyl epidemic in the USA: a predictive statistical modelling study. *Lancet Public Health*, 6(10): e720–e728.

Mattson, C. L.; Tanz, L. J.; Quinn, K.; Kariisa, M.; Patel, P.; and Davis, N. L. 2021. Trends and geographic patterns in drug and synthetic opioid overdose deaths - United States, 2013-2019. *MMWR Morb. Mortal. Wkly. Rep.*, 70(6): 202–207.

Mukherjee, S.; Becker, N.; Weeks, W.; and Ferres, J. L. 2020. Using internet search trends to forecast short term drug overdose deaths: a case study on Connecticut. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE. Miami, FL, USA.

Nam, Y.; Shea, D.; Shi, Y.; and Moran, J. 2017. State prescription drug monitoring programs and fatal drug overdoses. *Am. J. Manag. Care*, 23(5): 297–303.

Rawson, R. A.; Erath, T. G.; and Clark, H. W. 2023. The fourth wave of the overdose crisis: Examining the prominent

role of psychomotor stimulants with and without fentanyl. *Prev. Med.*, 176(107625): 107625.

Saloner, B.; Chang, H.-Y.; Krawczyk, N.; Ferris, L.; Eisenberg, M.; Richards, T.; Lemke, K.; Schneider, K. E.; Baier, M.; and Weiner, J. P. 2020. Predictive modeling of opioid overdose using linked statewide medical and criminal justice data. *JAMA Psychiatry*, 77(11): 1155–1162.

Santangelo, O. E.; Gentile, V.; Pizzo, S.; Giordano, D.; and Cedrone, F. 2023. Machine learning and prediction of infectious diseases: a systematic review. *Machine Learning and Knowledge Extraction*, 5(1): 175–198.

Schell, R. C.; Allen, B.; Goedel, W. C.; Hallowell, B. D.; Scagos, R.; Li, Y.; Krieger, M. S.; Neill, D. B.; Marshall, B. D. L.; Cerda, M.; and Ahern, J. 2022. Identifying predictors of opioid overdose death at a neighborhood level with machine learning. *Am. J. Epidemiol.*, 191(3): 526–533.

Srinivasan, S.; Pustz, J.; Marsh, E.; Young, L. D.; and Stopka, T. J. 2024. Risk factors for persistent fatal opioid-involved overdose clusters in Massachusetts 2011-2021: a spatial statistical analysis with socio-economic, accessibility, and prescription factors. *BMC Public Health*, 24(1): 1893.

Sumner, S. A.; Bowen, D.; Holland, K.; Zwald, M. L.; Vivolo-Kantor, A.; Guy, G. P., Jr; Heuett, W. J.; Pressley, D. P.; and Jones, C. M. 2022. Estimating weekly national opioid overdose deaths in near real time using multiple proxy data sources. *JAMA Netw. Open*, 5(7): e2223033.

Tatar, M.; Faraji, M. R.; Keyes, K.; Wilson, F. A.; and Jalali, M. S. 2023. Social vulnerability predictors of drug poisoning mortality: A machine learning analysis in the United States. *Am. J. Addict.*, 32(6): 539–546.

Thurston, H.; and Freisthler, B. 2020. The spatio-temporal distribution of naloxone administration events in rural Ohio 2010-16. *Drug Alcohol Depend.*, 209(107950): 107950.

Tran, C. T.; Zhang, M.; Andreae, P.; Xue, B.; and Bui, L. T. 2018. Improving performance of classification on incomplete data using feature selection and clustering. *Appl. Soft Comput.*, 73: 848–861.

University of Wisconsin Population Health Institute. 2024. County Health Rankings & Roadmaps 2024. www.countyhealthrankings.org. Accessed: August 1, 2024.

Wang, B.; Ning, L.; and Kong, Y. 2019. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment. *Expert Syst. Appl.*, 128: 301–315.

Wang, Y.; Yan, Z.; Wang, D.; Yang, M.; Li, Z.; Gong, X.; Wu, D.; Zhai, L.; Zhang, W.; and Wang, Y. 2022. Prediction and analysis of COVID-19 daily new cases and cumulative cases: times series forecasting and machine learning models. *BMC Infect. Dis.*, 22(1): 495.

Zhu, Z.; Li, Y.; and Kong, N. 2012. Clusterwise linear regression with the least sum of absolute deviations–an MIP approach. *International Journal of Operations Research*, 9(3): 162–172.