scientific reports

OPEN

Measuring network dynamics of opioid overdose deaths in the United States

Kushagra Tiwari[1✉], M. Amin Rahimian[1✉], Mark S. Roberts[2], Praveen Kumar[2] & Jeanine M. Buchanich[3]

The US opioid overdose epidemic has been a major public health concern in recent decades. There has been increasing recognition that its etiology is rooted in part in the social contexts that mediate substance use and access; however, reliable statistical measures of social influence are lacking in the literature. We use Facebook's social connectedness index (SCI) as a proxy for real-life social networks across diverse spatial regions that help quantify social connectivity across different spatial units. This is a measure of the relative probability of connections between localities that offers a unique lens to understand the effects of social networks on health outcomes. We use SCI to develop a variable, called "deaths in social proximity", to measure the influence of social networks on opioid overdose deaths (OODs) in US counties. Our results show a statistically significant effect size for deaths in social proximity on OODs in counties in the United States, controlling for spatial proximity, as well as demographic and clinical covariates. The effect size of standardized deaths in social proximity in our cluster-robust linear regression model indicates that a one-standard-deviation increase, equal to 11.70 more deaths per 100,000 population in the social proximity of ego counties in the contiguous United States, is associated with thirteen more deaths per 100, 000 population in ego counties. To further validate our findings, we performed a series of robustness checks using a network autocorrelation model to account for social network effects, a spatial autocorrelation model to capture spatial dependencies, and a two-way fixed-effect model to control for unobserved spatial and time-invariant characteristics. These checks consistently provide statistically robust evidence of positive social influence on OODs in US counties. Our analysis provides a pathway for public health interventions informed by social network structures. The statistical robustness of our primary variable of interest, deaths in social proximity, supports the hypothesis of a social network effect on OODs. Using agent-based modeling (ABM) to simulate social networks can offer an effective method to design interventions that incorporate the dynamics of social networks for maximum impact.

The opioid overdose epidemic is a major public health crisis in the US, with an exponentially increasing number of drug overdose deaths in the last four decades[1,2]. Alpert et al.[3] report that opioid overdose deaths (OODs) account for 75% of the increase in drug overdose deaths. Addressing this crisis requires planned interventions that focus on supply-side regulations and the dynamics of social connections. The rate of initiation of opioid misuse is known to increase due to social influence[4]. Costello et al.[5] report that of the 370 participants who entered an opioid withdrawal program, 97% identified knowing the individual with whom they initiated opioid use, with friendship being the most reported relationship between participants and their initiation partners. Similarly, Rigg et al.[6] note that two-thirds of misused drugs are obtained from friends and family. Guarino et al.[7] study of 539 young adults who misuse opioids and heroin indicates that their first experiences with the misuse of prescription opioids typically occur in a social setting with peers. The misuse of prescription opioids has been growing among young people[8]. Syvertsen et al.[9] make similar observations about young people who experiment with drugs and the initiation of drug use. Social networks can have positive and negative impacts on substance use. Empirical results have indicated that peer networks with subjects who do not use substances have a positive influence on curbing drug abuse; however, networks consisting of more substance users are likely to increase substance use[10]. In intervention designs, recovery support strategies, including peer recovery, have shown encouraging results. Peer workers who have completed their recovery help others in recovery from substance

[1]Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, USA. [2]Department of Health Policy and Management, University of Pittsburgh, Pittsburgh, USA. [3]Department of Biostatistics and Health Data Science, University of Pittsburgh, Pittsburgh, USA. ✉email: kut20@pitt.edu; rahimian@pitt.edu

addiction. This form of peer-supported recovery is found to be more effective in reducing the prevalence of opioid use disorder (OUD)[11].

Providing reliable measures of social influence on the opioid epidemic is complicated by the confounding factors that influence opioid misuse and social interactions, on the one hand, and the ethical barriers to randomized experiments, on the other. Studies measuring social influence use complex systems-based generative models to understand these phenomena in areas such as voting contagion[12]. Specifically, Braha and Aguiar[12] use a generalized social voter model combined with spatial-statistical analysis to examine how social influence has shaped voting behavior in the US presidential election over the past century. They distinguish between social contagion and external influences (e.g., media and opinion leaders) to assess their impact on county-level vote share distributions over time and geography. By analyzing spatial patterns, they demonstrate that social influence is geographically clustered and spreads like a contagion across county borders. In contrast, applying similar modeling techniques to the opioid epidemic to design targeted interventions has encountered certain limitations. Homer et al.[13] discuss the complexity of modeling OUD and highlight the limitations of their models in accurately capturing real-life scenarios due to their simplicity. Recognizing the complexities of modeling the opioid epidemic through generative processes, our research aims to address these gaps using controlled regressions.

From the traditional setting to the new digital era of social networks, we have witnessed a significant shift in informal social interactions. Facebook is the largest informal online social network globally, with 2.1 billion active users and 239 million active users in the US and Canada. Given its broad reach and prevalence, Facebook connections can provide insight into real-life social networks in many geographical regions. Facebook has released a social connectedness index dataset that measures how intensely geographical locations are connected according to the relative probability of connections[14]. In Supplementary Information section S1 we provide a review of SCI use cases as a proxy for real-life social connections in public health, epidemiology, economics and development applications. In the following paragraph, we justify our use of SCI for measuring the network dynamics of OODs in the US.

We use SCI to construct a measure of OODs in the social proximity of counties in the United States and investigate the statistical significance of its effect on county-level OODs after controlling for clinical, spatiotemporal, and socioeconomic confounders. Our analysis seeks to provide information on the role of social influence on OODs. Using SCI to represent the intensity of interpersonal networks between counties reflects the possibility of physical interactions between county residents. Bailey et al.[15,16] show that SCI is predictive of travel patterns within urban areas and throughout Europe. Coven et al.[17] show that counties with higher social connections to New York City are the most preferred destinations for those moving away from the city during the pandemic, highlighting the association of physical interaction with SCI. Kuchler et al.[18] use SCI to show that COVID-19 is more likely to spread between regions with stronger social networks and highlighted the potential of SCI to improve the prediction of epidemics. Although SCI provides a robust measure of social connectedness, some studies have explored alternative data sources, such as social media platforms, to examine trends in the context of the opioid epidemic. However, these approaches have faced significant challenges due to demographics and other data limitations[19–21]. For example, Pandrekar et al.[20] use Reddit data to analyze psychological categories and patterns of opioid abuse on a national scale. A major limitation of their study is that the data collected through the Reddit API do not provide access to user location information. In addition, the Reddit data do not indicate whether users are friends, which restricts the ability to analyze the structure of social networks and their association with OODs.

In contrast, SCI offers distinct advantages. SCI is location-based and provides detailed information about the structure of social networks in US counties. SCI measures friendship networks, serving as a proxy for real-world social connections. Unlike Reddit data, SCI allows for the analysis of location-specific friendship networks, which makes it particularly useful for studying how social networks influence the opioid epidemic on a population scale. Our choice of SCI as the network measure is informed by previous use cases that reflect the real-life dynamics of social connections in different domains such as education[22] and public health[18]. Our objective is to measure how these social connections contribute to heterogeneous patterns of opioid overdose deaths in US counties.

In our study, we present a novel perspective on analyzing the association between friendship networks and opioid overdose deaths at the population level in counties within three distinct geographical regions: the eastern United States, the central and western United States, and the entire contiguous United States. We provide statistical evidence linking the geographical spread of OODs with the structure of social connections.

To achieve this, we use a range of statistical methodologies. We use cluster-robust linear regression to account for intra-cluster correlation between counties within the same states, network and spatial autocorrelation methods to address the autocorrelation in error terms arising from unobserved factors shared among spatially or socially connected units, and two-way fixed effects models to control for unobserved spatial and temporal heterogeneities. Additionally, we perform robustness checks to account for the distance decay of proximity weights and apply a two-stage least squares method to jointly address spatial and network effects, discussed in the Supplementary Information sections S5 and S4.5. Our multifaceted statistical analysis demonstrates that our variable of interest, deaths in social proximity, is statistically significant across the three distinct geographical regions.

Understanding the role of social networks is important in designing interventions to combat opioid misuse behavior. Research measuring the network dynamics of opioid overdose death on the US population scale and at the resolution of counties remains limited, with the exception of[23,24] that measure social network drivers of the opioid epidemic and use natural experiments to support their claims. Mäckle and Reunzi[23] examine changes in county-level overdose deaths due to the reformulation of OxyContin and the must-access Prescription Drug Monitoring Program (PDMP). They analyze policy-induced shocks to estimate the indirect effect of friendship

networks (measured by SCI) on OODs. Their analysis includes the correlation of "social proximity" with OODs by using a two-way fixed-effects model. Compared to Mäckle and Reunzi[23], our study differs in the methodological approach. Along with cluster-robust linear regression and two-way fixed effects, our analysis uses a network and spatial autocorrelation model to account for autocorrelation in error terms, which Mäckle and Reunzi[23] do not consider when investigating the association of "social proximity" with OODs. We also include a robustness check to account for spatial and network autocorrelation together and provide evidence that the statistical significance of the "deaths in social proximity" variable is not sensitive to the choice of distance decay function. Furthermore, our regression includes domain-specific covariates such as availability of naloxone and buprenorphine, opioid dispensing rate, mental health distress rate and state-level fentanyl and analog seizure data to account for illicit opioids which Mäckle and Reunzi[23] do not consider when investigating the association between "social proximity" and OOD. Furthermore, Mäckle and Reunzi (2023) use the National Vital Statistics System (NVSS) public database and employ a backout procedure to recover mortality data points that the NVSS otherwise suppresses if the county has fewer than ten deaths. In contrast, our study uses mortality data from the National Center for Health Statistics (NCHS) which provides access to these suppressed data points. In this regard, our results complement their findings, as their coefficient for "social proximity" is also statistically significant. Concurrently with our work, Cuttler and Donahoe[24] have explored the dynamics of opioid death rates, focusing on SCI and the distance between counties to analyze spillovers. They posit that the increase in opioid demand is endogenous, resulting from spillovers between affected populations. Their study underscores the importance of social and spatial spillovers in estimating opioid demand, which is correlated with increased mortality rates. However, our work differs in its methodological approach, concentrating on constructing a socially lagged variable to assess the impact of social influence on overdose death rates.

In the following section, we formally introduce our variable of interest, "deaths in social proximity". The descriptive statistics serve as a starting point for later estimating the effect size of deaths in social proximity in the "Results" section.

OOD rates in social and spatial proximity

The root of the opioid epidemic is partially associated with social contexts that mediate substance use and accessibility. Existing studies integrating social network analysis with the geographical spread of overdose deaths have demonstrated how social characteristics influence the trajectory of substance use, for example, geographic discordance, which means that the community in which the overdose death occurs is different from the community of residence[25]. Using data on overdose deaths from 2017 to 2020 in Milwaukee County, Wisconsin, Forati et al.[25] build a social-spatial network framework to detect network interaction hotspots for overdose deaths and analyze their geographical discordance. However, their study is limited to Milwaukee County and does not extend to US population scales.

As discussed in the Introduction, the significant contribution of social influence in initiating substance use is well documented in the literature. Researchers have highlighted the impact of social networks on the regulation of substance use patterns based on the ego network[26]. However, applying these findings to a geographical context presents challenges, as geographical proximity substantially influences social connections and communication patterns across varying distances[27]. In our model, we control for the inherent spatial patterns of geographical proximity to refine our estimate of the effect of social networks on OODs.

Assessing the strength of social ties within every individual's network in a wide geographical area is very resource-intensive. To address this challenge, one potential approach is to aggregate and estimate the social networks of metapopulations residing in different localities, such as ZIP codes or counties in the United States. This offers a broader perspective and alleviates some of the constraints associated with individual-level analysis. In 2018, Meta Platforms, which operates the Facebook social network, released a data set that measures the distribution of social ties of location-specific networks globally. SCI is a surrogate for real-life friendship networks between registered Facebook users at each location[14]. It quantifies the strength of friendship ties in various locations using relative probability and is available at the ZIP code and county level for the United States. Formally, the SCI between two locations i and j is defined as follows[14]:

$$\mathrm{SCI}_{ij} = \frac{\mathrm{Facebook\ \ Connections}_{ij}}{\mathrm{Facebook\ \ Users}_i \times \mathrm{Facebook\ \ Users}_j}.$$

Here, $\mathrm{Facebook\ \ Users}_i$ represents the number of Facebook users in the county i. $\mathrm{Facebook\ \ Connections}_{ij}$ is the total number of Facebook friendship connections between individuals in counties i and j.

Building on the established link between SCI and real-world social connections, we introduce two variables, "deaths in social proximity" and "deaths in spatial proximity", to capture the influence of social and spatial factors on the distribution of OOD rates in counties in the United States. The term "deaths in social proximity" indicates the average number of death rates in alter locations weighted by their SCI to the focal node, also known as "ego". This variable operates as a socially lagged variable, accounting for the influence of death rates in "alter" locations, referring to the direct connections of the ego. On the other hand, "deaths in spatial proximity" measures the average number of death rates in the alters weighted by their inverse geographical distance to the ego. Unlike the socially lagged variable, this is a spatially lagged variable to account for the effects of deaths in the spatial vicinity of the ego. These factors work together to provide a comprehensive picture of how deaths are distributed and influenced by social and geographical factors in a given location. Quantitatively, deaths in social and spatial proximity, denoted by s_{-i} and d_{-i} for an ego location i, are defined as follows:

$$s_{-i} = \sum_{j \neq i} w_{ij} y_j, \quad \text{and} \quad d_{-i} = \sum_{j \neq i} a_{ij} y_j,$$

where y_i is OOD rate in US county *i*. The social and spatial proximity weights (w_{ij} and a_{ij}) are defined as follows:

$$w_{ij} = \frac{n_j \; \text{SCI}_{ij}}{\sum_{k \neq i} n_k \; \text{SCI}_{ik}}, \quad \text{and} \quad a_{ij} = \frac{1 + \frac{1}{d_{ij}}}{\sum_{k \neq i} \left(1 + \frac{1}{d_{ik}}\right)},$$

where n_j is the population of county *j* and d_{ij} is the distance between county locations *i* and *j*. Using $1 + 1/d_{ij}$ instead of $1/d_{ij}$ in the definition of a_{ij} improves numerical stability when dealing with long distances (large d_{ij} values) but does not change the decreasing nature of the spatial proximity weights with the increasing distances. To capture the effect of far away counties that would be otherwise down-weighted heavily by $1/d_{ij}$, in Supplementary Information section S5 we repeat our analysis with a slowly decaying distance function $\left(1/d_{ij}^{1/10}\right)$ and observe that our main conclusions remain unchanged and are not sensitive to the choice of the distance decay function. To visualize the social and spatial adjacency weights across the United States, we aggregate the county-level data to the state level. This aggregation allows us to effectively depict the dense networks. We give the details of the state level aggregation in Supplementary Information section S2. It is important to note that our analysis of estimating the effect of "deaths in social proximity" on OODs is still conducted at the county level. The visualization of the proximity weights and proximity values, however, is presented at the state level to help understand the geographical dispersion of the social and spatial proximity weights in relation to OOD rates. For comparison, in Supplementary Information section S2 we give the county-level visualization of social and spatial adjacency weights in Pennsylvania (PA). Figure 1A illustrates the state-level social network, measured by the social proximity weights (w_{xz}) in the contiguous United States. Figure 1B and C show the spatial dispersion of the proximity weights for the socially and spatially lagged variables for two ego states, California (1B) and Pennsylvania (1C), in the contiguous US.

Having formally introduced the socially and spatially lagged variables, we use mortality data from the National Center for Health Statistics (NCHS) for 2018-2019 to measure the state-level OOD rates. Figure 2 shows the state-level spatial distributions of our main variables of interest in the contiguous US: OOD rates (Fig. 2A), deaths in social proximity (Fig. 2B), deaths in spatial proximity (Fig. 2C), and their differences (Fig. 2D). Figure 3 shows the scatter plots of death rates (y_i), deaths in social proximity (s_{-i}), and deaths in spatial proximity (d_{-i}) for all counties in the contiguous US. The scatter plot matrix reveals a moderate linear dependence between death rates and spatially and socially lagged variables. In addition, there is a strong correlation between spatial and social proximity. However, the histograms that represent the distribution of these two variables exhibit differences. This contrast helps us identify the spatial effects of social influence on OOD and estimate an effect size for social proximity using controlled regressions described in the Results. Our choice of controls comes from domain knowledge consisting of clinical covariates and factors of social determinants of health (SDOH). The SDOH covariates are selected using the least absolute shrinkage and selection operator (LASSO) from an array of 17 variables. The details of SDOH and clinical covariates are in "Methods".

Results
Estimating the effect size of "Deaths in Social Proximity" variable
Our outcome of interest is the county-level OOD rates that we measure using NCHS data from 2018 to 2019. We use cluster-robust linear regression to estimate the coefficient of the socially lagged variable. Robust estimators and clustering by state help us correct for correlation among counties, which might be higher for counties in the same state than between different states. Figure 4 shows the significant coefficient of the socially lagged variable and provides statistical evidence for the effect of social networks on the spatial spread of OOD rates. The estimate of the effect size of the socially lagged variables is statistically significant across the eastern, western-central, and the entire contiguous US. The positive coefficient aligns with the theoretical proposition of the literature on the importance of social influence in the opioid epidemic[5]. The positive and significant magnitude of this effect size can originate from the dissemination of information through social networks about the initiation of use and availability of substances, leading to more OODs. The effect sizes for s_{-i} and d_{-i}, derived from the cluster-robust standard error model, along with other covariates, are in Supplementary Tables S2, S3 and S4 for the eastern, western-central, and contiguous US.

We standardize s_{-i} and d_{-i} prior to regression analysis. Consequently, the effect size for s_{-i} indicates that an increase of one standard deviation, equal to 11.69523 and 12.2417 more "deaths in social proximity" per 100, 000 population in the contiguous and eastern United States, respectively, is associated with an increase of nine deaths per 100, 000 population in ego counties within the eastern United States and thirteen deaths per 100, 000 population in the contiguous United States. For counties in the western and central United States, a similar increase of one standard deviation, equal to 5.7145 more deaths per 100, 000 population in the social proximity of ego counties, corresponds to six more deaths per 100, 000 population in ego counties. Despite observing a significant effect for the socially lagged variable on OOD rates, it is crucial to address potential issues arising from the inherent nature of the primary variable of interest. We recognize the methodological challenges due to correlated residuals when using statistical models to analyze social and spatial effects and perform robustness checks using network and spatial autocorrelation, as well as fixed-effects models to substantiate our findings.
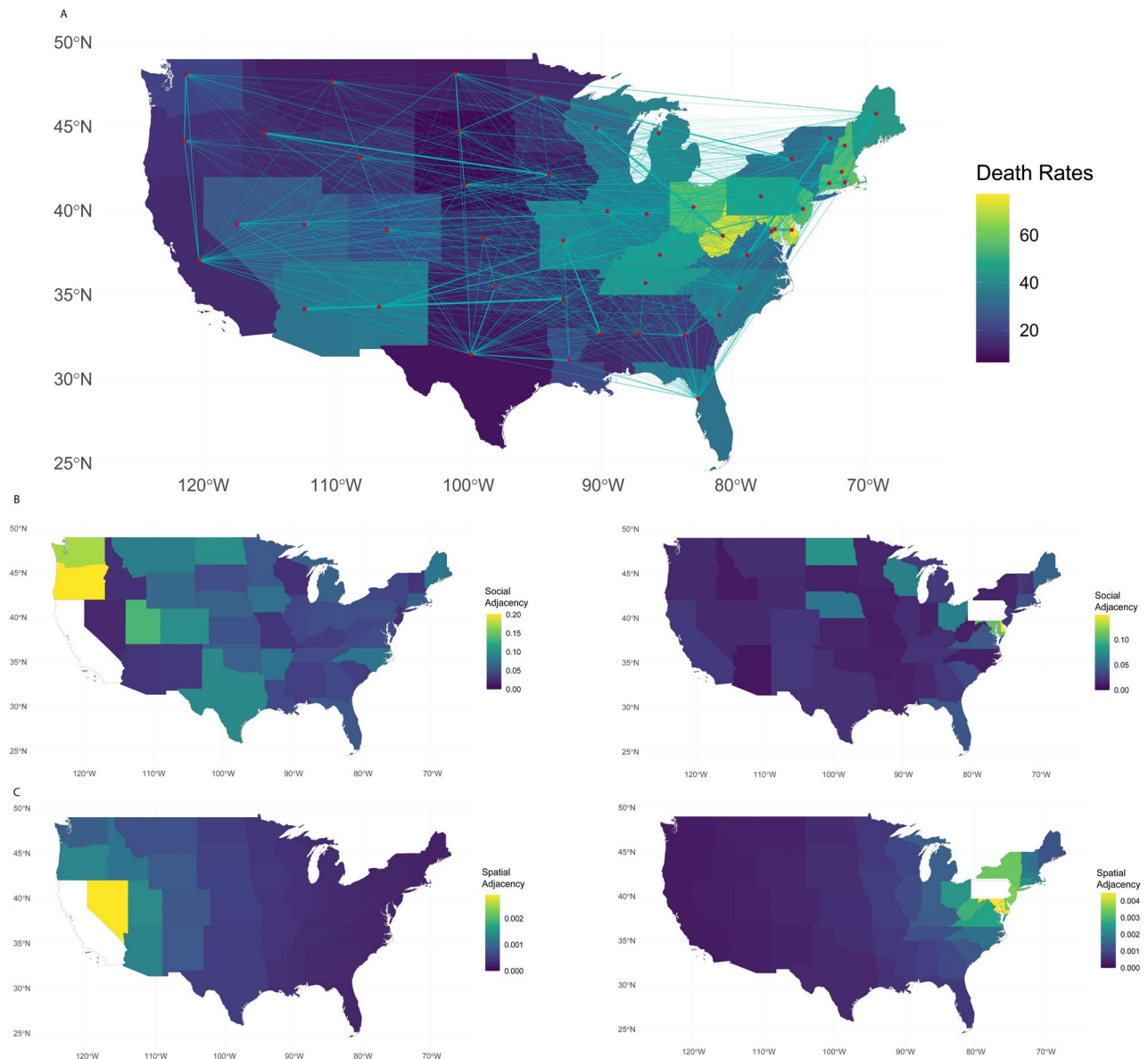
Fig. 1. (**A**) The spatial distribution of overdose death rates per 100,000 population in contiguous US from 2018 to 2019. Superimposed on this map is a social network diagram with edge widths representing the state-level social proximity weights. (**B**) The two middle maps show the social proximity weights of alter states to California (on the left) and Pennsylvania (on the right). (**C**) The bottom two maps show the spatial proximity weights of alter states to California (on the left) and Pennsylvania (on the right). Maps generated using R version 4.2.2 (2022-10-31, available at https://cran.r-project.org. Network diagram created using Gephi 0.10.1 (2023-01-17), available at https://gephi.org. Image editing performed with Adobe Illustrator 27.9.5.

Given that lag variables have a social and spatial component, we expect the error terms in our regression model to be correlated. To address this, we implemented a spatial error model (SEM) to test and correct for network and spatial autocorrelations in error terms. The autocorrelation in error terms arises from unobserved factors shared among spatially or socially connected units. Cluster-robust linear regression may not capture these autocorrelations originating from unobserved factors, which may lead to bias and inefficient estimates. Hence, to add robustness to our analysis, we utilize spatial and network aurocorrelation models. The methodological frameworks for the network and spatial autocorrelation models are explained in Supplementary Information sections S4.2 and S4.3. Supplementary Tables S5, S6 and S7 provide the SEM regression results for the network and spatial autocorrelation models for the eastern, western-central and contiguous United States, in which we find statistical evidence for correlated error. It is important to note that we performed two distinct models to test and correct for correlated error terms originating from both spatial and network dependence separately. Network autocorrelation might come from the structure of the socially lagged variable s_{-i}, while spatially correlated error terms could be attributed to the spatially lagged variable d_{-i}. Our autocorrelation models provide statistical evidence for the significance of social proximity in OODs.
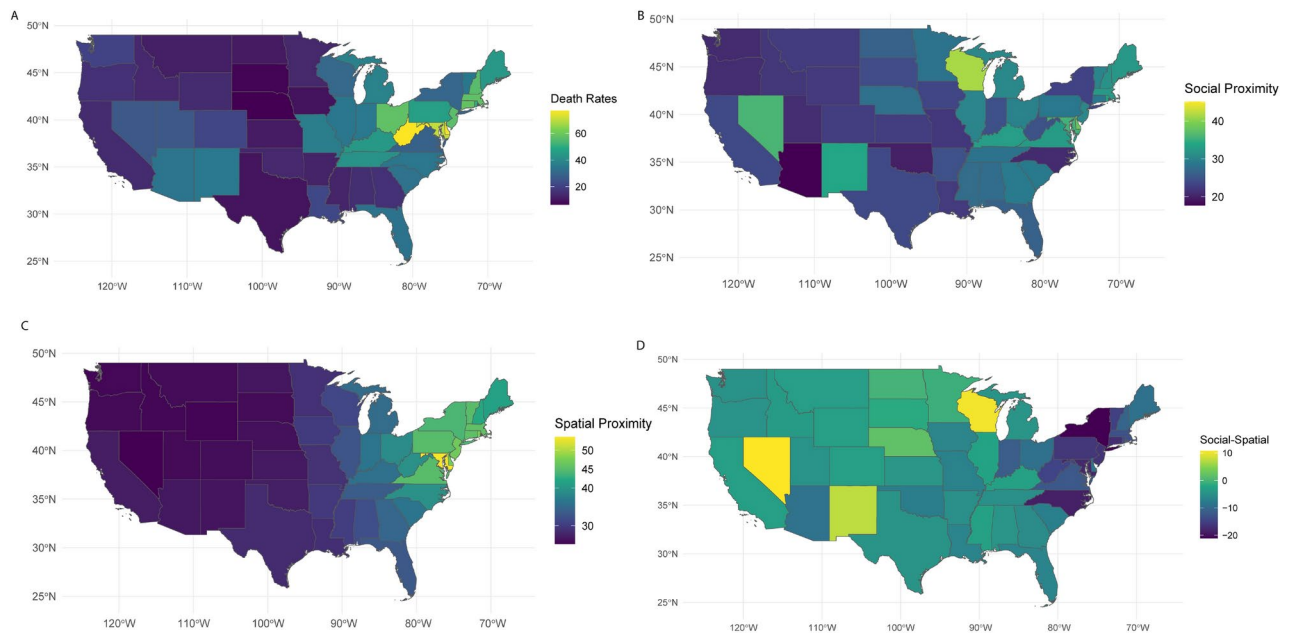
Fig. 2. (**A**) The top left map shows the spatial spread of state-level opioid overdose death rates in the contiguous US. (**B**) The top right map shows the spatial dispersion of "deaths in social proximity" for states in the contiguous US. (**C**) The bottom left map shows the geographical spread of "deaths in spatial proximity". (**D**) The bottom right map shows the difference between deaths in social and spatial proximity from top right and bottom left maps. Maps generated using R version 4.2.2 (2022-10-31, available at https://cran.r-project.org). Image editing performed with Adobe Illustrator 27.9.5.

Fig. 3. The figure shows the distribution and relationships between the primary variables of interest (death rates y_i, deaths in social proximity s_{-i}, and deaths in spatial proximity d_{-i}). The histograms on the main diagonal depict the distributions of y_i, s_{-i}, and d_{-i}. Moving to the upper triangle, we observe the degree of linear dependence between these variables, while the lower triangle displays scatter plots. Scatter plot matrix generated using the R version 4.2.2 (2022-10-31, available at https://cran.r-project.org). Image editing performed with Adobe Illustrator 27.9.5.
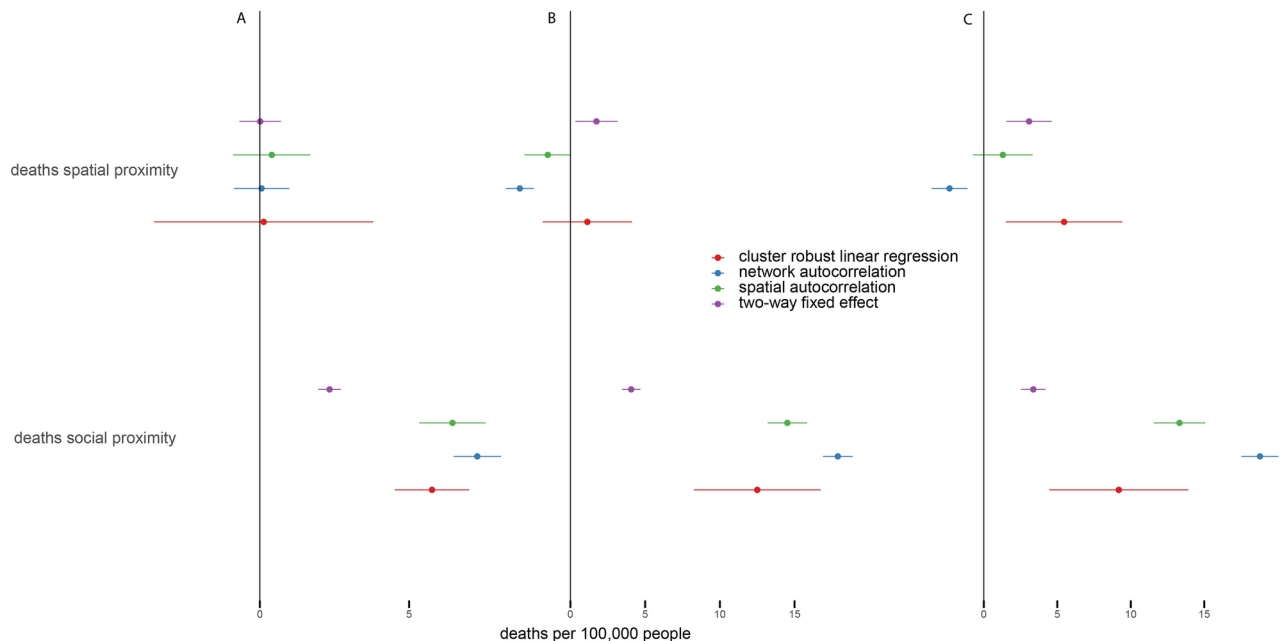
Fig. 4. (**A**) The plot shows the coefficient confidence interval plots for western and central US counties. The coefficient for s_{-i} for cluster-robust linear regression (Supplementary Table S3), network and spatial autocorrelation (Supplementary Table S6), and two-way fixed effects model (Supplementary Table S9), all indicate a positive, significant (p < 0.001) coefficient for s_{-i}. (**B**). Shows the coefficient plot for social and spatial proximity for counties in the contiguous US. The coefficient for s_{-i} for cluster-robust linear regression (Supplementary Table S4), network and spatial autocorrelation (Supplementary Table S7), and two-way fixed effects models (Supplementary Table S10) are all positive and significant (p< 0.001). (**C**) Shows the coefficient plot for s_{-i} and d_{-i} for counties in the eastern US. The coefficient for s_{-i} for cluster-robust linear regression (Supplementary Table S2), network and spatial autocorrelation models (Supplementary Table S5), and two-way fixed effects models (Supplementary Table S8) are all positive and significant (p < 0.001). The effect sizes for standardized s_{-i} in the cluster-robust linear regression models indicate that a one-standard-deviation increase, equal to 11.69523, 12.2417, and 5.7145 more deaths per 100, 000 population in the social proximity of the ego counties in contiguous, eastern and western-central United States, respectively, is associated with thirteen more deaths per 100, 000 population in contiguous and nine more deaths eastern and six more deaths in western-central US counties.

To address spatial and temporal heterogeneity and enhance robustness, we employed a two-way fixed-effects model. Specifically, we included state-fixed effects to control for unobserved state-specific characteristics, such as regulations and policy environments, that are constant over time, but vary between states. We also incorporate year-fixed effects to absorb nationwide shocks or trends that could influence our outcome of interest. This modeling strategy enables us to robustly assess the statistical significance of "deaths in social proximity" while accounting for unobserved spatial and temporal confounders. The results of this model are in Supplementary Tables S8, S9 and S10 for the eastern, western-central and contiguous United States. Our findings consistently demonstrate a significant positive effect size for "deaths in social proximity". For a detailed explanation of the two-way fixed effect model, refer to Supplementary Information section S4.4.

Figure 4 shows the confidence intervals (CI) for the cluster-robust standard error linear model, network autocorrelation, spatial autocorrelation, and the two-way, fixed-effects models. We consistently observe a positive and significant coefficient for s_{-i}, indicating the effect of social influence on the spread of OOD, while the effect size for d_{-i} has a varying CI, changes sign, and is not always significant. We also observe cluster-robust linear regression has a broader confidence interval compared to the other models, it is primarily because cluster-robust standard errors adjust for the intra-cluster correlation by accounting for the fact that there is less independent information than the total number of observations suggests. This adjustment often results in larger standard errors compared to conventional ones, reflecting the reduced amount of independent information. There is also a loss of statistical power when doing this analysis, as the effective sample size becomes closer to the number of clusters rather than the total number of observations. However, despite accounting for intra-cluster correlation, we observe a statistically significant coefficient for "deaths in social proximity". A key takeaway from this result is that social connections are predominantly more significant than the effect of spatial proximity on OOD rates. This robustness check adds to the consistency of our statistical evidence for the size of the effect of "deaths in social proximity". In our analysis, given the network and spatial configurations inherent in s_{-i} and d_{-i}, we suspected and addressed the correlated errors that stem from the endogeneity of both variables simultaneously, using a two-stage least squares approach. The results of our implementation are discussed in the Supplementary Information section S4.5 and confirm the positive and significant effect size of our social proximity variable in

the eastern, western-central and entire contiguous United States (Supplementary Figure S3 and Supplementary Table S11). To end our analysis, we show the robustness of the effect of "deaths in social proximity" by accounting for the rate at which spatial adjacency weights decay with increasing distance. By accounting for the "deaths in spatial proximity" decay rate, we allow the effect of distant counties far from the focal county to be more pronounced. We use cluster-robust linear regression to test the significance of the coefficient s_{-i}. We observe a statistically significant effect for the coefficient of "deaths in social proximity" between counties in the eastern, western-central, and contiguous United States. The results of this implementation are discussed in Supplementary Information section S5. The confidence interval plot for the model is illustrated in Supplementary Figure S4 and the effect sizes are shown in Supplementary Tables S12, S13 and S14.

Discussion

Our research underscores that social influence, measured by the coefficient of s_{-i}, exhibits a statistically significant impact on the spatial spread of OOD rates in US counties. This finding paves the way for a more in-depth exploration of the mechanisms driving the opioid epidemic, specifically within the social contagion framework. Harmon et al.[28] and Braha and Aguiar[12] have utilized complex generative modeling to understand social contagion in the context of economic crises and voting behavior. Building on their work in modeling social contagion, future studies can adopt a similar framework to model peer influence in the opioid epidemic using SCI data.

Our analysis demonstrates that deaths in social proximity, as measured by the SCI, are associated with opioid overdose deaths in counties in the United States. Although previous studies recognize the tendency for OODs to occur in isolated spaces[29,30], data from the CDC's State Unintentional Drug Overdose Reporting System (SUDORS) (https://www.cdc.gov/overdose-prevention/data-research/facts-stats/sudors-dashboard-fatal-overdose-data.html) indicate that almost half of overdose deaths had a potential bystander present, suggesting that the occurrence of OODs in isolation may be less pronounced than previously thought.

Furthermore, our measure of social influence is based on the social connections derived from the social network structure of Facebook's SCI across US counties and serves as a proxy for aggregated friendship network structures and their association with OODs; however, it does not capture the strength of individual friendship ties among Facebook users in the counties. Focusing on these social connections and their relation to the spatial patterns of OODs, such as geographic discordance (the community in which overdose death occurs being different from the community of residence), can provide novel insights into the complex interplay of social and spatial factors in perpetuating this public health crisis.

In addressing the opioid epidemic, accounting for social networks is crucial in understanding and predicting OOD patterns at the community level to aid policy intervention. Studies using complex generative models with spatial clustering have yielded valuable insight into the dynamics of social influence and can be crucial to designing targeted interventions. For example, Braha and Aguiar[12] use such models to identify spatial clusters of voting contagion. In the context of the opioid epidemic, Liao et al.[31] introduce the Spatio-TEMporal Mutually Exciting Point Process (STEMMED) model to quantify the interconnections between historical and future events, reflecting space-time clustering in OODs across various communities. This methodological approach can improve the prediction of OOD trends within localized settings by modeling unique community-specific OOD event streams, considering spatial and social influences. Our research complements this perspective by bridging the gap in understanding the role of social influence within these dynamics. Although STEMMED captures the spatiotemporal clustering of OOD events, our approach highlights the importance of social structures in shaping these patterns. By estimating the coefficient of the socially lagged variable, we measure the effect of social interactions in different counties in the US, which in conjunction with the spatial focus of Liao et al.[31] can offer a more comprehensive view of the factors driving OOD clusters. For policymakers, this comprehensive approach can provide a solid foundation for designing targeted interventions that address both the spatial and social dimensions of the opioid epidemic.

One of our motivations for this study was to facilitate the creation of proxy networks for social interactions in agent-based models of the US population based on SCI data. Agent-based models with SCI-calibrated social networks can provide valuable information about peer effects in the initiation of opioid use and the development of use disorder. When used in conjunction with powerful agent-based modeling tools such as FRED (Framework for Reconstructing Epidemic Dynamics)[32], such models can enable epidemiologists and policymakers to simulate the spread of the opioid epidemic in great mechanistic detail and evaluate interventions. Chu et al.[33] use FRED to study social contagion in the use of e-cigarettes, showing the potential of FRED to simulate social contagion of addictive behaviors. We expect that SCI-calibrated social networks, combined with agent-based models of OUD progression in FRED (cf., e.g.,[34]), can reveal important mechanistic details about the opioid epidemic in the US. Our results may be particularly useful for calibrating social contagion parameters. The contagion parameters can be tuned so that they reflect the same association between "deaths in social proximity" and OODs as we measure in the mortality data.

SCI can also help design and evaluate intervention strategies based on social networks. Macmadu et al.[35] highlight the need for interventions tailored to social networks. Their research shows that after an overdose, network members exhibit observable behavioral changes, including increased risk taking to manage feelings of bereavement and trauma, protective actions, and some cases showing no change in drug use behavior. Based on these findings, opioid reduction efforts can be optimized in areas with greater social connectivity. Measurement of the effects of social networks across locations can also help optimize the allocation of limited medications, such as naloxone, buprenorphine, and methadone. In all these cases, social network-based strategies offer important opportunities for policy makers to mitigate the spread of the opioid epidemic.

We acknowledge the limitations of using mortality data to measure the spread of this critical health crisis and propose to strengthen these findings with other OUD-related outcomes. Details on the limitations of the

mortality data are discussed in the "Methods". Future work can use SCI to investigate the social contagion of the opioid epidemic in mechanistic agent-based models and structural causal models for causal discovery and inference. We also plan to use SCI to improve our predictive accuracy in detecting OOD hotspots in the US. We have used SCI to measure the dynamics of social networks and their effect on the opioid epidemic. Our choice of SCI is guided by existing literature and our objective to study network effects at the population scale. We recognize the limitations of using SCI because Facebook users are not the same as the general population. SCI only acts as a proxy for real-life friendship networks, and it may not be a true representative of real-life friendship connections and their association with opioid use behavior. Other datasets that capture real-world interactions within a social network framework can also be valuable to understand the effects of social networks in the opioid epidemic. For example, the Add Health dataset (https://data.cpc.unc.edu/projects/2/view#public_li), which includes information on best friends during high school and opioid misuse in adulthood (14 years later), offers an alternative means of measuring these effects. Although this dataset allows analysis within the age-specific bracket of 25 to 35 years, potential problems of population selection bias may limit the generalizability of the findings to the broader population.

Methods

Data pipeline and preprocessing

Our data sources are shown in Fig. 5 and are explained in detail in the following paragraphs.

Mortality data and census demographics. We measure the OOD rate from mortality data obtained from the National Center for Health Statistics (NCHS) for the years 2018–2019. This data set includes demographic details of individuals who have lost their lives to opioid-related overdoses. To first identify overdose related deaths we utilise the following International Classification of Disease (ICD) codes "X40","X41", "X42", "X43", "X44","X60", "X61", "X62","X63", "X64", "X85", "Y10", "Y11", "Y12", "Y13", "Y14". The X and Y codes provide information about deaths that have occurred due to substance overdose. Furthermore, to specifically target opioid overdose deaths we use the ICD T codes "T400", "T401", "T402", "T403", "T404", and "T406". The T codes determine the cause of death in the specification of opioids from other substances. Despite the comprehensive nature of the NCHS mortality data and the use of specific ICD codes to identify opioid overdose deaths, there are important limitations associated with this dataset. Death certificates may not always specify the drugs involved in an overdose, and some overdose deaths involve multiple drugs, making it difficult to determine which substance was primarily responsible. In addition, the analysis also incorporates demographic data on the broader population of the United States of America. The data is stratified at the county levels. To extract this information, we utilize the R package "tidycensus" to systematically retrieve data from data.census.gov.

Clinical and mental health covariates. Clinical factors such as the opioid dispensing rate (ODR), availability of naloxone, and the access to buprenorphine for the treatment of opioid use disorder are used as controls in our regression. Morgan et al.[36] underscore the role of naloxone in reducing opioid-related harm, while Pendergrass
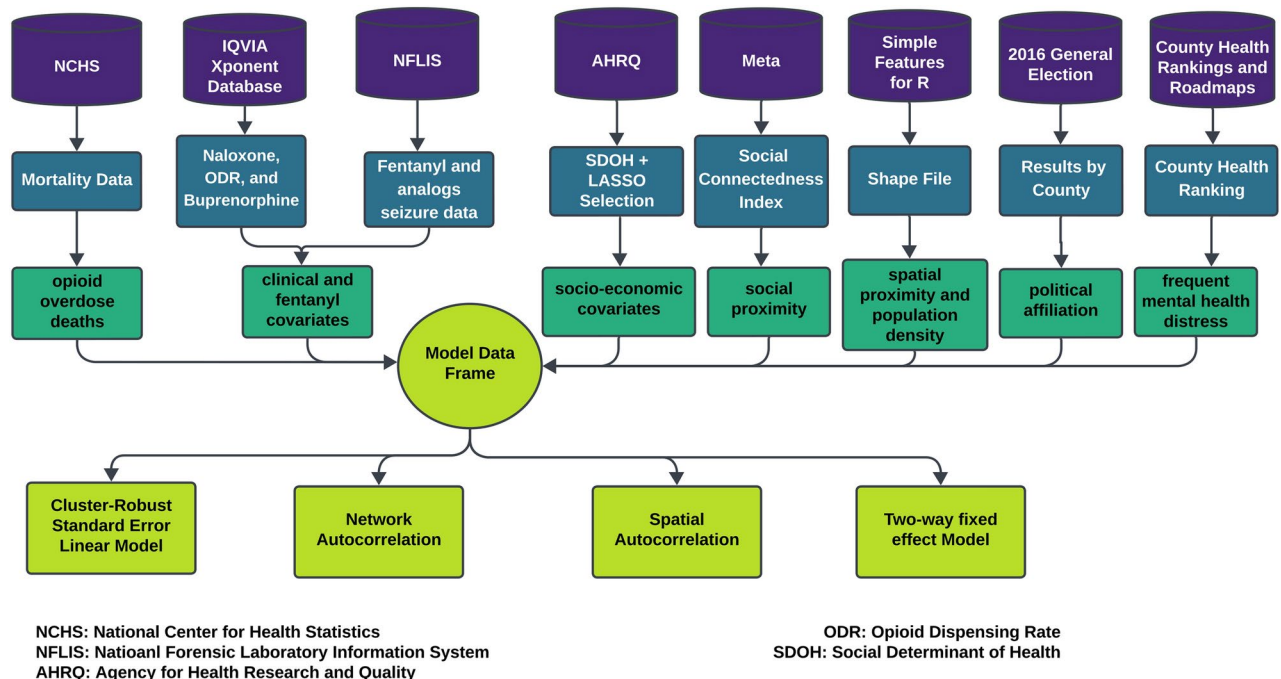

Fig. 5. The diagram depicts the data pipeline for our analysis, including the data streams for the primary variable of interests, s_{-i} and d_{-i}, as well as the relevant socioeconomic and clinical covariates. It also outlines our regression models for estimating the effect of peer influence as measured by SCI on county-level OOD outcomes.

et al.[37] emphasize the importance of buprenorphine availability in mitigating overdose fatalities. The data source for clinical factors such as naloxone, buprenorphine, and ODR is the IQVIA Xponent database, which provides the number of prescriptions for naloxone and buprenorphine distributed throughout counties in the US through retail pharmacy channels. The ODR, which captures the total number of opioid doses dispensed, is measured as morphine milligram equivalents (MME). The current wave of the opioid epidemic from 2013 to the present has witnessed an increase in the illicit use of synthetic opioids such as fentanyl. Furthermore, Kuehn[38] discusses the impact of fentanyl on overdose deaths, particularly among adolescents. Pergolizzi et al.[39] also discuss the role of fentanyl in exacerbating the opioid epidemic. To control for the supply of illegal substances that contribute to OODs, we incorporate state-level fentanyl and analog seizure data from the National Forensic Laboratory Information System as a control variable. Our selection of clinical and illicit-supply covariates is comprehensive based on the data sources available to us. In addition, several studies have shown associations between mental health and opioid overdose deaths[40–42]. Thus, we used data on frequent mental health distress from County Health Rankings and Roadmaps (CHRR). Frequent mental distress is the percentage of adults who reported poor mental health for more than 14 days in response to the question "How many days during the past 30 days was your mental health not good?"[43]. We use this measure to control for the effect of mental health-related issues in counties.

Population density and political affiliation. We also control for population density and political affiliation that are identified in the literature on the opioid epidemic and the structure of social networks. To account for the heterogeneity associated with SCI and opioid use among populations residing in urban and rural counties, we include population density in our regression model[15,44]. The risk status of opioid misuse is also associated with the political affiliations and liberal status of states[45]. Therefore, we control for the effect of political affiliation in our regression by accounting for counties' political leanings using 2016 general election data at the county level.

SCI data. SCI is available at the ZIP code and county levels in the United States. We chose counties as our analysis unit instead of ZIP codes because the latter had significant limitations. A considerable number of ZIP codes have missing SCI data. Typically, ZIP codes without SCI data are those with low populations or those designated exclusively for institutions. Institutional ZIP codes are assigned to areas predominantly occupied by specific institutions, such as hospitals, universities, or military bases. These institutional ZIP codes can also introduce spatial bias; for example, ZIP codes with hospitals are more likely to report higher overdose death rates. To ensure a more continuous and representative spatial framework, we use counties as our unit of analysis. This approach mitigates the issues of missing data and spatial bias, providing a more robust basis for analysis.

Social determinants of health. Socioeconomic factors shape social structures and ties; therefore, incorporating these factors is essential to interpret the influence of social networks on OOD. Shared socioeconomic conditions can also influence behaviors that mirror peer influence, and therefore, we include socioeconomic covariates as controls in our analysis to mitigate the risk of bias due to missing confounders when estimating the effect size of our deaths in the social proximity variable (s_{-i}). Social determinants of health (SDOH) encompass aspects of physical infrastructure, economic context, healthcare context, and social environment of counties. Our selection of SDOH covariates is based on socioeconomic predictors of the opioid epidemic. Liu et al.[46] demonstrate the effects of SDOH measures on drug overdose death locations. Therefore, we use SDOH variables to control for socioeconomic factors that may confound the impact of s_{-i} on county-level OOD rates. These covariates are selected from a pool of SDOH variables using the Least Absolute Shrinkage and Selection Operator (LASSO) technique to avoid multicollinearity issues. We select a subset of covariates from an array of 17 SDOH variables that are listed in Supplementary Table S1. We provide details of the LASSO selection process in the Supplementary Information section S3.

Statistical models. The model coefficients are evaluated for statistical significance at $p < 0.05$ level. Using county-level opioid overdose death rates as the outcome variable, we set up regression models to test the significance of deaths in social proximity to explain the county-level death rate after controlling for deaths in spatial proximity, clinical (mental health, availability of naloxone and buprenorphine and opioid dispensing rates in pharmacies) and fentanyl covariates, and socioeconomic covariates selected using LASSO from SDOH variables, as well as population density and political affiliation covariates. Linear regression and cluster-robust linear regression are analyzed for statistical significance (details in Supplementary Information section S4.1). The residuals in the regression are weighted by population size of the counties, so that the inference is reflecting the population residing in the counties across the US. To account for correlated error terms from the spatial and network structure of the variables s_{-i} and d_{-i}, respectively, we use network and spatial autocorrelation models (details in Supplementary Information sections S4.2 and S4.3). To account for unobserved space- and time-invariant characteristics that might be associated with the covariates in our regression, we use a two-way fixed-effects model. This model includes fixed temporal effects for 2018 and 2019 and fixed spatial effects for states in the eastern United States, the western and central United States, and the contiguous United States (details in Supplementary Information section S4.4). As a robustness check, we also use a two-stage least squares regression to account for correlated error terms that arise from the simultaneous estimation of effects from s_{-i} and d_{-i} (details in the Supplementary Information section S4.5). As a final robustness check, we use distance decay weight to account for the effect of faraway counties from the focal county (details in the Supplementary Information section S5).

Data availability

Mortality data was obtained from the National Center for Health Statistics (NCHS). Due to confidentiality concerns, this data set is not publicly accessible, but can be requested from NCHS at https://www.cdc.gov/nchs/nvss/nvss-restricted-data.htm. The clinical covariates were sourced from the IQVIA Xponent database, which is also not publicly available. Access can be requested through IQVIA at https://www.iqvia.com/insights/the-iqvia-institute/available-iqvia-data. Data on illicit fentanyl-related drugs were obtained from the National Forensic

Laboratory Information System (NFLIS) and can be accessed at https://www.nflis.deadiversion.usdoj.gov/. Data on frequent mental health distress are obtained from the County Health Rankings and Roadmaps (CHRR) at https://www.countyhealthrankings.org/sites/default/files/media/document/analytic_data2019. The 2016 General Election data can be accessed at https://raw.githubusercontent.com/tonmcg/US_County_Level_Election_Results_08-20/master/2016_US_County_Level_Presidential_Results. The social determinants of health covariates are available from the Agency for Healthcare Research and Quality (AHRQ) at https://www.ahrq.gov/sdoh/data-analytics/sdoh-data.html. The SCI data can be accessed using the Facebook Data for Good tools at https://dataforgood.facebook.com/dfg/tools/social-connectedness-index. Data on drug overdose deaths used in our LASSO covariate selection are available from the Bureau of Health Statistics and Registries of the Department of Health of the Commonwealth of Pennsylvania and can be requested as follows: https://www.pa.gov/content/dam/copapwp-pagov/en/health/documents/topics/documents/reporting-registries/vr-govt-researchers/Application%20for%20Access%20to%20Protected%20Data%20for%20Government%20Agencies%20-%20Request%20Form.doc. No new data has been collected in this study. All data availability questions can be addressed to K.T. (kut20@pitt.edu) or M.A.R. (rahimian@pitt.edu).

Code availability

Analysis code to reproduce figures and tables in the paper is available at https://github.com/kut97/ood-sci.

References

1. Jalal, H. et al. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science* **361**(6408), eaau1184 (2018).
2. Jalal, H. & Burke, D. S. Exponential growth of drug overdose poisoning and opportunities for intervention. *Addiction* **117**(5), 1200–1202 (2022).
3. Alpert, A., Evans, W. N., Lieber, E. M. & Powell, D. Origins of the opioid crisis and its enduring impacts. *Q. J. Econ.* **137**(2), 1139–1179 (2022).
4. Lim, T. Y. et al. Modeling the evolution of the US opioid crisis for national policy development. *Proc. Natl. Acad. Sci.* **119**(23), e2115714119 (2022).
5. Costello, B. J., Anderson, B. J. & Stein, M. Peer influence in initiation to heroin use. *J. Drug Issues* **51**(2), 323–339 (2021).
6. Rigg, K. K., McLean, K., Monnat, S. M., Sterner, G. E. III. & Verdery, A. M. Opioid misuse initiation: implications for intervention. *J. Addict. Dis.* **37**(3–4), 111–122 (2018).
7. Guarino, H., Mateu-Gelabert, P., Teubl, J. & Goodbody, E. Young adults' opioid use trajectories: From nonmedical prescription opioid use to heroin, drug injection, drug treatment and overdose. *Addict. Behav.* **86**, 118–123 (2018).
8. Jalal, H., Buchanich, J. M., Sinclair, D. R., Roberts, M. S. & Burke, D. S. Age and generational patterns of overdose death risk from opioids and other drugs. *Nat. Med.* **26**(5), 699–704 (2020).
9. Syvertsen, J. L., Paquette, C. E. & Pollini, R. A. Down in the valley: Trajectories of injection initiation among young injectors in california's central valley. *Int. J. Drug Policy* **44**, 41–49 (2017).
10. Valente, T. W. et al. Peer acceleration: effects of a social network tailored substance abuse prevention program among high-risk adolescents. *Addiction* **102**(11), 1804–1815 (2007).
11. Stringfellow, E. J. et al. Reducing opioid use disorder and overdose deaths in the United States: a dynamic modeling analysis. *Sci. Adv.* **8**(25), eabm8147 (2022).
12. Braha, D. & De Aguiar, M. A. Voting contagion: modeling and analysis of a century of US presidential elections. *PLoS ONE* **12**(5), e0177970 (2017).
13. Homer, J. & Wakeland, W. A dynamic model of the opioid drug epidemic with implications for policy. *Am. J. Drug Alcohol Abuse* **47**(1), 5–15 (2021).
14. Bailey, M., Cao, R., Kuchler, T., Stroebel, J. & Wong, A. Social connectedness: measurement, determinants, and effects. *J. Econ. Perspect.* **32**(3), 259–80 (2018).
15. Bailey, M., Farrell, P., Kuchler, T. & Stroebel, J. Social connectedness in urban areas. *J. Urban Econ.* **118**, 103264 (2020).
16. Bailey, M. et al. The determinants of social connectedness in Europe. In *Social Informatics: 12th International Conference, SocInfo 2020, Pisa, Italy, October 6–9, 2020, Proceedings 12* 1–14 (Springer, 2020).
17. Coven, J., Gupta, A. & Yao, I. JUE Insight: Urban flight seeded the COVID-19 pandemic across the United States. *J. Urban Econ.* **133**, 103489 (2023).
18. Kuchler, T., Russel, D. & Stroebel, J. JUE insight: The geographic spread of COVID-19 correlates with the structure of social networks as measured by facebook. *J. Urban Econ.* **127**, 103314 (2022).
19. Ge, Y. et al. Reddit-impacts: a named entity recognition dataset for analyzing clinical and social effects of substance use derived from social media. arXiv preprint arXiv:2405.06145 (2024).
20. Pandrekar, S. et al. Social media based analysis of opioid epidemic using reddit. In *AMIA Annual Symposium Proceedings, vol. 2018* 867 (American Medical Informatics Association, 2018).
21. Balsamo, D., Bajardi, P., Salomone, A. & Schifanella, R. Patterns of routes of administration and drug tampering for nonmedical opioid consumption: data mining and content analysis of reddit discussions. *J. Med. Internet Res.* **23**(1), e21212 (2021).
22. Diemer, A. & Regan, T. No inventor is an island: social connectedness and the geography of knowledge flows in the us. *Res. Policy* **51**(2), 104416 (2022).
23. Mäckle, K. & Ruenzi, S. Friends with drugs: The role of social networks in the opioid epidemic. *Available at SSRN* (2022).
24. Cutler, D. M. & Donahoe, J. T. Thick market externalities and the persistence of the opioid epidemic, tech. rep., National Bureau of Economic Research (2024).
25. Forati, A., Ghose, R., Mohebbi, F. & Mantsch, J. R. The journey to overdose: Using spatial social network analysis as a novel framework to study geographic discordance in overdose deaths. *Drug Alcohol Depend.* **245**, 109827 (2023).
26. Koram, N. et al. Role of social network dimensions in the transition to injection drug use: actions speak louder than words. *AIDS Behav.* **15**, 1579–1588 (2011).
27. Xu, Y., Santi, P. & Ratti, C. Beyond distance decay: Discover homophily in spatially embedded social networks. *Ann. Am. Assoc. Geogr.* **112**(2), 505–521 (2022).
28. Harmon, D. et al. Anticipating economic market crises using measures of collective panic. *PLoS ONE* **10**(7), e0131871 (2015).

29. Gicquelais, R. E., Genberg, B. L., Maksut, J. L., Bohnert, A. S. & Fernandez, A. C. Prevalence and correlates of using opioids alone among individuals in a residential treatment program in michigan: implications for overdose mortality prevention. *Harm Reduct. J.* **19**(1), 135 (2022).
30. Li, Y., Miller, H. J., Root, E. D., Hyder, A. & Liu, D. Understanding the role of urban social and physical environment in opioid overdose events using found geospatial data. *Health & Place* **75**, 102792 (2022).
31. Liao, C.-Y. et al. Tides need stemmed: A locally operating spatio-temporal mutually exciting point process with dynamic network for improving opioid overdose death prediction. arXiv:2211.07570 (2022).
32. Grefenstette, J. J. et al. Fred (a framework for reconstructing epidemic dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC Public Health* **13**, 1–14 (2013).
33. Chu, K.-H. et al. Integrating social dynamics into modeling cigarette and e-cigarette use. *Health Educ. Behav.* **47**(2), 191–201 (2020).
34. Ahmed, A. A., Rahimian, M. A. & Roberts, M. S. Estimating treatment effects using costly simulation samples from a population-scale model of opioid use disorder. In *2023 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)* 1–4 (IEEE, 2023).
35. Macmadu, A. et al. Drug use behaviors, trauma, and emotional affect following the overdose of a social network member: a qualitative investigation. *Int. J. Drug Policy* **107**, 103792 (2022).
36. Morgan, J. & Jones, A. L. The role of naloxone in the opioid crisis. *Toxicol. Commun.* **2**(1), 15–18 (2018).
37. Pendergrass, S. A., Crist, R. C., Jones, L. K., Hoch, J. R. & Berrettini, W. H. The importance of buprenorphine research in the opioid crisis. *Mol. Psychiatry* **24**(5), 626–632 (2019).
38. Kuehn, B. M. Fentanyl drives startling increases in adolescent overdose deaths. *JAMA* **329**(4), 280–281 (2023).
39. Pergolizzi, J. V., LeQuang, J. A., Taylor, R., Raffa, R. B. & Group, N. R. Going beyond prescription pain relievers to understand the opioid epidemic: the role of illicit fentanyl, new psychoactive substances, and street heroin. *Postgrad. Med.* **130**(1), 1–8 (2018).
40. Dinwiddie, A. T. Reported non-substance-related mental health disorders among persons who died of drug overdose-United States. *MMWR Morbid. Mortal. Week. Rep.* **2024**, 73 (2024).
41. Santo, T. Jr. et al. Prevalence of mental disorders among people with opioid use disorder: a systematic review and meta-analysis. *Drug Alcohol Depend.* **238**, 109551 (2022).
42. van Draanen, J. et al. Mental disorder and opioid overdose: a systematic review. *Soc. Psychiatry Psychiatr. Epidemiol.* **1–25**, 2022 (2022).
43. Yang, T.-C., Matthews, S. A., Sun, F. & Armendariz, M. Modeling the importance of within- and between-county effects in an ecological study of the association between social capital and mental distress. *Prev. Chronic Dis.* **16**, E75 (2019).
44. Peters, D. J., Monnat, S. M., Hochstetler, A. L. & Berg, M. T. The opioid hydra: Understanding overdose mortality epidemics and syndemics across the rural-urban continuum. *Rural. Sociol.* **85**(3), 589–622 (2020).
45. Haffajee, R. L., Lin, L. A., Bohnert, A. S. & Goldstick, J. E. Characteristics of US counties with high opioid overdose mortality and low capacity to deliver medications for opioid use disorder. *JAMA Netw. Open* **2**(6), e196373–e196373 (2019).
46. Liu, M. et al. Geo-spatial risk factor analysis for drug overdose death in south florida from 2014 to 2019, and the independent contribution of social determinants of health. *Drug Alcohol Depend.* **248**, 109931 (2023).

Acknowledgements

Author contributions

K.T., M.A.R., M.S.R., P.K. and J.M.B. conceived the research. Author K.T. led the project with supervision from M.A.R. and M.S.R. Authors M.S.R., P.K. and J.M.B. contributed to data acquisition pipeline. Author K.T. managed data processing with input from M.A.R., M.S.R., P.K. and J.M.B. Authors K.T. and M.A.R. led the analysis with input from all the authors. Author K.T. wrote the analysis code with input from M.A.R. Authors K.T. and M.A.R. led the writing of the paper, with input from all the authors. All authors approved the final paper.

Competing interests

K.T. has been an associate specialist at Merck Sharp & Dohme LLC from May to August 2023. M.A.R. has served on the advisory committee of a vaccine confidence fund created by Meta and Merck; some of his research has also been funded by Meta. The other authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-80627-4.

Correspondence and requests for materials should be addressed to K.T. or M.A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.