The Annals of Applied Statistics
2021, Vol. 15, No. 1, 88–101
https://doi.org/10.1214/20-AOAS1384

POINT PROCESS MODELING OF DRUG OVERDOSES WITH HETEROGENEOUS AND MISSING DATA

BY XUEYING LIU[1,*], JEREMY CARTER[2,‡], BRAD RAY[2,§] AND GEORGE MOHLER[1,†]

[1]*Department of Computer and Information Science, Indiana University–Purdue University Indianapolis,* [*]*xl17@iu.edu;* [†]*gmohler@iupui.edu*

[2]*O'Neill School of Public and Environmental Affairs, Indiana University–Purdue University Indianapolis,* [‡]*carterjg@iupui.edu;* [§]*bradray@iupui.edu*

Opioid overdose rates have increased in the United States over the past decade and reflect a major public health crisis. Modeling and prediction of drug and opioid hotspots, where a high percentage of events fall in a small percentage of space–time, could help better focus limited social and health services. In this work we present a spatial-temporal point process model for drug overdose clustering. The data input into the model comes from two heterogeneous sources: (1) high volume emergency medical calls for service (EMS) records containing location and time but no information on the type of nonfatal overdose, and (2) fatal overdose toxicology reports from the coroner containing location and high-dimensional information from the toxicology screen on the drugs present at the time of death. We first use nonnegative matrix factorization to cluster toxicology reports into drug overdose categories, and we then develop an EM algorithm for integrating the two heterogeneous data sets, where the mark corresponding to overdose category is inferred for the EMS data and the high volume EMS data is used to more accurately predict drug overdose death hotspots. We apply the algorithm to drug overdose data from Indianapolis, showing that the point process defined on the integrated data out-performs point processes that use only coroner data (AUC improvement 0.81 to 0.85). We also investigate the extent to which overdoses are contagious, as a function of the type of overdose, while controlling for exogenous fluctuations in the background rate that might also contribute to clustering. We find that drug and opioid overdose deaths exhibit significant excitation with branching ratio ranging from 0.72 to 0.98.

1. Introduction. Over 500,000 drug overdose deaths have occurred in the United States since 2000, and over 70,000 of these deaths occurred in 2017 (Seth et al. (2018)). Opioids are a leading cause in these deaths, and these trends are characterized by three distinct time periods (Ciccarone (2017)). In the 1990s, overdose deaths were driven by prescription opioid-related deaths (Cicero et al. (2014)), whereas reduced availability of prescriptions led to an increase of heroin-related deaths beginning in the 2010s (Cicero et al. (2014), Rudd et al. (2014), Strickler et al. (2019)). Illicit fentanyl, a synthetic opioid 50 to 100 times more potent than morphine (Gladden, Martinez and Seth (2016)), has become a major cause of opioid-related deaths since around 2013. It is estimated that in 2016 around half of opioid-related deaths contained fentanyl (Jones, Einstein and Compton (2018)), and fentanyl mixed into heroin and cocaine is likely contributing to many of these overdose deaths (Jones, Baldwin and Compton (2017), Kandel et al. (2017)).

Criminology and public health disciplines have leveraged spatiotemporal event modeling in attempts to predict social harm for effective interventions (Mohler, Carter and Raje (2018), Tsui et al. (2011), Yu et al. (2007)). Fifty percent of crime has been shown to concentrate

Received November 2019; revised July 2020.
Key words and phrases. Point process, expectation maximization algorithm, semisupervised learning, nonnegative matrix factorization, opioid overdose.

within just five percent of an urban geography (Weisburd (2015)). Geographic concentrations of drug-related emergency medical calls for service (Hibdon and Groff (2014)), drug activity (Hibdon, Telep and Groff (2017)) and opioid overdose deaths mirror those of crime (Carter, Mohler and Ray (2018)). In particular, over half of opioid overdose deaths in Indianapolis occur in less than 5% of the city (Carter, Mohler and Ray (2018)).

Patterns of repeat and near-repeat crime in space and time further suggest that not only does crime concentrate in place but also that such events are an artifact of a contagion effect resulting from an initiating criminal event (Townsley, Homel and Chaseling (2003)). Similar observations have also explained the diffusion of homicide events (Zeoli et al. (2014)). Experiments of predictive policing models, using spatiotemporal Hawkes and self-exciting point processes, demonstrates that such empirical realities can be harnessed to direct police resources to reduce crime (Mohler et al. (2015)). Thus, the interdependence and chronological occurrence of event types in crime and public health lend promise to how to best predict other social harm events, such as opioid overdoses.

In this work we consider the modeling of two datasets of space–time drug and opioid overdose events in Indianapolis. The first dataset consists of emergency medical calls for service (EMS) events. These events are nonfatal overdoses and include a date, time and location but no information on the cause of the overdose. The second dataset consists of overdose death events (including location) and are accompanied by a toxicology report that screens for substances present or absent in the overdose event. We develop a marked point process model for the heterogeneous dataset that uses nonnegative matrix factorization to reduce the dimension of the toxicology reports to several categories. We then use an expectation-maximization algorithm to jointly estimate model parameters of a Hawkes process and, simultaneously, infer the missing overdose category for the nonfatal overdose EMS data.

We show that the point process defined on the integrated, heterogeneous data outperforms point processes that use only homogeneous coroner data. We also investigate the extent to which overdoses are contagious, as a function of the type of overdose, while controlling for exogenous fluctuations in the background rate that might also contribute to clustering. We find that opioid overdose deaths exhibit significant excitation with branching ratio ranging from 0.72 to 0.98.

The outline of the paper is as follows. In Section 2 we provide an overview of our modeling framework. In Section 3 we run several experiments on synthetic data to validate the model and also on Indianapolis drug overdose data to demonstrate model accuracy on the application. We discuss several policy implications and directions for future research in Section 4.

2. Methods.

2.1. *Self-exciting point processes.* In this work we consider a self-exciting point process of the form (Mohler et al. (2011a))

$$\lambda(x, y, t) = \mu_0 \nu(t) u(x, y) + \sum_{i:t_i<t} g(x - x_i, y - y_i, t - t_i), \tag{1}$$

where $g(x, y, t)$ is a triggering kernel modeling the extent to which risk following an event increases and spreads in space and time. The background Poisson process modeling spontaneous events is assumed separable in space and time, where $u(x, y)$ models spatial variation in the background rate and $\nu(t)$ may reflect temporal variation arising from time of day, weather, seasonality, etc. The point process may be viewed as a branching process (or superposition of Poisson processes), where the background Poisson process with intensity $\mu_0 \nu(t) u(x, y)$ yields the first generation, and then each event (x_i, y_i, t_i) triggers a new generation, according to the Poisson process $g(x - x_i, y - y_i, t - t_i)$.

We allow for self-excitation in the model to capture spatiotemporal clustering of overdoses present in the data. For example, a particular supply of heroin may contain an unusually high amount of fentanyl, leading to a cluster of overdoses in a neighborhood where the drug is sold and within a short time period.

Model (1) can be estimated via an expectation-maximization algorithm (Mohler et al. (2011b), Veen and Schoenberg (2008)), leveraging the branching process representation of the model. Let L be a matrix where $l_{ij} = 1$ if event i is triggered by event j in the branching process and $l_{ii} = 1$ if event i is a spontaneous event from the background process. Then, the complete data log-likelihood is given by

$$\sum_i l_{ii} \log\big(\mu_0 v(t_i) u(x_i, y_i)\big) - \int \mu_0 v(t) u(x, y,) \, dx \, dy \, dt \tag{2}$$

$$+ \sum_{ij} l_{ij} \log\big(g(x_i - x_j, y_i - y_j, t_i - t_j)\big) \tag{3}$$

$$- \sum_j \int g(x - x_j, y - y_j, t - t_j) \, dx \, dy \, dt. \tag{4}$$

Thus, estimation decouples into two density estimation problems, one for the background intensity and one for the triggering kernel. Because the complete data is not observed, we introduce a matrix P with entries p_{ij} representing the probability that event i is triggered by event j.

Given an initial guess P_0 of matrix P, a nonparametric density estimation procedure can be used to estimate u and v from $\{t_k, x_k, y_k, p_{kk}\}_{k=1}^N$, providing estimates u_0, v_0 in the maximization step of the algorithm.

More specifically, we estimate u and v using leave-one-out kernel density estimation,

$$\begin{aligned} v(t_i) &= \frac{1}{N_b} \sum_{i \neq j} \frac{p_{jj}}{2\pi {b_1}^2} \exp\left\{-\frac{(t_i - t_j)^2}{2{b_1}^2}\right\}, \\ u(x_i, y_i) &= \frac{1}{N_b} \sum_{i \neq j} \frac{p_{jj}}{2\pi {b_2}^2} \exp\left\{-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2{b_2}^2}\right\}, \end{aligned} \tag{5}$$

where $N_b = \sum_i p_{ii}$ is the estimated number of background events and b_1, b_2 are the kernel bandwidths that can be estimated via cross-validation or based on nearest neighbor distances. Because u and v are chosen to integrate to 1, we then have the ML estimate $\hat{\mu}_0 = N_b$.

We assume the triggering kernel is given by a separable function, that is, exponential in time (Figure 1) with parameter ω and Gaussian in space with parameter σ (Mohler (2014)):

$$g(x, y, t) = K_0\big(w \cdot \exp\{-wt\}\big) \tag{6}$$

$$\cdot \frac{1}{2\pi\sigma^2} \cdot \exp\left\{-\frac{1}{2\sigma^2}\big(x^2 + y^2\big)\right\}. \tag{7}$$

We then obtain an estimate for the parameters using weighted sample averages from the data $\{t_i - t_j, x_i - x_j, y_i - y_j, p_{ij}\}_{t_i > t_j}$,

$$\begin{aligned} \hat{K}_0 &= \sum_{t_i > t_j} p_{ij} \Big/ \sum_{i,j} p_{ij}, \\ \hat{w} &= \sum_{t_i > t_j} p_{ij} \Big/ \sum_{t_i > t_j} p_{ij} \cdot (t_i - t_j), \\ \hat{\sigma} &= \sqrt{\sum_{t_i > t_j} p_{ij} \cdot \big[(x_i - x_j)^2 + (y_i - y_j)^2\big] / 2 \cdot \sum_{t_i > t_j} p_{ij}}. \end{aligned} \tag{8}$$

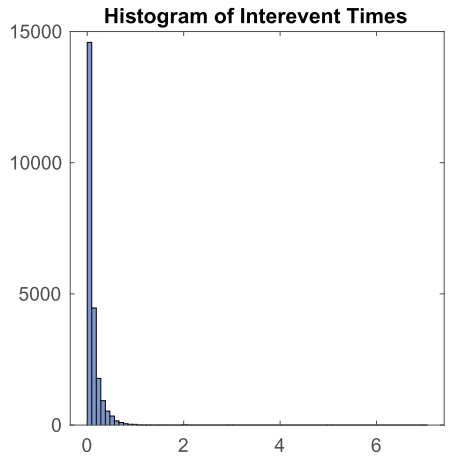

FIG. 1. *Histogram of interevent times of real data, suggests that time triggering function is exponential.*

In the estimation step we estimate the probability that event i is a background event via the formula,

$$p_{ii} = \frac{\mu_0 u(x_i, y_i) v(t_i)}{\lambda(x_i, y_i, t_i)}, \tag{9}$$

and the probability that event i is triggered by event j as,

$$p_{ij} = \frac{g(x_i - x_j, y_i - y_j, t_i - t_j)}{\lambda(x_i, y_i, t_i)}, \tag{10}$$

(Zhuang, Ogata and Vere-Jones (2002)). We then iterate for $n = 1, \ldots, N_{em}$ between the expectation and maximization steps until a convergence criteria is met:

1. Estimate u_n, v_n, and g_n using (5) and (8).
2. Update P_n from u_n, v_n, and g_n using (9) and (10).

2.2. *Modeling with heterogeneous event data.* In this work we assume that we are given two datasets, A and B, though our modeling framework extends more generally to three or more. Event dataset A contains low-dimensional, unmarked space–time events, whereas dataset B contains space–time events with high-dimensional marks. In our application, drug overdoses that do not result in death comprise dataset A, whereas those overdoses that do result in death are accompanied by a high-dimensional mark, namely, the toxicology screen conducted by the coroner. Event dataset B, therefore contains a much smaller number of events compared to A.

Next, we use nonnegative matrix factorization (NMF) (Lee and Seung (2001)) to reduce the dimension of the high-dimensional mark of dataset B into an indicator for K groups. Each toxicology report consists of an indicator (presence or absence) for each one of 133 drugs for which the test screens. These reports then are input into a overdose-drug matrix analogous to a document-term matrix in text analysis using NMF. We then use NMF to factor overdose-drug matrices into the product of two nonnegative matrices, one of them representing the relationship between drugs and topic clusters and the other one representing the relationship between topic clusters and specific overdose events in the latent topic space. The second matrix yields the cluster membership of each event (the cluster is the argmax of the column corresponding to each event).

2.3. *Estimation of a marked point process with missing data.* Merging dataset A and B, we now have marked event data (x_i, y_i, t_i, k_i) where the mark k_i is one of $k = 1, \ldots, K$ clusters and is unknown for event data coming from A but is known for event data from B.

Model (1) can be extended by adding in the group labels

$$\lambda^k(x, y, t) = \mu_0^k u^k(x, y) v^k(t) + \sum_{\substack{i: t_i < t \\ k_i = k}} g^k(x - x_i, y - y_i, t - t_i), \tag{11}$$

where g^k is modeled as follows:

$$\begin{aligned} g^k(x, y, t) &= K_0^k \left(w^k \cdot \exp\{-w^k t\}\right) \\ &\quad \cdot \frac{1}{2\pi {\sigma^k}^2} \cdot \exp\left\{-\frac{1}{2{\sigma^k}^2}\left[x^2 + y^2\right]\right\}. \end{aligned} \tag{12}$$

Here, we assume each cluster k has its own parameters $(\omega^k, \mu_0^k, \sigma^k, K_0^k)$.

We then extend the branching structure matrix P to a set of K matrices, P^k, with initial guess P_0^k and entries:

$$p_{ij}^k = \begin{cases} \dfrac{1}{K} & \text{if } i = j \text{ and event } i \text{ from } A, \\ 1 & \text{if } i = j, \text{ event } i \text{ from } B \text{ and belongs to group } k, \\ 0 & \text{otherwise.} \end{cases}$$

Then, P^k can be updated similarly for each cluster $k = 1, \ldots, K$:

$$p_{ii}^k = \frac{u^k(x_i, y_i) v^k(t_i)}{\lambda^k(x_i, y_i, t_i)} \tag{13}$$

and

$$p_{ij}^k = \frac{g^k(x_i - x_j, y_i - y_j, t_i - t_j)}{\lambda^k(x_i, y_i, t_i)}, \tag{14}$$

where, for each event i from dataset A, we have that $\sum_{k=1}^K (\sum_{t_i \geq t_j} p_{ij}^k) = 1$, and, for event i from dataset B, we have that $p_{ij}^{\tilde{k}} = 0$ for all events j where $t_i \geq t_j$ and $\tilde{k}$ is not the group to which event i belongs.

The parameters are then estimated using P^k,

$$\begin{aligned} K_0^k &= \sum_{t_i > t_j} p_{ij}^k \Big/ \sum_{i,j} p_{ij}^k, \\ w^k &= \sum_{t_i > t_j} p_{ij}^k \Big/ \sum_{t_i > t_j} p_{ij}^k \cdot (t_i - t_j), \\ \sigma^k &= \sqrt{\sum_{t_i > t_j} p_{ij}^k \cdot \left[(x_i - x_j)^2 + (y_i - y_j)^2\right] \Big/ \left(2 \cdot \sum_{t_i > t_j} p_{ij}^k\right)}, \\ \mu_0^k &= \sum p_{ii}^k \end{aligned}$$

and the EM algorithm is iterated to convergence.

3. Results.

3.1. *Synthetic data.* To validate our methodology, we simulate point process data where dataset B has $K = 4$ groups with parameters given by those in Tables 1 and 2. The back-

TABLE 1
Background rates of synthetic data

Group	$bg(1)$	$bg(2)$	$bg(3)$	$bg(4)$
1	0.1	0.2	0.3	0.4
2	0.4	0.3	0.2	0.1
3	0.4	0.4	0.1	0.1
4	0.1	0.4	0.1	0.4

ground rate for each group is heterogeneous in space, with different rates in each quadrant in the unit square and homogeneous in time. Figure 2 and Table 1 illustrate how the background events are simulated: different background rates are assigned to each of the four different regions. Table 2 contains the true parameters for each group.

We then simulate the missing data process by assigning 30% of the data to dataset A (no label) and 70% to B. We find that the EM algorithm detailed above converges within 50 iterations.

We simulate 50 synthetic datasets and then estimate the true parameters, where the results are displayed in Figure 3. In the figure the histograms of w, K_0, σ and μ correspond to the estimates from the EM algorithm, where the red reference lines represent the average of the 50 results and the true value of the parameters are in blue. We find that our model is able to accurately recover both the true parameters and the event cluster membership up to the standard errors of the estimators.

In Table 3 we display the estimated number of events of each group (along with their actual values) when A has 30% of events as well as when 90% of events are assigned to A (and thus unknown). We find in both experiments that the model is able to recover the cluster sizes accurately.

In Figure 4 we compare baseline models estimated only on A or B individually against the combined model. We also analyze the difference in performance vs. the percentage of events assigned to dataset A. Here, we find that the model estimated on both datasets always has higher likelihood than the models estimated only on one dataset.

3.2. *Emergency data and toxicology report.* Next, we analyze a dataset of drug overdose data from Marion County, Indiana (Indianapolis). The data spans the time period from January 14, 2010 to December 30, 2016. The fatal drug overdose dataset with toxicology reports (dataset B) consists of 969 events, and the nonfatal, emergency medical calls for service dataset is 24 times bigger, with 22,049 unlabeled events.

We use NMF, as described above, to cluster the toxicology report data. We use coherence (Stevens et al. (2012)) to select the number of clusters, which we find to be $K = 4$ for our data (see Figure 5). In Table 4 we show the top 24 most frequent drugs, and their frequencies present in the fatal overdose dataset and in Table 5 we display the top five most frequent

TABLE 2
True parameters of synthetic data

Group	w	K_0	σ	μ
1	0.1	0.9	0.01	67
2	0.5	0.8	0.001	28
3	1	0.6	0.02	55
4	0.3	0.75	0.003	132

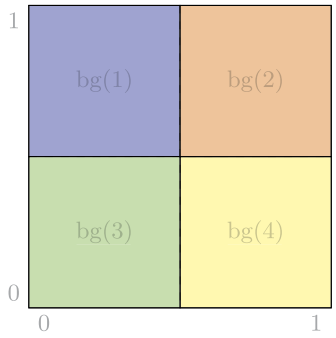

FIG. 2. *Simulation of events' location: for each background event, probabilities of falling in the purple, orange, green and yellow regions are* $bg(1)$, $bg(2)$, $bg(3)$, $bg(4)$, *respectively.*

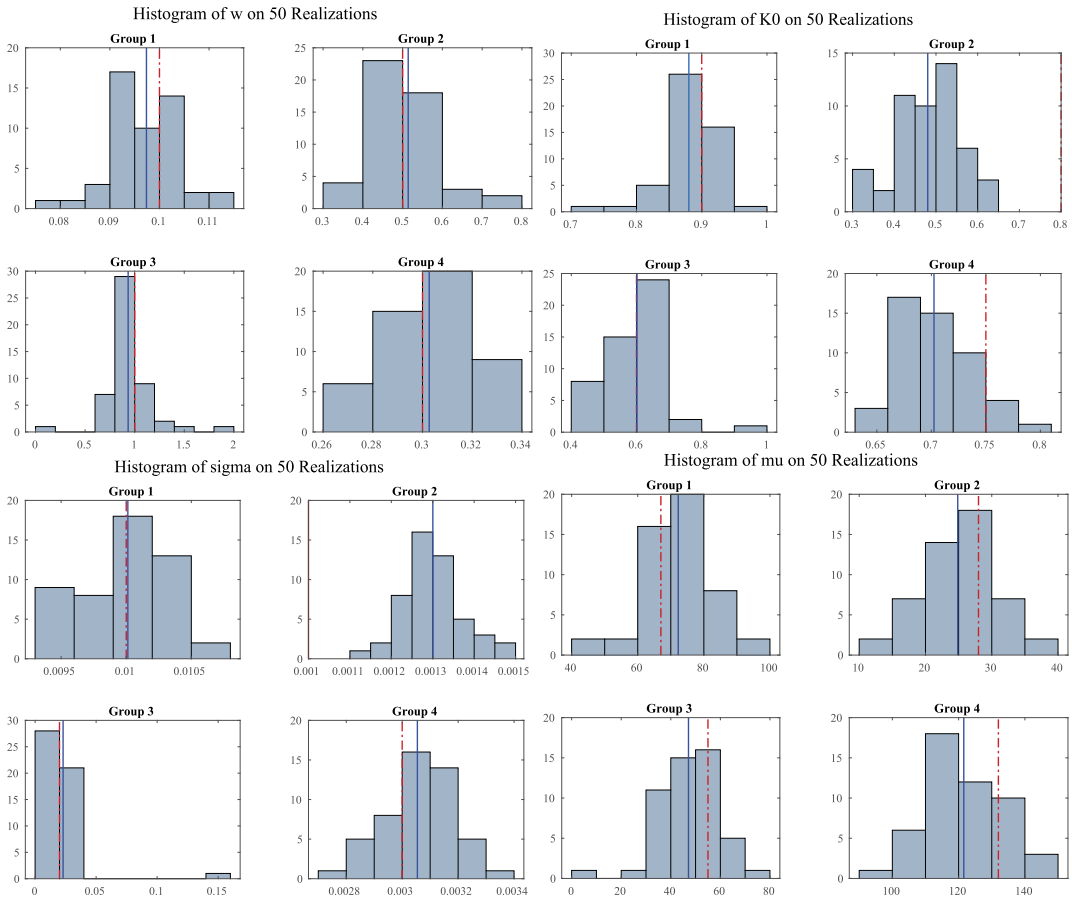

FIG. 3. *Parameters' true value (in red dash-dot line) and average of converged values (in blue solid line).*

TABLE 3

Number of events from each group vs. estimated number while dataset A is 30% (*left*) *and* 90% (*right*) *of all data*

Group	True #	Estimated #
Group 1	570	581
Group 2	154	145
Group 3	173	168
Group 4	431	434

Group	True #	Estimated #
Group 1	1197	1195
Group 2	71	56
Group 3	134	113
Group 4	380	418

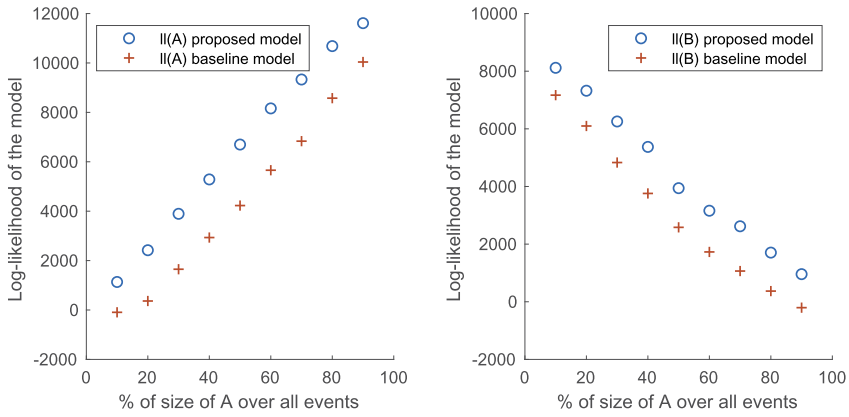

FIG. 4. *Log-likelihood of the model vs. baseline model on individual datasets with different percentage of A. Left: Likelihood evaluated on dataset A. Right: Likelihood evaluated on dataset B.*

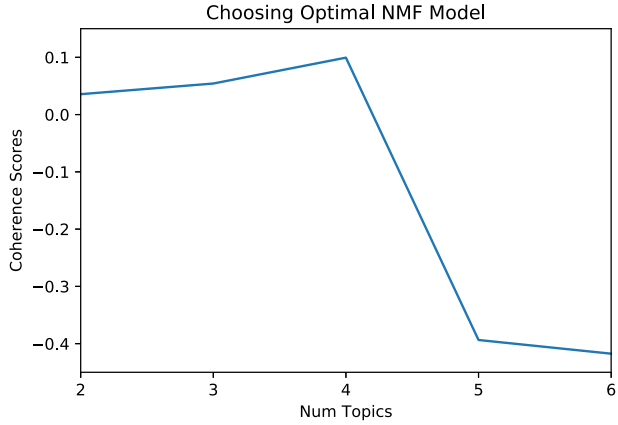

FIG. 5. *NMF coherence scores of drug overdose clusters vs. number of topic clusters K.*

TABLE 4
24 *most frequently present drugs*

Drug	Frequency	Drug	Frequency
Hypnotic	0.9617	11-Nor-9-carboxy-THC	0.8113
Lidocaine	0.5588	11-Hydroxy-THC	0.4856
Phenobarbital	0.4762	Gastrointestinal	0.3841
Eszopiclone	0.3841	THC-Aggregate	0.3580
Promethazine	0.3566	Alcohol	0.2451
Ethanol	0.2451	Opioids	0.2263
Illicit	0.2189	Norfentanyl	0.1773
Amphetamine	0.1760	Acetylfentanyl	0.1605
Fentanyl	0.1571	Acetyl	0.1343
Methamphetamine	0.1162	Morphine	0.1162
Delta-9-THC	0.0907	6-MAM	0.0604
Diazapam	0.0537	THC	0.0524

TABLE 5
Top 5 drugs from each group

Drug	Group 1	Group 2	Group 3	Group 4
1	6-MAM	Benzodiazepine	Ethanol	Fentanyl
2	Heroin	Hydrocodone	Alcohol	Norfentanyl
3	Codeine	Oxycodone	Cocaine	Opioids
4	Morphine	Hydromorphone	Illicits	Amphetamine
5	Illicit	Oxymorphone	Benzodiazepine	Methamph.

drugs found in each NMF group. In Table 5 we find that the first group consists of illicit drugs (6-MAM and heroin), whereas group 2 consists of mostly opioids that can be obtained via a prescription. Group 3 overdoses involve alcohol, whereas group 4 is fentanyl related overdoses.

Next, we fit the point process model to the fatal and nonfatal overdose data. In Figure 6 we plot a heatmap of the inferred background events in space, disaggregated by group, along with the temporal trend of background events in Figure 7. We find that, in time, the frequency of prescription opioid overdoses went down in Indianapolis, whereas illicit opioid overdoses, including the fentanyl group, increased over the same time period. In space the illict drug hotspots are focused downtown, whereas the prescription opioid hotspots are more spread out in the city.

In Table 6 we display the estimated point process parameters. We see that for each group selfexcitation plays a large role, where the branching ratio ranges from 0.72 to 0.98. In Tables 7 and 8 we compare the log-likelihood values of the combined heterogeneous point process to baseline models, estimated only on EMS or overdose death data. Here, we find that including the EMS data improves the AIC values of the model for opioid overdose death, and the overdose death data improves the AIC of the model for EMS events.

To assess the model with a metric that better mirrors how interventions might work, we run the following experiment. For each day from January 15, 2010 to December 30, 2016, we estimate the point process intensity in each of 50×50 grid cells covering Indianapolis. We then rank the cells by the intensity and assign labels for whether an overdose occurs (1) or does not occur (0) during the next day. We then compute the area under the curve (AUC) of this ranking for the baseline and the proposed method. In practice, a point process model could be used to rank the top hotspots where overdoses are likely to occur, and then those areas could be the focus of targeted interventions, such as distribution of naloxone that reverses the effects of an overdose.

In Tables 7 and 8 we find that the AUC of the combined model evaluated on overdose death data is 0.85, compared to 0.81 for the model utilizing only overdose data. However, adding overdose death data to the EMS data impairs the model in terms of AUC. The heterogeneous model has an AUC of 0.72 compared to 0.8 for the EMS data model (though the overdose death data does improve the AIC of the EMS data model).

4. Discussion. Heterogeneous data integration for model improvement promotes several policy and intervention benefits. Research using emergency medical services data has shown that persons who experience repeat nonfatal drug overdoses have a significantly higher mortality rate, as compared to individuals without repeat events (Ray et al. (2018)). As our results suggest, toxicology data can be leveraged to model overdose diffusion across space and time, and diffusion varies across geographies. Taken together, integration of large-scale event data and overdose diffusion can sharpen policy interventions designed to reduce substance abuse

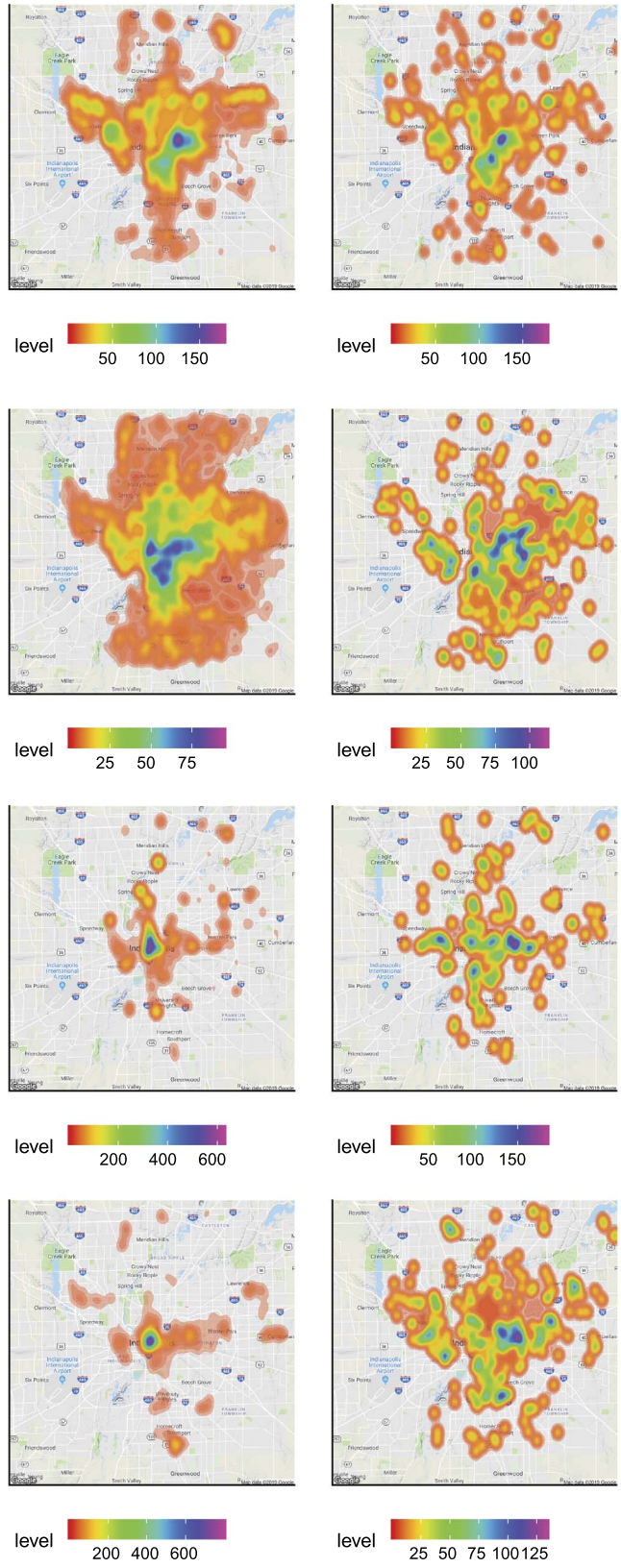

FIG. 6. *Heatmaps of nonfatal overdose events* (*left*) *and fatal overdose events* (*right*). *Top to bottom*: *groups* 1–4.

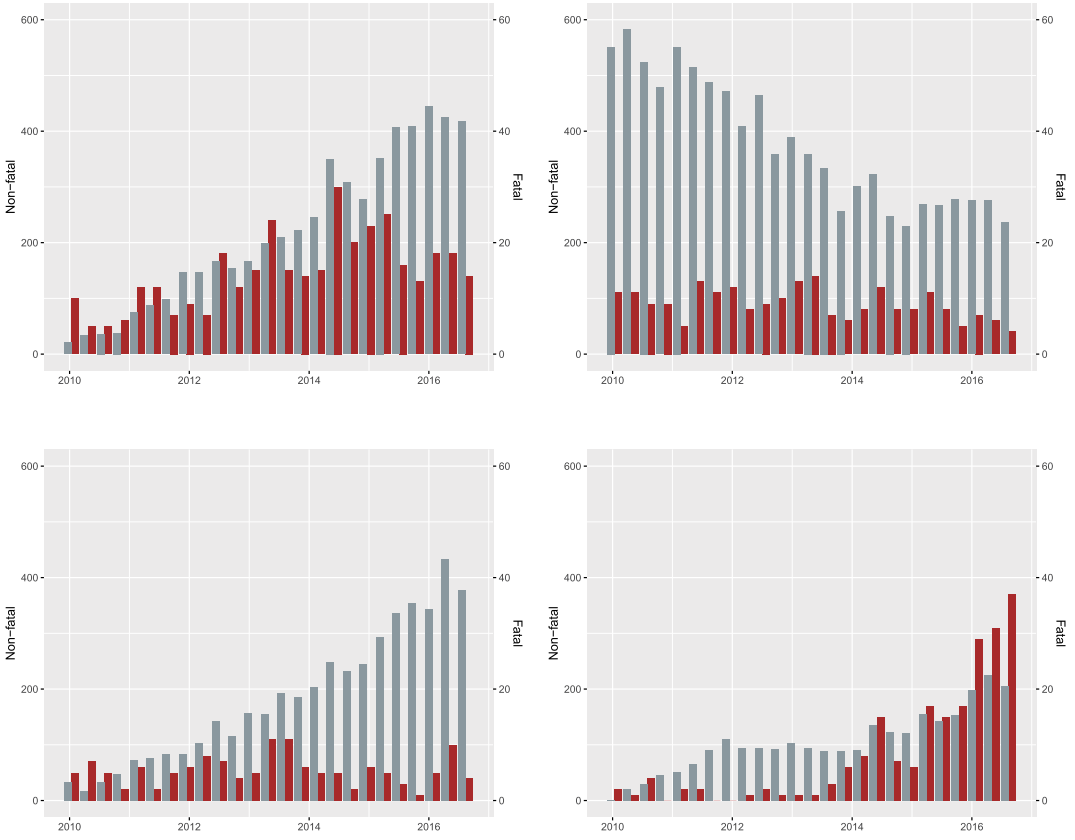

FIG. 7. *Histograms of nonfatal* (*grey*) *and fatal* (*red*) *overdose events for each group over time*: *group* 1 (*top left*), *group* 2 (*top right*), *group* 3 (*lower left*), *and group* 4 (*lower right*).

TABLE 6
Parameters of estimated model for each group

Group #	K_0	w	μ	σ
1	0.9609	0.0153	4.0517	0.0148
2	0.9864	0.0170	2.8304	0.0313
3	0.7257	0.0094	28.7279	0.0044
4	0.9214	0.0143	4.4550	0.0091

TABLE 7
Different measurement results on EMS data

Model	Log-likelihood	df	AIC	AUC
Baseline model	4.9892×10^4	4	-9.9774×10^4	**0.8032**
Proposed model	5.5752×10^4	16	$\mathbf{-1.1147 \times 10^5}$	0.7159

TABLE 8
Different measurement results on opioid overdose death data

Model	Log-likelihood	df	AIC	AUC
Baseline model	-3.6110×10^3	16	7.2540×10^3	0.8088
Proposed model	1.7165×10^3	16	$\mathbf{-3.4009 \times 10^3}$	**0.8524**

and substance-related deaths. One such policy example is the deployment of nasal naloxone by police and EMS agencies which mitigates overdose effects (Fisher et al. (2016)).

Integration of heterogeneous data sources also help to contextualize and better understand the nuances of how social harms may affect different populations of people. As our study illustrates, prescription drug overdoses occur at higher rates in areas further from downtown Indianapolis, while illicit drug overdoses are more concentrated around the urban core of the city. These results underscore societal differences of opioid drug use. Consistent with community explanations of crime and social disadvantage (Sampson and Groves (1989)), we observe that illicit drugs, which are more likely to result in mortality, may disproportionately impact minority communities. Current evidence indicates these trends are driven by heroin and synthetic opioid-related deaths as well as growing use of fentanyl-laced cocaine among African Americans (Alexander, Kiang and Barbieri (2018), Jalal et al. (2018)). Moreover, these trends persist despite evidence that African Americans are less likely to be prescribed opioids for pain relative to Caucasians (Meghani, Byun and Gallagher (2012)) which has been identified as a primary pathway to illicit opioid use (Mars et al. (2014)). Together, current evidence suggests the epidemiology of opioid use, especially illicit opioid use, is not well defined for racial-ethnic minorities. Heterogeneous data integration is likely the most appropriate path forward to improve our understanding of this issue.

Our work here is also related to the analysis of free text data that accompanies crime reports (Kuang, Brantingham and Bertozzi (2017), Mohler and Brantingham (2018), Pandey and Mohler (2018)) and other types of incidents, for example, railway accidents (Heidarysafa et al. (2018)). While the majority of point process focused studies of crime and social harm use only location, time and incident category as input into the model, we believe future research efforts on incorporating auxilliary, high-dimensional information into these models may yield improvements in model accuracy and also provide insight into the underlying causal mechanisms in space–time event contagion.

We do note that disentangling contagion patterns from other types of spatiotemporal clustering is challenging due to seasonal and exogeneous trends (Mohler (2013), Zhuang and Mateu (2019)). Future work should also focus on investigating the extent to which drug overdose triggering found in the present study can be detected across cities and model specifications.

Acknowledgments. This work was supported in part by NSF Grants ATD-1737996 and SCC-1737585.

REFERENCES

ALEXANDER, M. J., KIANG, M. V. and BARBIERI, M. (2018). Trends in black and white opioid mortality in the United States, 1979–2015. *Epidemiology* **29** 707.

CARTER, J. G., MOHLER, G. and RAY, B. (2018). Spatial concentration of opioid overdose deaths in Indianapolis: An application of the law of crime concentration at place to a public health epidemic. *J. Contemp. Crim. Justice*. https://doi.org/10.1177/1043986218803527

CICCARONE, D. (2017). Fentanyl in the US heroin supply: A rapidly changing risk environment. *Int. J. Drug Policy* **46** 107–111.

CICERO, T. J., ELLIS, M. S., SURRATT, H. L. and KURTZ, S. P. (2014). The changing face of heroin use in the United States: A retrospective analysis of the past 50 years. *JAMA Psychiatr.* **71** 821–826.

FISHER, R., O'DONNELL, D., RAY, B. and RUSYNIAK, D. (2016). Police officers can safely and effectively administer intranasal naloxone. *Prehosp. Emerg. Care* **20** 675–680. https://doi.org/10.1080/10903127.2016.1182605

GLADDEN, R. M., MARTINEZ, P. and SETH, P. (2016). Fentanyl law enforcement submissions and increases in synthetic opioid-involved overdose deaths—27 states, 2013–2014. *Morb. Mort. Wkly. Rep.* **65** 837–843. https://doi.org/10.15585/mmwr.mm6533a2

HEIDARYSAFA, M., KOWSARI, K., BARNES, L. and BROWN, D. (2018). Analysis of railway accidents' narratives using deep learning. In 2018 17*th IEEE International Conference on Machine Learning and Applications* (*ICMLA*) 1446–1453. IEEE.

HIBDON, J. and GROFF, E. R. (2014). What you find depends on where you look: Using emergency medical services call data to target illicit drug use hot spots. *J. Contemp. Crim. Justice* **30** 169–185.

HIBDON, J., TELEP, C. W. and GROFF, E. R. (2017). The concentration and stability of drug activity in Seattle, Washington using police and emergency medical services data. *J. Quant. Criminol.* **33** 497–517.

JALAL, H., BUCHANICH, J. M., ROBERTS, M. S., BALMERT, L. C., ZHANG, K. and BURKE, D. S. (2018). Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science* **361** eaau1184.

JONES, C. M., BALDWIN, G. T. and COMPTON, W. M. (2017). Recent increases in cocaine-related overdose deaths and the role of opioids. *Am. J. Publ. Health* **107** 430–432. https://doi.org/10.2105/AJPH.2016.303627

JONES, C. M., EINSTEIN, E. B. and COMPTON, W. M. (2018). Changes in synthetic opioid involvement in drug overdose deaths in the United States, 2010–2016. *JAMA* **319** 1819–1821. https://doi.org/10.1001/jama.2018.2844

KANDEL, D. B., HU, M.-C., GRIESLER, P. and WALL, M. (2017). Increases from 2002 to 2015 in prescription opioid overdose deaths in combination with other substances. *Drug Alcohol Depend.* **178** 501–511.

KUANG, D., BRANTINGHAM, P. J. and BERTOZZI, A. L. (2017). Crime topic modeling. *Crime Sci.* **6** 12.

LEE, D. D. and SEUNG, H. S. (2001). Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems* 556–562.

MARS, S. G., BOURGOIS, P., KARANDINOS, G., MONTERO, F. and CICCARONE, D. (2014). "Every 'never' I ever said came true": Transitions from opioid pills to heroin injecting. *Int. J. Drug Policy* **25** 257–266.

MEGHANI, S. H., BYUN, E. and GALLAGHER, R. M. (2012). Time to take stock: A meta-analysis and systematic review of analgesic treatment disparities for pain in the United States. *Pain Med.* **13** 150–174.

MOHLER, G. (2013). Modeling and estimation of multi-source clustering in crime and security data. *Ann. Appl. Stat.* **7** 1525–1539. MR3127957 https://doi.org/10.1214/13-AOAS647

MOHLER, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *Int. J. Forecast.* **30** 491–497.

MOHLER, G. and BRANTINGHAM, P. J. (2018). Privacy preserving, crowd sourced crime Hawkes processes. In 2018 *International Workshop on Social Sensing* (*SocialSens*) 14–19. IEEE.

MOHLER, G., CARTER, J. and RAJE, R. (2018). Improving social harm indices with a modulated Hawkes process. *Int. J. Forecast.* **34** 431–439.

MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and TITA, G. E. (2011a). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106** 100–108. MR2816705 https://doi.org/10.1198/jasa.2011.ap09546

MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and TITA, G. E. (2011b). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106** 100–108. MR2816705 https://doi.org/10.1198/jasa.2011.ap09546

MOHLER, G. O., SHORT, M. B., MALINOWSKI, S., JOHNSON, M., TITA, G. E., BERTOZZI, A. L. and BRANTINGHAM, P. J. (2015). Randomized controlled field trials of predictive policing. *J. Amer. Statist. Assoc.* **110** 1399–1411. MR3449035 https://doi.org/10.1080/01621459.2015.1077710

PANDEY, R. and MOHLER, G. O. (2018). Evaluation of crime topic models: Topic coherence vs spatial crime concentration. In 2018 *IEEE International Conference on Intelligence and Security Informatics* (*ISI*) 76–78. IEEE.

RAY, B. R., LOWDER, E. M., KIVISTO, A. J., PHALEN, P. and GIL, H. (2018). EMS naloxone administration as non-fatal opioid overdose surveillance: 6-year outcomes in Marion County, Indiana. *Addiction* **113** 2271–2279. https://doi.org/10.1111/add.14426

RUDD, R. A., PAULOZZI, L. J., BAUER, M. J., BURLESON, R. W., CARLSON, R. E., DAO, D., DAVIS, J. W., DUDEK, J., EICHLER, B. A. et al. (2014). Increases in heroin overdose deaths—28 states, 2010 to 2012. *Morb. Mort. Wkly. Rep.* **63** 849.

SAMPSON, R. J. and GROVES, W. B. (1989). Community structure and crime: Testing social-disorganization theory. *Am. J. Sociol.* **94** 774–802.

SETH, P., SCHOLL, L., RUDD, R. A. and BACON, S. (2018). Overdose deaths involving opioids, cocaine, and psychostimulants—United States, 2015–2016. *Morb. Mort. Wkly. Rep.* **67** 349–358. https://doi.org/10.15585/mmwr.mm6712a1

STEVENS, K., KEGELMEYER, P., ANDRZEJEWSKI, D. and BUTTLER, D. (2012). Exploring topic coherence over many models and many topics. In *Proceedings of the* 2012 *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* 952–961. Association for Computational Linguistics.

STRICKLER, G. K., ZHANG, K., HALPIN, J. M., BOHNERT, A. S., BALDWIN, G. and KREINER, P. W. (2019). Effects of mandatory prescription drug monitoring program (PDMP) use laws on prescriber registration and use and on risky prescribing. *Drug Alcohol Depend.*.

TOWNSLEY, M., HOMEL, R. and CHASELING, J. (2003). Infectious burglaries. A test of the near repeat hypothesis. *Br. J. Criminol*. **43** 615–633.

TSUI, K.-L., WONG, S. Y., JIANG, W. and LIN, C.-J. (2011). Recent research and developments in temporal and spatiotemporal surveillance for public health. *IEEE Trans. Reliab.* **60** 49–58.

VEEN, A. and SCHOENBERG, F. P. (2008). Estimation of space–time branching process models in seismology using an EM-type algorithm. *J. Amer. Statist. Assoc*. **103** 614–624. MR2523998 https://doi.org/10.1198/016214508000000148

WEISBURD, D. (2015). The law of crime concentration and the criminology of place. *Criminology* **53** 133–157.

YU, H.-L., KOLOVOS, A., CHRISTAKOS, G., CHEN, J.-C., WARMERDAM, S. and DEV, B. (2007). Interactive spatiotemporal modeling of health systems: The SEKS-GUI framework. *Stoch. Environ. Res. Risk Assess*. **21** 555–572. MR2380675 https://doi.org/10.1007/s00477-007-0135-0

ZEOLI, A. M., PIZARRO, J. M., GRADY, S. C. and MELDE, C. (2014). Homicide as infectious disease: Using public health methods to investigate the diffusion of homicide. *Justice Q*. **31** 609–632.

ZHUANG, J. and MATEU, J. (2019). A semiparametric spatiotemporal Hawkes-type point process model with periodic background for crime data. *J. Roy. Statist. Soc. Ser. A* **182** 919–942. MR3955503

ZHUANG, J., OGATA, Y. and VERE-JONES, D. (2002). Stochastic declustering of space–time earthquake occurrences. *J. Amer. Statist. Assoc*. **97** 369–380. MR1941459 https://doi.org/10.1198/016214502760046925