

RESEARCH ARTICLE

Modeling opioid overdose events recurrence with a covariate-adjusted triggering point process

Fenglian Pan¹, You Zhou², Carolina Vivas-Valencia³, Nan Kong^{2,4}, Carol Ott⁵,
Mohammad S. Jalali⁶ , Jian Liu^{1*} 

1 Department of Systems and Industrial Engineering, University of Arizona, Tucson, Arizona, United States of America, **2** Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana, United States of America, **3** Department of Biomedical and Chemical Engineering, The University of Texas at San Antonio, San Antonio, Texas, United States of America, **4** Edwardson School of Industrial Engineering, Purdue University, West Lafayette, Indiana, United States of America, **5** College of Pharmacy, Purdue University, Indianapolis, Indiana, United States of America, **6** MGH Institute for Technology Assessment, Harvard Medical School, Boston, Massachusetts, United States of America

* jianliu@arizona.edu



OPEN ACCESS

Citation: Pan F, Zhou Y, Vivas-Valencia C, Kong N, Ott C, Jalali MS. et al. (2025) Modeling opioid overdose events recurrence with a covariate-adjusted triggering point process. *PLoS Comput Biol* 21(5): e1012889. <https://doi.org/10.1371/journal.pcbi.1012889>

Editor: Gregg S. Gonsalves, Yale School of Public Health, UNITED STATES OF AMERICA

Received: May 17, 2024

Accepted: February 19, 2025

Published: May 5, 2025

Copyright: © 2025 Pan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Substance use disorder, particularly opioid-related, is a serious public health challenge in the U.S. Accurately predicting opioid overdose events and stratifying the risk of having such an event are critical for healthcare providers to deliver effective interventions in patients with opioid overdose. Despite a large body of literature investigating various risk factors for the prediction, the existing research to date has not explicitly investigated and quantitatively modeled how an individual's past opioid overdose events affect future occurrences. In this paper, we proposed a covariate-adjusted triggering point process to simultaneously model the effect of various risk factors on opioid overdose events and the triggering mechanism among opioid overdose events. The prediction performance was assessed by the U.S. state-wise Medicaid reimbursement claims data. Compared with commonly used prediction models, the proposed model achieved the lowest Mean Absolute Errors and Mean Absolute Percentage Errors on 30-, 60-, 90, 120-, 150-, and 180-day-ahead predictions. In addition, our results showed the statistical significance of considering the triggering mechanism for recurrent opioid overdose events prediction. On average, around 47% of the event recurrence were explained by the triggering mechanism.

Author summary

In complex natural systems, discrete event sequences often occur stochastically, such as earthquakes, financial transactions, and drug overdoses. The occurrence of these events is typically influenced by various factors (e.g., the occurrence of earthquakes is

influenced by tectonic zones experiencing deformation). Given these known factors, however, such events do not appear independently. Instead, the occurrence of one event additionally influences the rate of subsequent events (e.g., a mainshock increasing the rate of subsequent aftershocks). A similar pattern can be observed in drug overdose events, where occurrences are related to various risk factors, and an just occurred overdose increases the risk of subsequent overdoses. Existing methods that identify associations between overdoses and the risk factors provide limited insight into inter-event dependencies. Recently, the triggering event point process has been used to capture these complex inter-event dependencies. In this study, we propose a covariate-adjusted triggering point process, extending the conventional models to simultaneously account for the impacts of various factors and inter-event dependency on event occurrence. The results of the real data analysis demonstrates that proposed model has the potential to provide unique insights into the nature of inter-event dependencies and significantly improve prediction accuracy.

Introduction

The United States (U.S.) is experiencing a nationwide opioid overdose epidemic, which witnesses a noticeable mortality rate and socioeconomic burden [1]. This situation has been exacerbated in recent years due to the emergence of fentanyl in the illicit opioid market since 2013 [2]. In 2020, opioids were involved in about 70,000 overdose deaths in the U.S. [3], and this number rose to over 80,400 in 2021 [4]. Opioid overdose is the most common cause of death among various substance overdoses. For example, it accounted for 81.7% of the total drug overdose deaths in 2016 [5]. In response to this crisis, the U.S. Centers for Disease Control and Prevention (CDC) has added opioid overdose prevention to its list of top five public health challenges in 2014 [6].

In this context, identifying individuals at high risk of opioid overdose has been shown as a key opportunity to connect patients with essential services for potentially lifesaving prevention [7]. In addition, predicting the temporal trends of recurrent opioid overdoses may help evaluate the impact of prevention efforts, particularly given the predominance of fentanyl and the exacerbation of the overdose crisis by the COVID-19 pandemic [8]. In this paper, we aim to establish a general statistical model to model the intensity of recurrent opioid overdose events for each patient, which can be viewed as the instantaneous rate of opioid overdose occurrence over time, is an important indicator of the risk of opioid overdose. Based on the estimated intensity, we can predict the risk of opioid overdose and the temporal trends of the recurrent opioid overdoses in the future time intervals.

While understanding and predicting the occurrence of opioid overdoses is crucial, it faces some challenges due to the complex opioid overdose occurrence processes. First, anyone with opioid use disorder (OUD) may experience an opioid overdose at any moment randomly, and the intrinsic tendency of having an opioid overdose can differ significantly from one patient to another. This difference may be related to various risk factors, including age, sex, substance amount used, and treatment received [9]. Second, opioid overdose often occurs recurrently in the majority of patients who suffer from chronic OUD [10]. Given all of the known risk factors, the recurrent overdose events may not be independent. With the chronic condition, an already occurred opioid overdose may trigger the occurrence of subsequent opioid overdoses, which is regarded as an event-to-event triggering mechanism in this paper. The event-to-event triggering mechanism exists in biobehavioral health, making opioid overdose

recurrence history-dependent and influencing the occurrence rate of subsequent opioid overdoses. Intuitively, this is because a patient with OUD consumes an excessive amount of opioids, leading to an opioid overdose, which could create a pattern of repeated opioid use and recurring opioid overdoses in the future. Empirically, we explored the distribution of time intervals between two consecutive opioid overdoses among recipients in our data. As shown in Fig 1, around 72% of time intervals between two consecutive opioid overdoses are within thirty days. This also suggested that an individual who has experienced an overdose is at a high risk of subsequent overdoses in the near future, aligning with the finding in the existing literature [11,12]. Our research was motivated by the need to understand and quantify the additional event-to-event triggering mechanism among overdoses, apart from effects captured by the known risk factors. This work is enabled by the opportunities and challenges in analyzing opioid overdose records among Medicaid recipients, using reimbursement claims data collected by a state Medicaid office.

In the literature, considerable methods, including clinical assessment rules, statistical models, and machine learning (ML) and deep learning (DL) methods, have been developed to examine risk factors that are associated with opioid overdose, such as demographic factors [9] and clinical characteristics [13]. For example, at the early stage, clinical assessment criteria were proposed to identify patients at risk by examining whether high-risk criteria (e.g., need for intubation and seizures) were present in a patient's medical record (e.g., emergency room notes, progress notes by doctors and nurses, and vital signs sheets) [14]. Subsequently, clinical prediction rules were developed to synthesize potentially influencing clinical characteristics (e.g., blood pressure and comorbid illnesses) associated with overdose [15]. These clinical prediction rules are knowledge-based and generally rely on limited clinical test data. Normally, they can only make reasonable predictions on high-risk encounters *within one event episode*. This is unlikely applicable to our prediction problem, which studies the tendency and the recurrence of opioid overdoses.

With increasing data available, statistical models, including survival analysis [16] and Poisson regression-based models [17], have emerged in recent years for understanding the quantitative relationship between opioid overdose and various risk factors. Such methods model the risk of opioid overdose as a function of risk factors in six categories, including patient's demographics, mental health comorbidities, substance use disorders, physical health comorbidities, characteristics of opioids prescribed, and non-opioid medications prescribed [9]. For instance, the relationship between the risk of opioid overdose and the risk factors can be

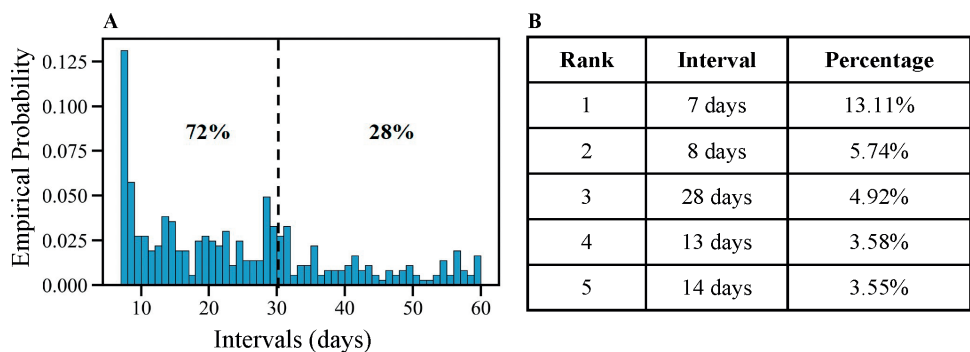


Fig 1. Distribution of time intervals between two consecutive opioid overdoses. (A) Histogram of the time intervals. (B) Top 5 time intervals.

<https://doi.org/10.1371/journal.pcbi.1012889.g001>

represented as a hazard function in a Cox Proportional Hazards Regression (CPHR) [18]. A five-variable CPHR (i.e., age, age-squared, long-acting and extended-release formulation, tobacco use, and mental health and substance abuse/dependence diagnoses) was developed to identify high-risk opioid overdose individuals [16]. However, conventional CPHR-based models are mainly designed to model survival data, referring to a type of data where the primary outcome of interest is the time until the event of interest (e.g., death or disease diagnosis) occurs. They only take into account the first overdose occurrence and discard subsequent event information, which likely leads to modeling inaccuracy due to interdependence between the recurrent events.

To analyze recurrent events, the Andersen-Gill (A-G) model [19] that generalizes CPHR was proposed to allow for the analysis of recurrent event data. For example, a multivariate A-G model was used to identify demographic, health, social, and criminal justice predictors of non-fatal overdose [20,21]. The Poisson regression models were also widely used in modeling recurrent event data, which focus on the number of times that an event of interest has occurred [22]. For instance, log-linear and logistic Poisson models with a Bayesian space-time framework were used to predict the future course of overdose mortality [17]. A Poisson regression model was used to examine the relationship between prescription opioid rates and prescription opioid overdose deaths using spatial cluster and regression analyses [23]. Such models are specially designed to count data and assume that the opioid overdoses are independent. Thus, they cannot model the event-to-event triggering mechanism among recurrent events.

Recently, the development of interoperable health IT technology and the establishment of large data repositories with sufficient data harmonization has helped realize longitudinal environments [24]. For example, a large quantity of longitudinal data related to overdose, including hundreds of potential risk factors, were acquired from electronic health records (EHRs) [25]. With such data readily available, multiple ML techniques, including random forest [26], decision tree [1], logistic regression [27], and *K*-means [28] were used to predict an individual's risk of overdose. In addition, deep learning methods gained popularity for EHR-based predictive modeling. A long short-term memory (LSTM) model was built to identify whether a patient is at high risk of overdose in the next hospital visit [29]. These studies have demonstrated that ML- and DL-based predictive models can achieve promising results in learning a patient's risk of opioid overdose from EHR history. However, there are two main differences between the existing ML- and DL-based method and our proposed method. First, our method enables reliable identification and quantification of the triggering mechanism, which is not explicitly considered in the existing methods. Second, our research focuses on predicting the occurrence rate of future recurring opioid overdoses. The classification-based methods focus on determining whether a patient faces a high risk of opioid overdose in the future.

The proposed method is based on the Hawkes process [30], which is a special type of point process and models self- and mutual-exciting patterns among recurrent events. It has drawn much attention in various fields, such as seismology [31], criminology [32], epidemiology [33], neural activity [34], and engineering [35]. The intensity function of the Hawkes process consists of two terms: a baseline intensity, which is independent of past events, and a triggering intensity, which captures the triggering effects of past events on future events. For example, each earthquake usually triggers a sequence of aftershock activities, and the occurrence rate of aftershocks can be represented in the triggering intensity function. The self- and mutual-exciting point processes align well with the aforementioned event-to-event triggering mechanism among recurrent opioid overdoses. Motivated by the need of understanding

the additional triggering mechanism among overdose events and better predicting the occurrence of overdose events, we propose a covariate-adjusted triggering point process to model the intensity function of recurrent opioid overdoses. Specifically, the intensity function is controlled by two components: the baseline event rate, which is used to capture the heterogeneity of individuals such as demographics factors, and is independent of past opioid overdoses, and the triggering intensity, which is collected from the positive triggering effect of every past opioid overdoses.

The rest of this paper is organized as follows. **Methods** section introduces a methodological framework that includes study cohorts, model formulation, and model estimation. **Results** section presents the model implementation and demonstrates the advantages of our proposed method in terms of model prediction accuracy and interpretability. The conclusion and future work will be discussed in **Discussion** section and the **Limitations** section summarizes the limitations of our methods.

Methods

Ethics statement

This study was approved by the Purdue University Institutional Review Board (2019-118). The IRB was an umbrella IRB that was used by a recently awarded grant from the United States Centers for Medicare and Medicaid (CMS) through the Indiana Family and Social Service Administrations (FSSA) so all Purdue researchers were able to utilize the data for their studies. Co-authors Kong and Vivas-Valencia (then as a PhD student at Purdue) were included on the original IRB. Co-author Zhou was added once he became a graduate student at Purdue in 2021. The data was physically transferred to Purdue University and housed on a secured data server. The data was further de-identified by a centrally appointed data management technician. Then only the de-identified data were made available to the Purdue campus-wide researchers. The data used for our analyses reported in this manuscript were prepared by Vivas-Valencia under the guidance of co-authors Kong's and Ott's. The reporting of Medicaid data by health care providers was mandated by the United States Centers for Medicare and Medicaid (CMS) for all Medicaid beneficiaries. Because of the de-identified nature of the data, we did not own the data and none of the authors were designated custodian of the data for data sharing with Purdue. At any point of the research project, non-Purdue researchers were granted the access to the data set. The workflow was such that the University of Arizona researchers prepared the codes based on the discussion about the data structure and they sent the code to Purdue so co-author Zhou could run the code on the de-identified data.

Cohort selection, index date, and covariates

We identified recipients enrolled in the a state-wise Medicaid program who experienced two or more opioid overdoses between January 1, 2016, and December 31, 2018. Opioid overdoses were determined using ICD-10-CM diagnosis codes (see supplementary material for the list of codes). For each patient, we recorded the earliest overdose date following a baseline period of 90 days with no observed overdoses. The date 90 days after the earliest overdose date was defined as the individual's *index date*. Patients were excluded from the study if there were no overdoses after the index date. This is because individuals with only one recorded overdose do not include recurrence information. Additionally, our dataset only recorded overdoses between January 1, 2016, and December 31, 2018. For individuals with only one recorded overdose, it is unclear whether this reflects a genuinely single event or a truncated record of their history. Including individuals assumed to have only one overdose (but who might have

experienced additional overdoses outside the observation period) could introduce bias into the model estimation. Patients were excluded from the study if they were under 18 years old on the index date or had missing demographic information. A detailed flowchart of the data extraction and preprocessing process is shown in Fig 2. After preprocessing, we identified 1,186 patients who met these criteria. These patients were followed from the index date until they were censored on December 31, 2018. The model was trained on 949 patients (80%) and externally validated in the remaining 237 patients (20%).

We included six known covariates that may influence the baseline intensity of opioid overdoses, including demographics, opioid withdrawal, and medications for opioid use disorder (MOUD). For demographic information, we considered the following categorical variables: (i) sex, classified as male or female; (ii) age, grouped into young adults (18-34), middle-aged adults (35-49), older adults (50-65), and seniors (66+); (iii) race, categorized into four groups: Caucasian, Black, Hispanic, and Others; and (iv) geographic classification, where the state was divided into rural, urban, and mixed rural-urban areas based on 3-digit ZIP codes. These demographic covariates were observed at the start of the study and were treated as static covariates over the observation period. The opioid withdrawal and MOUD were treated as time-dependent binary variables in our model.

Model formulation

The stochastic nature of randomly recurring opioid overdoses can be modeled as a point process, where the recurring opioid overdoses are treated as a series of random points distributed in time [36]. Consider a data set that includes the opioid overdoses recorded from a cohort of N recipients, in which each recipient i , $i = 1, \dots, N$, can experience opioid overdoses recurrently during the observed window $[0, T)$. Let \mathbf{t}_i denote a collection of timestamps of opioid overdoses of recipient i , i.e., $\mathbf{t}_i = \{t_{ij} | j = 1, \dots, n_i\}$, where $t_{i1} < \dots < t_{in_i}$ and n_i denotes the total number of opioid overdoses of recipient i . In addition to the timestamps of opioid overdoses, this paper considers the recipient's demographic factors associated with the opioid overdoses, such as sex, age, and race. Let $\mathbf{x}_i \in \mathbb{X} \subset \mathbb{R}^p$ denote a vector of demographic factors of recipient i . Thus, the historical data of recipient i during $[0, T)$, can be denoted as $\mathcal{H}_i(T) = (\mathbf{t}_i, \mathbf{x}_i)$ and the

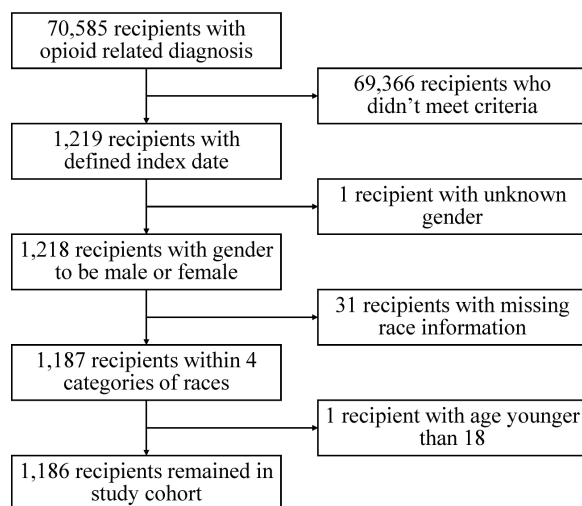


Fig 2. Flowchart of data extraction and preprocessing.

<https://doi.org/10.1371/journal.pcbi.1012889.g002>

overall historical data from N recipients, during $[0, T)$, can be denoted as $\mathcal{H}(T) = \{\mathcal{H}_i(T) | i = 1, \dots, N\}$.

From a modeling point of view, the opioid overdoses of recipient i , i.e., \mathbf{t}_i , can be treated as a series of random points occurring in time $[0, T)$ and be modeled by a point process. Specifically, for each recipient i , its expected instantaneous occurrence probability of an opioid overdose at time t can be modeled by a conditional intensity function (CIF, “intensity” hereafter) given a set of historical event data, $\mathcal{H}_i(t)$, during $[0, t)$, i.e.,

$$\lambda_i(t | \mathcal{H}_i(t)) = \lim_{dt \rightarrow 0} E[N([t, t + dt)) | \mathcal{H}_i(t)] / dt, \tag{1}$$

where $N([t_1, t_2))$ denotes a non-negative integer-valued point process counting the number of opioid overdoses occurring in the time interval $[t_1, t_2)$, and $E[N([t, t + dt)) | \mathcal{H}_i(t)]$ is the expected number of opioid overdoses occurring in the time interval $[t, t + dt)$, given $\mathcal{H}_i(t)$.

As discussed in [Introduction](#) section, each recipient using opioids may experience an opioid overdose at any moment randomly, and the probability of having such an opioid overdose may be related to various risk factors or the event-to-event triggering mechanism. Such occurrence characteristics of opioid overdose provide us a hint to decompose the total intensity of opioid overdoses into two parts. Specifically, for each recipient i , an opioid overdose may be caused by various risk factors, whose occurrence randomness is modeled by a baseline intensity λ_i^b , with the superscript “ b ” denoting a baseline intensity, or it may be triggered by an occurred opioid overdose, which can be modeled as λ_i^t , with the superscript “ t ” denoting the triggering intensity, and be combined with the baseline intensity function, λ_i^b . Thus, the overall opioid overdose occurrence intensity of recipient i can be formulated as:

$$\lambda_i(t | \mathcal{H}_i(t)) = \lambda_i^b(t) + \lambda_i^t(t), i = 1, 2, \dots, N, \tag{2}$$

where $\mathcal{H}_i(t)$ represents historical data of recipient i . It is worth noting that there is no label indicating whether an overdose is due to risk factors or triggered by a specific mechanism.

Baseline intensity. It has been shown that the risk of opioid overdose is time-varying [\[37\]](#) and related to various risk factors, such as demographics and health conditions. For example, evidence shows that opioid dosage was the factor most consistently associated with increased risk of opioid overdose [\[38\]](#). In this context, we model the baseline intensity, $\lambda_i^b(t)$, in a log-linear form, as often used in Cox-type regression models in public health [\[39,40\]](#), i.e.,

$$\lambda_i^b(t) = \lambda_0(t) \cdot \exp(\alpha^\top \mathbf{x}_i), \tag{3}$$

where $\lambda_0(t)$ is a function used to model the temporal dependencies of baseline rate. It can be time-independent or time-dependent. Without loss of generality, we adopt a Power Law function, i.e., $\lambda_0(t) = \theta \cdot t^{\theta-1}$, which is mathematically simple and flexible to capture the constant (when $\theta = 1$), decreasing (when $\theta < 1$), and increasing (when $\theta > 1$) event rates. Additionally, model selection was performed using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The results demonstrated that the model with one-parameter Power Law function outperformed the model with three-parameter Weibull function, as it achieved lower AIC and BIC values. In addition, we incorporate individualized covariates $\mathbf{x}_i \in \mathbb{X} \subset \mathbb{R}^p$ into the baseline intensity with a log-linear form with the corresponding coefficient vector $\alpha \in \mathbb{A} \subset \mathbb{R}^p$, to model the heterogeneity of individuals.

Triggering intensity. In addition, an occurred opioid overdose may trigger the occurrence of subsequent opioid overdoses. The existence of the event-to-event triggering

mechanism additionally increases the intensity of future opioid overdoses to the baseline intensity. Motivated by Hawkes process [30], this paper models the increased intensity, i.e., triggering intensity, $\lambda_i^t(t)$, as:

$$\lambda_i^t(t) = \sum_{j:t_{ij}<t} \beta \cdot g(t - t_{ij}), \tag{4}$$

where the term $\beta \cdot g(t - t_{ij})$, $\beta > 0$ represents the triggering effect on overdose intensity at time t caused by a past overdose occurring at time t_{ij} . Specifically, β quantifies the magnitude of the increase in overdose intensity following an overdose at t_{ij} , which subsequently decays according to the decay function $g(\cdot)$ until the next overdose occurs to raise the intensity again. In this study, without loss of generality, we employed an exponential decay function, $g(t) = \gamma \cdot \exp(-\gamma \cdot t)$, $\gamma > 0$. This function was chosen for its simplicity and effectiveness in capturing the natural decay of the triggering effect over time. It has been well-established and widely applied in fields such as epidemiology [33] and health analytics [41].

With the above specification, the total intensity function of recipient i in Eq (2) is expressed as:

$$\lambda_i(t|\mathcal{H}(t)) = \theta \cdot t^{\theta-1} \cdot \exp(\alpha^T \mathbf{x}_i) + \sum_{j:t_{ij}<t} \beta \cdot \gamma \cdot \exp(-\gamma \cdot (t - t_{ij})). \tag{5}$$

With parameters $\Theta = \{\theta, \alpha, \beta, \gamma\}$ in Eq (5) estimated, we can calculate the cumulative intensity of opioid overdose (also known as the cumulative hazard) by $\Lambda_i(t) = \int_0^t \lambda_i(u) du$ for recipient i . Thus, the risk of opioid overdose by time t can be calculated by $1 - \exp(-\Lambda_i(t))$ [42].

Model estimation

This paper uses the maximum likelihood estimation (MLE) method for parameter estimation. Let T denote the length of the observation window, given the historic observational data $\mathcal{H}(T) = \{(\mathbf{t}_i, \mathbf{x}_i) | i = 1, \dots, N\}$, the likelihood function associated with overall parameters $\Theta = \{\theta, \alpha, \beta, \gamma\}$ defined in Eq (5), can be formulated as,

$$L(\Theta) = \prod_{i=1}^N \left\{ \prod_{j=1}^{n_i} \lambda_i(t_{ij}; \mathbf{x}_i, \Theta) \right\} \times \exp\{-\Lambda_i(T; \mathbf{x}_i, \Theta)\}, \tag{6}$$

where the intensity function $\lambda_i(t; \mathbf{x}_i, \Theta)$ is defined in Eq (5), and the term $\Lambda_i(T; \mathbf{x}_i, \Theta) = \int_0^T \lambda_i(t; \mathbf{x}_i, \Theta) dt$ is the cumulative intensity function. The log-likelihood function is obtained by taking the logarithm of $L(\Theta)$ in Eq (6). That is,

$$\ell(\Theta) = \sum_{i=1}^N \left\{ \sum_{j=1}^{n_i} \log(\lambda_i(t_{ij}; \mathbf{x}_i, \Theta)) \right\} - \Lambda_i(T; \mathbf{x}_i, \Theta), \tag{7}$$

where the first term depends on the historical observation data, i.e., $\mathcal{H}(T) = \{(\mathbf{t}_i, \mathbf{x}_i) | i = 1, \dots, N\}$ and can be easily computed by plugging the historical data into Eq (5). The second term is a definite integral of Eq (5), which can be explicitly expressed and depends on observed demographic factors, i.e., \mathbf{x}_i , $i = 1, \dots, N$. Thus, the MLE of Θ , denoted by $\hat{\Theta}$, can be obtained by finding the value of Θ that maximizes $\ell(\Theta)$ in Eq (7), i.e.,

$$\Theta = \operatorname{argmax}_{\Theta} \ell(\Theta). \tag{8}$$

To address the bounded constraints on some of the parameters in Θ , we use the “L-BFGS-B” option in the Python `minimize()` function. The “L-BFGS-B” algorithm allows users to specify an interval for each variable to be optimized [43].

Results

Model implementation

We applied the proposed method to the study cohort. The cohort was randomly divided into a training set, comprising 80% of the data, and a test set, comprising the remaining 20%, with 100 replications. In each replication, r , the proposed model was trained based on the training set to obtain the estimated parameters $\hat{\Theta}^r$ and predicted intensity $\hat{\lambda}_i^r(t, \hat{\Theta}^r)$. Given $\hat{\lambda}_i^r(t, \hat{\Theta}^r)$, the model was externally validated in the test set. Specifically, the number of opioid overdoses in the M recipients in the test set during time interval $[0, L]$ can be predicted as [42]:

$$\hat{m}_{[0,L]}^r = \sum_{i=1}^M \int_0^L \hat{\lambda}_i^r(t, \mathbf{x}_i, \hat{\Theta}^r) dt, \quad (9)$$

where L is the length of the prediction window. The predicted number above would be compared to the actual number of opioid overdoses in the test set for the same time interval, denoted as $m_{[0,L]}^r$. For a certain length of the prediction window, L , this comparison was conducted for 100 times, i.e., $r = 1, 2, \dots, 100$.

Model evaluation

To measure the prediction accuracy, we define the mean absolute error (MAE) over the R replications as [44]:

$$\text{MAE} = \frac{\sum_{r=1}^R |\hat{m}_{[0,L]}^r - m_{[0,L]}^r|}{R}. \quad (10)$$

A smaller MAE implies better model performance. However, the MAE is sensitive to the scale of actual numbers of opioid overdoses and can not provide a standardized measure when dealing with the prediction results at different prediction windows. In response, the mean absolute percentage error (MAPE) over the R prediction windows is considered and calculated by:

$$\text{MAPE} = \frac{100\%}{R} \cdot \sum_{r=1}^R \frac{|\hat{m}_{[0,L]}^r - m_{[0,L]}^r|}{m_{[0,L]}^r}. \quad (11)$$

MAPE is expressed as a percentage, making it more interpretable for modelers who may not know the magnitude of actual opioid overdoses. Both MAE and MAPE were used to evaluate the performance of the proposed method.

To investigate the advantage of the proposed method that explicitly considers the event-to-event triggering mechanism, we compared the prediction performance of the proposed method and the benchmarks that ignore the triggering mechanism. The benchmarks and the proposed method are listed below:

- **Homogeneous Poisson Process (HPP)** assumes that opioid overdoses of each individual are independent, and the occurrence rate is constant, i.e., the intensity of subject i is $\lambda_i(t) = \theta, \theta > 0$.

- **Non-homogeneous Poisson Process (NHPP)** assumes that opioid overdoses of each individual are independent, and the occurrence rate follows a predefined time-varying function [45]. In this paper, the adopted intensity of subject i is: $\lambda_i(t) = \theta \cdot t^{\theta-1} \cdot \exp(\alpha^\top \mathbf{x}_i)$, where \mathbf{x}_i denotes the individualized covariates.
- **Andersen-Gill (A-G) model** allows for the analysis of recurrent event data [19].
- **The proposed method** explicitly considers the impacts of the demographic factors and the event-to-event triggering mechanism among opioid overdoses. The intensity of subject i is defined in Eq (5), i.e., $\lambda_i(t|\mathcal{H}_i(t)) = \theta \cdot t^{\theta-1} \cdot \exp(\alpha^\top \mathbf{x}_i) + \sum_{j:t_{ij}<t} \beta \cdot \gamma \cdot \exp(-\gamma \cdot (t - t_{ij}))$.

Without loss of generality, we trained different models in the same training set and predicted the number of opioid overdoses at prediction windows of 30, 60, 90, 120, 150, and 180 days. (i.e., $L = 30, 60, 90, 120, 150, \text{ and } 180$), respectively. The MAE and MAPE defined in Eqs (10)–(11) were used to evaluate the accuracy of the prediction.

Model accuracy

Fig 3 illustrated the MAEs of the proposed method and the benchmarks on 30-, 60-, 90-, 120-, 150-, and 180-days-ahead predictions, providing the average absolute differences between the true opioid overdose counts and predicted counts. There are several notable findings. First, the proposed method consistently achieves the lowest MAE across all prediction windows, highlighting its superior predictive accuracy compared to the benchmark models. Second, the MAE values increase as the prediction window lengthens, which can be attributed to the corresponding increase in true opioid overdose counts. Third, when compared to the HPP and A-G models, the proposed method reduces the MAE by a value ranging between 30%-49%, demonstrating a substantial improvement in prediction accuracy. When compared to the NHPP model, the proposed method reduces the MAE by a value ranging between 3%-30%, with performances converging for the longer prediction windows. Specifically, the proposed method lowers the MAE by 6% for the 150-day prediction window and by 3% for the 180-day prediction window.

Fig 4 displayed the MAPEs of the proposed method and the benchmarks across 30-, 60-, 90-, 120-, 150-, and 180-day prediction windows, measuring the percentage deviation between actual and predicted opioid overdose counts. The proposed method consistently

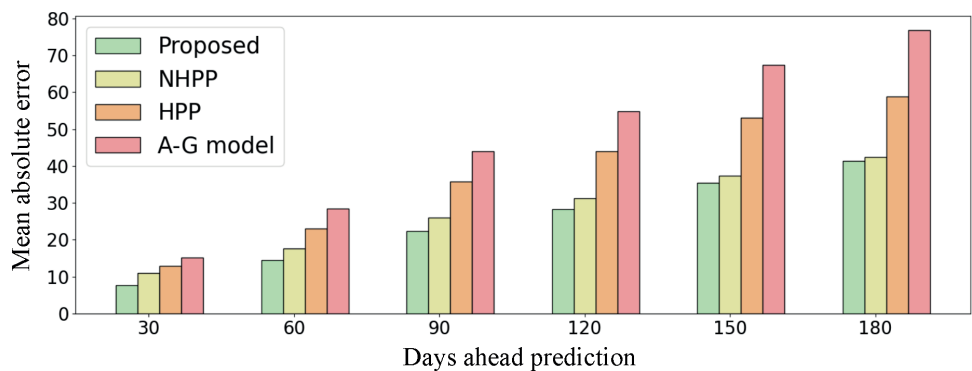


Fig 3. Mean absolute errors of the proposed method and the benchmarks on 30-, 60-, 90-, 120-, 150-, 180-day ahead predictions.

<https://doi.org/10.1371/journal.pcbi.1012889.g003>

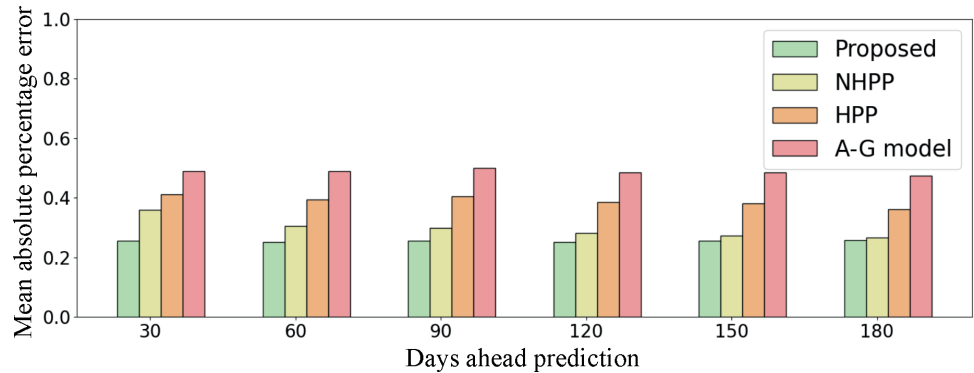


Fig 4. Mean absolute percentage errors of the proposed method and the benchmarks on 30-, 60-, 90-, 120-, 150-, 180-day ahead predictions.

<https://doi.org/10.1371/journal.pcbi.1012889.g004>

achieves the lowest MAPE across all prediction windows, maintaining an approximate 25% error rate, which demonstrated its robustness in accurately predicting overdose counts.

In addition to predicting the number of opioid overdoses, we also wanted to understand and quantify the effect of the triggering mechanism on opioid overdose intensity. The quantification result can help healthcare providers predict the likelihood of some individual to have repeated overdoses and thus tailor intervention plans for these individuals based on the predicted likelihood. We first evaluated the significance of the triggering mechanism by testing whether the estimated triggering parameter, $\hat{\beta}$, across 100 repeated estimates, was significantly different from zero. Specifically, we conducted a t-test for $\hat{\beta}$ with the null hypothesis $H_0: \hat{\beta} = 0$ and the alternative hypothesis $H_a: \hat{\beta} > 0$. The test yielded a p-value of less than 0.001. Based on our significance level, $\alpha_c = 0.05$, we rejected the null hypothesis and included the triggering mechanism in our model.

After testing the significance of the triggering effect, we calculated the proportion of the expected number of opioid overdoses explained by the triggering intensity and the baseline intensity within the test set. As shown in Fig 5, the grey bar represented the expected number of opioid overdoses explained by the baseline intensity, and the blue bar represented the expected number of opioid overdoses explained by the triggering intensity. We observed the following findings. First, different from benchmarks, the proposed method can explicitly quantify the triggering mechanism among opioid overdoses, providing a more nuanced interpretation of different types of effects. Thus, public health officials or practitioners can obtain a clearer view of the causes of opioid overdoses and assess the efficacy of interventions and policies implemented in response to the opioid crisis. Second, the proportion of the expected numbers of opioid overdoses explained by the triggering intensity consistently remained at 47% across all prediction windows, indicating that our model is robust to variations in prediction window length. Finally, the consistent 47% contribution from the triggering mechanism highlights its significant role in opioid overdose prediction.

Discussion

Accurately predicting the risk of opioid overdose is essential for delivering effective treatment to individuals with OUD and for guiding public health policy development. The occurrence of opioid overdose is influenced not only by various risk factors but also by the event-to-event triggering mechanism inherent among opioid overdoses. This paper introduced a covariate-adjusted triggering point process that models both the underlying risk factors and

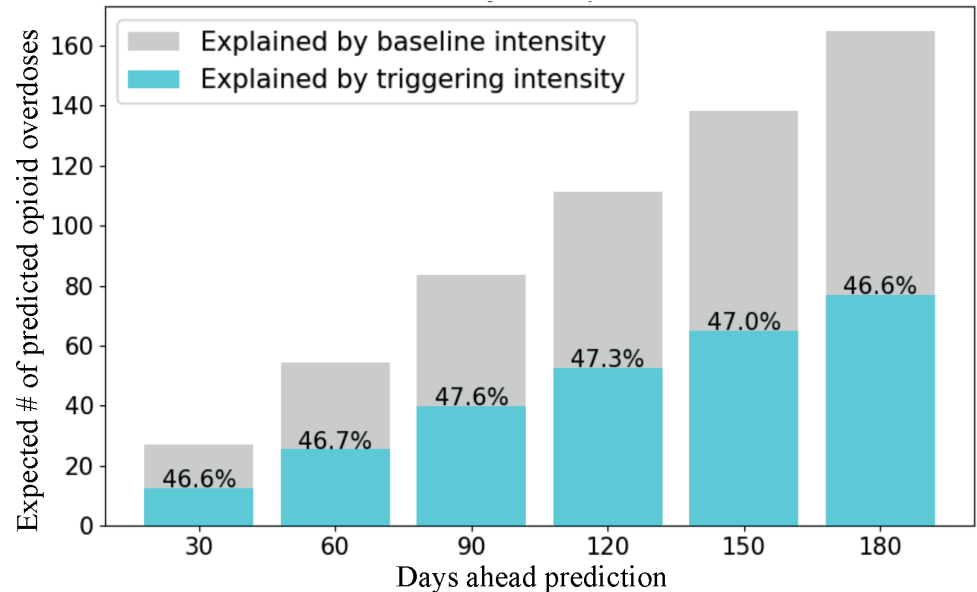


Fig 5. Expected opioid overdoses explained by baseline intensity and triggering intensity.

<https://doi.org/10.1371/journal.pcbi.1012889.g005>

the triggering effects on opioid overdose risk. For the case study, we utilized Medicaid reimbursement claims data collected from a U.S. state. The results demonstrated that the proposed method offers enhanced predictive accuracy by incorporating the triggering mechanism, which has not been considered in the existing methods. The results of our model could be used by healthcare providers to mitigate the risk of future overdoses. For example, based on the predicted risk, healthcare providers can stratify patients into higher or lower risk categories which allows for more efficient use of resources, focusing interventions on those most at risk for repeated overdoses. In addition, the proposed method allows for quantification of the triggering effect, offering a more nuanced interpretation of how different factors contribute to overdose risk, which can help healthcare providers predict the likelihood of a repeated overdose and tailor intervention plans based on individual needs.

There are several directions for our future studies. First, an immediate task will be to investigate the adaptation of the proposed method to the same data collected from other states in the U.S. Second, the proposed parametric modeling approach can be extended to allow the triggering mechanism to vary based on covariates. Third, this study primarily focused on exploring the triggering effects inherent in recurrent overdoses, which required individuals to have experienced multiple overdose events. However, in practice, some individuals may experience only a single overdose, and these cases were not included in our analysis. In the future, developing a joint modeling framework that incorporates single-overdose events, censored single-overdose events, and recurrent overdoses within a unified structure could enhance the model's applicability to practical clinical settings.

Limitations

While the proposed method demonstrated its good performance using Medicaid reimbursement claims data from a single state, covering the period from January 1, 2016, to December 31, 2018, several limitations should be noted. First, the study population was limited to a single state, which may reduce the generalizability of our findings to other states due to

differences in drug availability, healthcare services, and prescribing patterns. Second, we were unable to account for opioid overdoses that occurred prior to January 1, 2016 and after December 31, 2018, as these data were not available. These unobserved events may have influenced subsequent overdose risk, potentially impacting our model's results. Third, we excluded individuals with only one recorded overdose during the study period, which may introduce potential bias. This exclusion limits the generalizability of our findings to clinical settings where single-overdose patients constitute a significant proportion of the population. Fourth, Medicaid claims data may not fully capture all OUD-related events. For instance, illicit opioid use, non-fatal overdoses where medical care was not sought, or opioid prescriptions paid for out-of-pocket may not be recorded, leading to potential underestimation of overdose risk. Additionally, patients who changed insurers, relocated, or died from fatal overdoses during the study period could result in incomplete longitudinal data, which may have led to an overestimation of the risk. Unfortunately, these gaps in information could not be fully addressed. Lastly, claims data are primarily used for billing purposes and thus lack detailed clinical information and important covariates, which may have limited the predictive performance of our model.

Supporting information

S1 Supplementary Material. ICD-10-CM codes used to identify study population.
(PDF)

S1 Sample Data. Data for model training and testing in Results section.
(XLSX)

S2 Sample Data Codebook. This document serves as a reference guide for understanding sample data.
(PDF)

Author contributions

Conceptualization: Fenglian Pan, Carolina Vivas-Valencia, Nan Kong, Carol Ott, Mohammad S. Jalali, Jian Liu.

Data curation: You Zhou, Carolina Vivas-Valencia.

Formal analysis: Fenglian Pan, Jian Liu.

Investigation: Fenglian Pan, You Zhou, Nan Kong, Carol Ott, Mohammad S. Jalali, Jian Liu.

Methodology: Fenglian Pan, Nan Kong, Jian Liu.

Supervision: Jian Liu.

Writing – original draft: Fenglian Pan, Jian Liu.

Writing – review & editing: You Zhou, Carolina Vivas-Valencia, Nan Kong, Carol Ott, Mohammad S. Jalali, Jian Liu.

References

1. Lee S, Wei S, White V, Bain PA, Baker C, Li J. Classification of opioid usage through semi-supervised learning for total joint replacement patients. *IEEE J Biomed Health Inform.* 2021;25(1):189–200. <https://doi.org/10.1109/JBHI.2020.2992973> PMID: 32386170
2. Althoff KN, Leifheit KM, Park JN, Chandran A, Sherman SG. Opioid-related overdose mortality in the era of fentanyl: monitoring a shifting epidemic by person, place, and time. *Drug Alcohol Depend.* 2020;216:108321. <https://doi.org/10.1016/j.drugalcdep.2020.108321> PMID: 33007700

3. National Center for Health Statistics (NCHS). Multiple cause of death 1999–2020 on cdc wonder online database, 2021. Data are compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. 2021.
4. National Center for Health Statistics N. Wide-ranging online data for epidemiologic research (WONDER). 2022.
5. Hedegaard H, Bastian B, Trinidad J, Spencer M, Warner M. Drugs most frequently involved in drug overdose deaths: United States, 2011–2016. 2018.
6. Kolodny A, Courtwright DT, Hwang CS, Kreiner P, Eadie JL, Clark TW, et al. The prescription opioid and heroin crisis: a public health approach to an epidemic of addiction. *Annu Rev Public Health*. 2015;36:559–74. <https://doi.org/10.1146/annurev-publhealth-031914-122957> PMID: 25581144
7. Naeger S, Mutter R, Ali MM, Mark T, Hughey L. Post-discharge treatment engagement among patients with an opioid-use disorder. *J Subst Abuse Treat*. 2016;69:64–71. <https://doi.org/10.1016/j.jsat.2016.07.004> PMID: 27568512
8. Yangchen T, Rodriguez M, Baird J, Hallowell BD, Daly MM, Berk J, et al. Trends in recurrent overdose and treatment initiation following emergency department visits for opioid overdose between 2016 and 2021. *Drug Alcohol Depend*. 2024;262:111379. <https://doi.org/10.1016/j.drugalcdep.2024.111379> PMID: 38968835
9. Tseregounis IE, Henry SG. Assessing opioid overdose risk: a review of clinical prediction models utilizing patient-level data. *Transl Res*. 2021;234:74–87. <https://doi.org/10.1016/j.trsl.2021.03.012> PMID: 33762186
10. Boyes AP. Repetition of overdose: a retrospective 5-year study. *J Adv Nurs*. 1994;20(3):462–8. <https://doi.org/10.1111/j.1365-2648.1994.tb02382.x> PMID: 7963051
11. Coffin PO, Tracy M, Bucciarelli A, Ompad D, Vlahov D, Galea S. Identifying injection drug users at risk of nonfatal overdose. *Acad Emerg Med*. 2007;14(7):616–23. <https://doi.org/10.1197/j.aem.2007.04.005> PMID: 17554010
12. Caudarella A, Dong H, Milloy MJ, Kerr T, Wood E, Hayashi K. Non-fatal overdose as a risk factor for subsequent fatal overdose among people who inject drugs. *Drug Alcohol Depend*. 2016;162:51–5. <https://doi.org/10.1016/j.drugalcdep.2016.02.024> PMID: 26993373
13. Langabeer JR, Chambers KA, Cardenas-Turanzas M, Champagne-Langabeer T. County-level factors underlying opioid mortality in the United States. *Subst Abus*. 2022;43(1):76–82. <https://doi.org/10.1080/08897077.2020.1740379> PMID: 32186475
14. Brett AS, Rothschild N, Gray R, Perry M. Predicting the clinical course in intentional drug overdose. Implications for use of the intensive care unit. *Arch Intern Med*. 1987;147(1):133–7. <https://doi.org/10.1001/archinte.1987.00370010131028> PMID: 3800515
15. Christenson J, Etherington J, Grafstein E, Innes G, Pennington S, Wanger K, et al. Early discharge of patients with presumed opioid overdose: development of a clinical prediction rule. *Acad Emerg Med*. 2000;7(10):1110–8. <https://doi.org/10.1111/j.1553-2712.2000.tb01260.x> PMID: 11015242
16. Glanz JM, Narwaney KJ, Mueller SR, Gardner EM, Calcaterra SL, Xu S, et al. Prediction model for two-year risk of opioid overdose among patients prescribed chronic opioid therapy. *J Gen Intern Med*. 2018;33(10):1646–53. <https://doi.org/10.1007/s11606-017-4288-3> PMID: 29380216
17. Sumetsky N, Mair C, Wheeler-Martin K, Cerda M, Waller LA, Ponicki WR, et al. Predicting the future course of opioid overdose mortality: an example from two US States. *Epidemiology*. 2021;32(1):61–9. <https://doi.org/10.1097/EDE.0000000000001264> PMID: 33002963
18. Fox J, Weisberg S. Cox proportional-hazards regression for survival data. An R and S-PLUS companion to applied regression. 2002.
19. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Statist*. 1982;10(4). <https://doi.org/10.1214/aos/1176345976>
20. Keen C, Young JT, Borschmann R, Kinner SA. Non-fatal drug overdose after release from prison: a prospective data linkage study. *Drug Alcohol Depend*. 2020;206:107707. <https://doi.org/10.1016/j.drugalcdep.2019.107707> PMID: 31757517
21. Keen C, Kinner SA, Borschmann R, Young JT. Comparing the predictive capability of self-report and medically-verified non-fatal overdose in adults released from prison: A prospective data linkage study. *Drug Alcohol Depend*. 2020;206:107742. <https://doi.org/10.1016/j.drugalcdep.2019.107742> PMID: 31778949
22. Yang H, Pan F, Tong D, Brown HE, Liu J. Measurement error-tolerant Poisson regression for Valley Fever incidence prediction. In: *IISE transactions on healthcare systems engineering*. 2024; p. 1–13.
23. Romeiser J, Labriola J, Meliker J. Geographic patterns of prescription opioids and opioid overdose deaths in New York State, 2013–2015. *Drug and Alcohol Dependence*. 2019;195:94–100.

24. Xia S, Wung S-F, Chen C-C, Coompon JLK, Roveda J, Liu J. Data-fusion-based quality enhancement for HR measurements collected by wearable sensors. *Sensors (Basel)*. 2024;24(10):2970. <https://doi.org/10.3390/s24102970> PMID: 38793825
25. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*. 2012;13(6):395–405. <https://doi.org/10.1038/nrg3208> PMID: 22549152
26. Zhao S, Browning J, Cui Y, Wang J. Using machine learning to classify patients on opioid use. *J Pharm Health Serv Res*. 2021;12(4):502–8. <https://doi.org/10.1093/jphsr/rmab055> PMID: 35003334
27. Badger J, LaRose E, Mayer J, Bashiri F, Page D, Peissig P. Machine learning for phenotyping opioid overdose events. *J Biomed Inform*. 2019;94:103185. <https://doi.org/10.1016/j.jbi.2019.103185> PMID: 31028874
28. Mullin S, Zola J, Lee R, Hu J, MacKenzie B, Brickman A, et al. Longitudinal K-means approaches to clustering and analyzing EHR opioid use trajectories for clinical subtypes. *J Biomed Inform*. 2021;122:103889. <https://doi.org/10.1016/j.jbi.2021.103889> PMID: 34411708
29. Dong X, Deng J, Hou W, Rashidian S, Rosenthal RN, Saltz M, et al. Predicting opioid overdose risk of patients with opioid prescriptions using electronic health records based on temporal deep learning. *J Biomed Inform*. 2021;116:103725. <https://doi.org/10.1016/j.jbi.2021.103725> PMID: 33711546
30. Hawkes AG. Point spectra of some mutually exciting point processes. *J Roy Statist Soc Ser B: Statist Methodol*. 1971;33(3):438–43. <https://doi.org/10.1111/j.2517-6161.1971.tb01530.x>
31. Ogata Y. Space-time point-process models for earthquake occurrences. *Annals Inst Statist Math*. 1998;50(2):379–402. <https://doi.org/10.1023/a:1003403601725>
32. Mohler G. Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *Int J Forecast*. 2014;30(3):491–7. <https://doi.org/10.1016/j.ijforecast.2014.01.004>
33. Rizoia MA, Mishra S, Kong Q, Carman M, Xie L. SIR-Hawkes: Linking epidemic models and Hawkes processes to model diffusions in finite populations. In: *Proceedings of the 2018 World Wide Web Conference*. 2018;2018:419–28.
34. Reynaud-Bouret P, Rivoirard V, Tuleau-Malot C. Inference of functional connectivity in Neurosciences via Hawkes processes. In: *2013 IEEE Global Conference on Signal and Information Processing*. 2013:317–20. <https://doi.org/10.1109/globalsip.2013.6736879>
35. Pan F, Zhang Y, Liu J, Head L, Elli M, Alvarez I. Reliability modeling for perception systems in autonomous vehicles: A recursive event-triggering point process approach. *Transp Res Part C: Emerg Technol*. 2024;169:104868. <https://doi.org/10.1016/j.trc.2024.104868>
36. Xu L, Duan JA, Whinston A. Path to purchase: a mutually exciting point process model for online advertising and conversion. *Manag Sci*. 2014;60(6):1392–412. <https://doi.org/10.1287/mnsc.2014.1952>
37. Brandt L, Hu M-C, Liu Y, Castillo F, Odom GJ, Balise RR, et al. Risk of experiencing an overdose event for patients undergoing treatment with medication for opioid use disorder. *Am J Psychiatry*. 2023;180(5):386–94. <https://doi.org/10.1176/appi.ajp.20220312> PMID: 36891640
38. Park TW, Lin LA, Hosanagar A, Kogowski A, Paige K, Bohnert ASB. Understanding risk factors for opioid overdose in clinical populations to inform treatment and policy. *J Addict Med*. 2016;10(6):369–81. <https://doi.org/10.1097/ADM.0000000000000245> PMID: 27525471
39. Cox DR. Regression models and life-tables. *J Roy Statist Soc*. 1972.
40. Prentice RL, Self SG. Asymptotic distribution theory for cox-type regression models with general relative risk form. *Ann Statist*. 1983;11(3). <https://doi.org/10.1214/aos/1176346247>
41. Choi E, Du N, Chen R, Song L, Sun J. Constructing disease network and temporal progression model via context-sensitive Hawkes process. In: *2015 IEEE International Conference on Data Mining*. 2015:721–6.
42. van Lieshout M-CNM. On estimation of the intensity function of a point process. *Methodol Comput Appl Probab*. 2011;14(3):567–78. <https://doi.org/10.1007/s11009-011-9244-9>
43. Min J, Hong Y, King CB, Meeker WQ. Reliability analysis of artificial intelligence systems using recurrent events data from autonomous vehicles. *J Roy Statist Soc Ser C: Appl Statist*. 2022;71(4):987–1013. <https://doi.org/10.1111/rssc.12564>
44. Pan F, Zhang Y, Head L, Liu J, Elli M, Alvarez I. Quantifying error propagation in multi-stage perception system of autonomous vehicles via physics-based simulation. In: *2022 Winter Simulation Conference (WSC)*. IEEE; 2022. p. 2511–22.
45. Ryan CW. An inhomogeneous Poisson model of the spatial intensity of opioid overdoses, as represented by EMS naloxone use. *bioRxiv*. 2019:529206.