# Illuminating the Hidden Elements and Future Evolution of Opioid Abuse Using Dynamic Modeling, Big Data and Particle Markov Chain Monte Carlo

Xiaoyan Li, Bryce Keeler, Rifat Zahan, Lujie Duan, Anahita Safarishahrbijari, Jason Goertzen, Yuan Tian, Juxin Liu, and Nathaniel Osgood

University of Saskatchewan, Saskatoon, SK, Canada
xiaoyan.li@usask.ca, keeler.bryce@usask.ca, rifat.zahan@usask.ca, lujie.duan@usask.ca, anahita.safari@usask.ca
jason.goertzen@usask.ca, yut473@mail.usask.ca, juxin.liu@usask.ca, osgood@cs.usask.ca

**Abstract.** Annual deaths from opioid overdoses have increased continuously over the past 20 years in the United States. This modern plague has precipitated serious adverse social impacts at both the individual and social level. Estimating and predicting the opioid epidemic trend is helpful for public health agencies making intervention policies but is greatly complicated by the rapid evolution of the epidemic and the scattered, cross-sectoral nature of the relevant evidence. Using Cincinnati as a case study, we incorporate the machine learning algorithm of Particle Markov Chain Monte Carlo with a multiply stratified SD modeling of opioid dynamics, traditional reported demographic and health data, and search volume data from Google Trends to provide a framework to support rapid learning from the evolving epidemic. The results show that the PMCMC model can readily incorporate diverse unfolding lines of cross-sectoral evidence to estimate the evolving underlying system state, key parameters, and predict opioid epidemic trends. Finally, we demonstrate how the model can be used to evaluate the impact of an intervention strategy to decrease opioid-related deaths and overdoses.

**Keywords:** Particle Markov chain Monte Carlo (PMCMC) · Opioid · Particle Filtering.

## 1 Introduction

Annual deaths resulting from drug overdoses have increased continuously over the past 20 years in the United States [2]. As a case in point, while there have been more than 350,000 opioid overdose deaths between 1999 and 2016, the number of opioid overdoses was five times higher in 2016 than in 1999 [2]. This modern plague has precipitated serious adverse social impacts at both the individual and social level [2, 4], including the deterioration of physical health, the breakdown of families, and burgeoning health-care costs [4, 5].

Dynamic models can serve as potent tools for learning about complex systems [15]. In recent years, System Dynamics (SD) models have become popular for addressing opioid abuse in the public health domain. Several studies using System Dynamics modelling have been conducted so far to address the scope of opioid abuse, and to identify high-level policy interventions which might reduce the negative consequences of the opioid crisis (including deaths resulting from overdose) [9, 14, 18, 17, 16]. While much of this work is promising, existing dynamic models are hampered by lack of availability of certain key elements of empirical data [14]. Researchers have particularly suggested that securing additional data related to opioid abuse would enhance the validity of one recent dynamic model and will help to understand the dynamic nature of such complex system [14]. Given the rapidly shifting nature of the opioid crisis, it is also notable that such models are hampered by their use of data from a snapshot in time, rapid obsolescence because of technological, societal and health system change, and an inability to incorporate new evidence. In this research, we seek to provide a framework to inform policy discourse that addresses key gaps above, namely the need to 1) Fill data gaps of the system. 2) Enhance understanding of the past and current status of the opioid epidemic. 3) Anticipate coming trends so as to leverage responsive resources. 4) Perhaps most importantly, support learning faster and more reliably in light of the fast-moving nature of the epidemic. Within this work, we address such needs by providing a framework that can, on an ongoing basis as data arrives, collectively leverage insights from model structure and diverse, cross-sectoral data sources to probablistically estimate the current status of the entire system (including latent factors). On the basis of that probabilistically estimated current state, the framework is capable of projecting that state forward and examining the impact of alternative intervention options. This framework leverages both traditional data sources and big data (including from municipal sources and online communicational data), machine learning (in the form of Particle Markov Chain Monte Carlo techniques) and dynamic modeling (capturing underlying theory concerning the structure of the system). The remainder of this paper is as follows: Within the next section, we discuss the methods employed, including PMCMC, the dynamic model design, and scenarios conducted. The penultimate section discusses model findings. The paper closes with synthesis of modeling project contributions, and notes model limitations.

## 2 Methodology

In this project, two crucial methods are employed – An SD population model representing opioid administration, use and abuse, and the Particle Markov Chain Monte Carlo (PMCMC) machine learning method. The PMCMC method is used to link the SD model to a broad set of empirical data resources from Cincinnati, Ohio. The data used for this model includes values used for model parameter estimates, as well as time series drawn from traditional sources (such as that related to monthly mortality and prescription levels), municipal open data repositories, and from online communicational behaviour. This combination provides a framework for ongoing rapid learning in light of the unfolding availability of evidence in time series, and supports recurrent estimation of the state of the entire system based on collective implications of the model structure together with the data. As outlined below, this combination further allows for the pursuit of policy analysis and scenario projection in light of this recurrently refreshed evidence.

### 2.1 Model Structure

This project used SD modelling to capture the dynamic behavior of the complex opioid system. Readers interested in the detailed structure and mathematical specification of the opioid SD model to be employed in this project are referred to the supplementary file. To capture the behaviour of population in key categories of individuals associated with distinct dynamics, the stocks in this SD model (Fig. **??** in the supplemental information) are further stratified into distinct groups; specifically, most stocks are stratified into three dimensions. Given the central role of chronic pain management as an entry into initiation and maintenance of opioid use, the status of pain is divided into with chronic pain and without chronic pain. So as to capture the escalation in tolerance upon chronic use, the high risk of disorder among those with high tolerance, and the high risk of overdose for those with low tolerance, the population is further divided into three categories: low, medium and high tolerance. Finally, the dimension of disorder history captures whether the included persons have a history of opioid use disorder or not. This final type of stratification is motivated by the fact that it has a profound impact on risk of occurrence (or recurrence) of opioid dependence, will aid in the understanding of model dynamics, and may prove useful for comparison against empirical data. In light of the rapidly evolving nature of the opioid epidemic, it is further notable that the SD model allowed for stochastic evolution with respect to certain factors that might otherwise be characterized as parameters – factors such as the risk associated with illicit opioids, or the hazard rate with which those travel from prescription sources to dealer sourcing.

### 2.2 Algorithm: PMCMC

In order to make the opioid SD model more robust and to support the model in learning from one or more incoming empirical time series, the PMCMC method [1] is employed to recurrently reground this model using such data. PMCMC is a modern methodology that fits into the broader statistical filtering tradition. As time passes, the method combines estimates generated by a dynamic model (recognized as fallible) with one or more types of arriving empirical observations (recognized as noisy), yielding a consensus posterior distribution (here, captured by sampling from a joint distribution over the values for all parameters and the entire state of the SD model over time). For each single data set, only a part of an underlying dynamic model is informed (grounded). However, by utilizing a large number of datasets, with each capturing the character of different parts, and in light of the structure and logic of the model, the data collectively illuminate even latent areas of the model. This reflects the fact that, because the different areas of the model are coupled, and because the nature of the coupling is well-defined, data about one part of the model will generally provide considerable information regarding closely coupled areas of that model, or at least constrains an understanding as to the situation in those coupled regions. With sufficient arriving empirical data, underlying state of the opioid dynamic system can be estimated with increased certainty – including latent components not directly illuminated by specific data items. The PMCMC algorithm is well specified in the seminal work of Andreiu et al. [1], as is the Marginal Metropolis-Hastings (MMH) variant of that algorithm specifically employed here. Due to space constraints, we confine our coverage here to a very high-level specification of this algorithm, noting how it was adapted to address the specific needs of our dynamic model. The MMH PMCMC algorithm samples from a joint distribution of parameter values and model state over time (for our model, the latter represent trajectories over time in terms of the state of the SD model). It accomplishes this by alternating between a stage in which it samples parameter values conditional on the previously sampled trajectory of the model using the random walk Metropolis Hastings Markov Chain Monte Carlo algorithm [6], and a stage in which it performs the Sequential Monte Carlo algorithm of Particle Filtering on the state equation model (here, the SD model). For our case, the application of particle filtering associated with the latter stage is very similar to that described in [12], with the exception of the fact that resampling occurs following each observation except the last. In addition, particle filtering is followed by sampling a complete particle lineage (spanning across each previous observation point) from the set of lineages of all particles, with the probability of picking a given lineage being proportional to the weight associated with each particle at the final timepoint [1].

### 2.3 Data Description

To use with in conjunction with the opioid PMCMC SD model, this work drew on a large and diverse set of empirical datasets related to the geographic region of interest – Cincinnati, Ohio, and the enclosing Hamilton county. These empirical datasets could be generally divided into two categories: surveillance and demographic time series from the local and state public departments, and a large amount of online search behavior information, most notably search volume time series related to opioids and opioid misuse. These two categories are introduced and listed as follows. The surveillance and demographic datasets have relatively high quality and standardization, and are as follows:

- The time series count of EMS responses to cases of Heroin overdoses, as reported by the Cincinnati Fire Department and the Cincinnati Police Department [3].
- The time series count of Police Calls for Service related to drug complaints, as reported by the Cincinnati Police Department [13].

- The yearly number of Hamilton county drug overdose deaths, as reported by the Ohio Department of Health [11].
- The monthly number of Hamilton county drug overdose deaths, as publicly reported by the *WONDER* system of the United States Centres for Disease Control and Prevention [19]. Monthly time series data for the years 2015 and 2016 on drug-induced overdose deaths were obtained, including overdose deaths specific to unintentional classified as suicide (ICD-11 codes X40-X44), intentional suicide (X60-X64), homicide (X85) and undetermined causes (Y10-Y14).
- The quarterly Hamilton county report of opioid prescriptions [10].

While the datasets above provide high-quality evidence related to important processes and elements of the broader opioid system, they offer less detail with respect to certain factors, including occurrence of chronic pain, latent abuse of prescription opioids during active prescriptions, and use of drug rehabilitation services. They further provide limited resolution related to the complex factors of illicit drug abuse, which is a key driver of overdose deaths. To better illuminate such factors, we additionally made use of four Google Trends [7] weekly time series. Specifically, we downloaded weekly datasets from google trends, using that service to report relative search counts for either a certain search term or for a topic for Cincinnati, Ohio. The specific searches whose volumes were studied were for Back pain (search topic), Drug rehabilitation (search topic), Dark web (search topic), and Naloxone (search Medication). It is noted that data obtained from Google trends are scaled to lie within the range of 0 to 100 for a particular topic [7]. This represents relative search volume for a topic within a geographically defined region. Data with respect to such online communicational behaviour lack the quality and standardization of the surveillance and demographic data. However, they still provide a valuable independent source of reasonably high temporal resolution information to ground the values of the stocks and parameters in the PMCMC opioid model. Our conviction in the value of such information borne of success other lines of our work that have demonstrated that even low-quality datasets offering distinct information can confer high value in grounding dynamic models using particle filtering.

### 2.4 Combination of Dynamic Model and Datasets

The model combined each empirical datum with corresponding model elements according to likelihood functions described in Table 2 of the Supplemental Information.

### 2.5 Scenarios

The work here used the model characterized above to execute several scenarios. In the baseline scenario, the PMCMC algorithm was simulated over the whole timeframe of the opioid SD model, incorporating the traditional public health surveillance empirical datasets and the social media datasets (from Google

Trend), with ongoing learning (via PMCMC) from past and current opioid epidemic trends. Specifically, we run 30 chains with 1000 particles, and each chain with 2000 iterations. In this case, we set our burn-in period as the initial 1000 iterations. For each iteration, we recorded dichotomous acceptance, and jointly sampled values of parameters and state variables. Then, we have run four prediction experiments. To support evaluation of the predictive accuracy of the model, the prediction was undertaken for time periods at which data was available, but during the prediction experiments, that data was not used within the prediction algorithms. Specifically, prediction started from week 61, week 81, week 101 and week 121.

## 3 Results

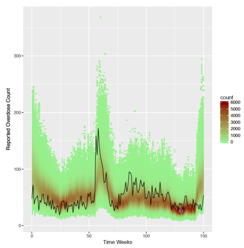
The results of the baseline scenario are shown in Fig.1, which compares the 2D histogram plot of the model result and the line plot of the empirical datasets. It indicates that, informed by data the model could broadly effectively estimate the epidemic trend of Cincinnati during the past three years. Beyond the baseline scenario, we tested the predictive accuracy of the model. Fig.2 depicts the prediction and intervention results. While formal evaluation is not conducted here, comparison of the prediction results with estimation results using empirical data suggests that the model offers significant predictive accuracy over the span of months; however, the stochastics within the model capture the evolving character of the opioid crisis, and make infeasible tightly bounded predictions over the long term. Finally, as a demonstration of the ability of the framework to support intervention investigation, we propose here an intervention strategy – a harm reduction providing the disordered individuals with access to legal opioids. This strategy could prevent these disordered persons seeking illegal opioid dealer sources, which is more dangerous and harmful to their health due to strong variability in dosing, and the potential for intermittent availability. The results of intervention suggests that this intervention strategy could decrease the death population related to opioid and overdose count effectively.

## 4 Discussion and Future Work

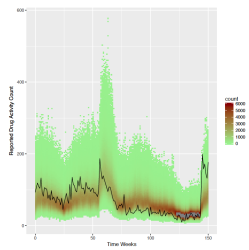
This research project has contributed several substantial findings. A methodological contribution more generally in modeling is that, in light of the fact that PMCMC allows for sampling from a joint distribution between model parameters and latent state of the model over time-based on incoming data, the model will further afford data-informed probabilistic projection of the evolution of the opioid epidemic – supporting rapid learning from shifts in a deadly and fast-moving epidemic. The projection will rigorously capture uncertainty associated with estimates of both model parameters and latent model state, in a fashion that is rarely possible within dynamic modeling sphere. This project may be the first contribution exploring leveraging the cutting-edge machine learning algorithm PMCMC with dynamic modeling in the health or social domains, and one

Fig. 1: The comparison of the results between the model output and empirical datasets

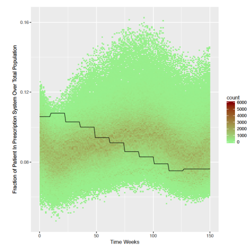
(a) Overdose



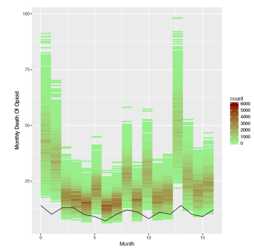
(b) Drug activity



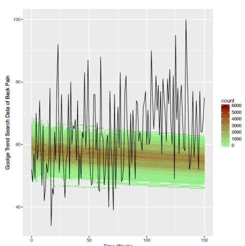
(c) Opioid prescription



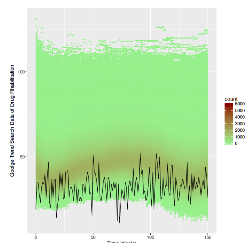
(d) Opioid deaths



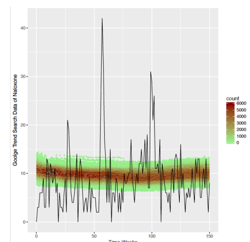
(e) Back pain



(f) Drug rehabilitation



(g) Naloxone



(h) Dark Web

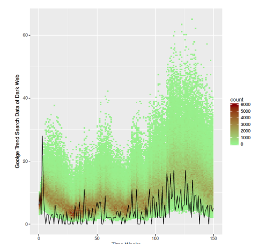
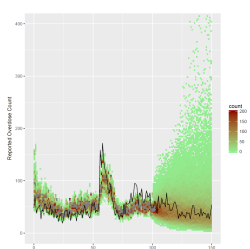
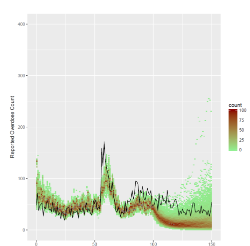


Fig. 2: The prediction and intervention results

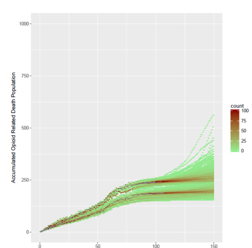
(a) Predicted overdose from $t = 101$



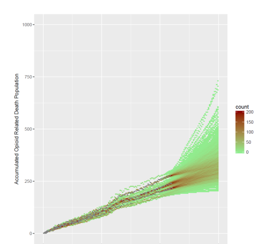
(b) Intervention overdose from $t = 101$



(c) Intervention deaths from $t = 81$



(d) Intervention deaths from $t = 101$



of very few examples in other domains. Thirdly, based on a review of the literature, there are very few researchers who have incorporated so many datasets within a PMCMC model (and, per the above, none for dynamic modeling). Therefore, this project offered a good opportunity to assess the performance of the powerful, contemporary PMCMC model when combining a relatively large number of empirical datasets, including social media data which generally have a high noise and uncertainty. An advantage in terms of policy planning reflects the fact that the model will support probabilistic projections of policy tradeoffs, for the Cincinnati context, in a fashion that takes into account uncertainties regarding the latent state of the model.

Despite the advantages noted above, the work here suffers from an important set of limitations. Firstly, the convergence of the PMCMC to the asymptotic distribution associated with the underlying MCMC needs to be assessed more rigorously – such as with the Heidelberg and Welch convergence tests [8], or such as that proposed by Gelman involving multiple walkers [6]; it is likely that additional iterations will be necessary to achieve effective convergence. Secondly, some of the model parameters that are supported by smaller bodies of evidence but are assumed here as fixed (e.g., the rate of disordered individuals transit to treatment per unit time) or given placeholder values (e.g., rates of evolution of the two stochastic parameters) might be additionally informed by values in the literature which were not considered, or (lacking appropriate estimates) better sampled via PMCMC. Thirdly, the model would benefit greatly from additional data sources in key areas, such as those related to prescription dynamics, from police intelligence related to the drug trade, and from drug rehabilitation service-seeking. Fourthly, the model estimates could be enhanced by including data from additional sources of online behaviour, most notably time series extracted from validated tweets and (particularly) dark web-related activity.

While the findings presented here are substantial, the current work offers many lines of natural extension. In terms of refinement of the model, a high priority should be placed on including some poorly-evidenced parameters as part of the vector of parameters sampled in each iteration. As a contrast to positioning a fixed constant value for such parameters, specifying a prior distribution of possible values in which the values of those parameters are informed by incoming data may allow for a model that better accounts for the empirical data observed. An alternative point of expansion would operate within the sphere of the current model, and would center on enhancing the policy repertoire to be examined in the model. Policy questions of particular interest for investigation include broadening naloxone (Narcan) availability, harm-reduction strategies centered on supervised injection as well as heroin prescriptions for stabilization, impact of promotion of alternative pain-management strategies, changes related to cannabis legalization, and enhanced monitoring of disordered individuals concurrently drawing on multiple providers. A separate line of work would focus on including social media data, with Twitter being an obvious source of evidence for high-temporal resolution data for the Cincinnati area. Data to be examined here include time series of counts of key phrase and hash-tag mentions, and possibly counts resulting from machine-learning based classification of tweet contents. While we examined and deferred consideration of Twitter counts due to lack of an appropriate historic archive of Twitter data that include this geographic region, we have begun archiving tweets from that region. We are confident that, between this archive and other archives that could be available via collaborators. Twitter could serve as an important source of evidence, particularly with respect to alternative management strategies for chronic pain and comments on levels of illicit drug use and diversion. Finally, we believe that a variant of the model created here that was adapted instead to the context of a single-payer system could allow for informing the model with additional lines of population-wide evidence – and thereby support the deduction of narrow ranges of parameter values that would be transferable to the Cincinnati context.

## 5 Conclusion

We have demonstrated here a promising framework supporting rapid learning from diverse lines of emerging evidence related to the opioid crisis. By leveraging a wide variety of evidence and theory as captured in a dynamic model, this approach provides a rigorous means of estimating on an ongoing basis the state of the underlying system – even areas of the system on which little direct data is available, and of probabilistically anticipating future trends and evaluating intervention trade-offs in a way that leverages the latest evidence. This capacity for the model to adapt and learn from emerging evidence – evidence automatically incorporated into the model – is of vital importance given the large uncertainties still associated with the processes underlying opioid epidemic, its cross-sectoral and pervasive elements, and the fast-shifting nature of its dynamics. While dynamic models such as that introduced here and by past contributors [9, 14, 18, 17, 16] offer much value, and while machine learning techniques applied to lines of evidence can yield great insight, we posit that the capacity enabled here to learn quickly from emerging evidence and to update models and understanding accordingly will provide far more gain than would be secured by any one or approach.

**Acknowledgement.** This research was enabled in part by support provided by the Saskatchewan Ministry of Policing and Corrections, Westgrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca), and the National Science and Engineering Research Council of Canada.

## References

1. Andrieu, C., Doucet, A., Holenstein, R.: Particle markov chain monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **72**(3), 269–342 (2010)
2. Understanding the epidemic, centers for disease control and prevention. https://www.cdc.gov/drugoverdose/epidemic/index.html

3. Cincinnati Fire Incidents (cad), City of Cincinnati. https://data.cincinnati-oh.gov/Safer-Streets/Cincinnati-Fire-Incidents-CAD-including-EMS-ALS-BL/vnsz-a3wp
4. Fiellin, D.A., Rosenheck, R.A., Kosten, T.R.: Office-based treatment for opioid dependence: reaching new patient populations. American Journal of Psychiatry **158**(8), 1200–1204 (2001)
5. Florence, C.S., Zhou, C., Luo, F., Xu, L.: The economic burden of prescription opioid overdose, abuse, and dependence in the united states, 2013. Medical care **54**(10), 901–906 (2016)
6. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian data analysis. Chapman and Hall/CRC (1995)
7. Google trends. https://www.google.com/trends
8. Heidelberger, P., Welch, P.D.: Simulation run length control in the presence of an initial transient. Operations Research **31**(6), 1109–1144 (1983)
9. Nielsen, A., Wakeland, W., Zimam, A.: A public health model for simulating policy interventions to reduce nonmedical opioid use. In: Simulation and Modeling Methodologies, Technologies and Applications, pp. 239–254. Springer (2015)
10. Ohio Automated Rx Reporting System, Ohio Board of Pharmacy. https://www.ohiopmp.gov/County.aspx
11. Ohio public health data warehouse, Ohio department of health. http://publicapps.odh.ohio.gov/EDW/DataBrowser/Browse/Mortality
12. Osgood, N., Liu, J.: Towards closed loop modeling: evaluatng the prospects for creating recurrently regrounded aggregate simulation models using particle filtering. In: Proceedings of the 2014 Winter Simulation Conference. pp. 829–841. IEEE Press (2014)
13. PDI (police data initiative) police calls for service (cad), city of cincinnati. https://data.cincinnati-oh.gov/Safer-Streets/PDI-Police-Data-Initiative-Police-Calls-for-Servic/gexm-h6bt
14. Schmidt, T.D., Haddox, J.D., Nielsen, A.E., Wakeland, W., Fitzgerald, J.: Key data gaps regarding the public health issues associated with opioid analgesics. The Journal of Behavioral Health Services & Research **42**(4), 540–553 (2015)
15. Sterman, J.D.: Learning from evidence in a complex world. American Journal of Public Health **96**(3), 505–514 (2006)
16. Wakeland, W., Nielsen, A., Schmidt, T.D.: Gaining policy insight with a system dynamics model of pain medicine prescribing, diversion and abuse. Systems Research and Behavioral Science **33**(3), 400–412 (2016)
17. Wakeland, W., Schmidt, T., Gilson, A.M., Haddox, J.D., Webster, L.R.: System dynamics modeling as a potentially useful tool in analyzing mitigation strategies to reduce overdose deaths associated with pharmaceutical opioid treatment of chronic pain. Pain Medicine **12**(suppl_2), S49–S58 (2011)
18. Wakeland, W.W., Schmidt, T.D., Haddox, J.D.: A system dynamics model of pharmaceutical opioids: medical use diversion and nonmedical use. In: 29th International Conference of the System Dynamics Society. Washington, DC. Citeseer (2011)
19. About underlying cause of death, 1999 - 2016. https://wonder.cdc.gov/ucd-icd10.html