Early Detection of Opioid Over-Procurement: A Machine Learning Approach

Dimitris Bertsimas, Mohammad M. Fazel-Zarandi, Joshua Ivanhoe
MIT Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA
{dbertsim, fazel, jivanhoe}@mit.edu

Over the past three decades, the misuse of prescription opioids has been the leading cause of the decline in life expectancy in the United States. In this study, we present machine learning models to identify the illegal diversion of controlled substances across the prescription drug supply chain in the US. The models use a combination of state-of-the-art machine learning algorithms and a vast volume of administrative data to discover suspicious opioid purchasing behavior. We show that the algorithms can detect illicit opioid transactions with high accuracy, and are capable of early detection up to several years in advance. We estimate that the implementation of our machine learning models could have prevented the illegal distribution of as many as 600 million pills between 2006 and 2012. Our findings hold promise for using data-driven machine learning algorithms to detect, prevent, and investigate the diversion of prescription opioids.

Key words: opioid crisis, machine learning, early detection

1. Introduction

Prescription opioid abuse is a severe public health crisis in modern United States history. Escalating from the late 1990s, it has resulted in the death of more than 400,000 people, exceeding those attributed to guns or car accidents during the same period (US Government Accountability Office (2020)), and has contributed to the first decline in US life expectancy since the height of the AIDS epidemic in 1993 (Scholl et al. (2018)). The first decade of the twenty-first century saw a dramatic increase in the prescription and consumption of opioids, notably Oxycodone and Hydrocodone, with the nationwide per capita distribution of these drugs more than doubling over this decade (Kenan et al. (2012)). By 2013, the yearly cost attributable to the crisis, due to healthcare and criminal justice expenditure, as well as reduced productivity, was estimated at almost $80 billion (Brownstein et al. (2010)).

In this paper, we present machine learning methods to identify the illegal diversion of controlled substances by licensed, consumer-facing buyers, such as pharmacies and doctors. Our models use state-of-the-art machine learning algorithms and vast amounts of administrative data to discover complex patterns in the transactions of controlled substances across the prescription drug supply chain in the US. The models utilize these patterns to predict and flag suspicious pharmacies and doctors diverting prescription opioids into the illegal marketplace.

We show that our machine learning algorithms can detect suspicious opioid buyers engaged in improper purchasing behavior with high accuracy (achieving AUC scores of over 0.97), and are capable of early violation detection up to several years in advance. Overall, we estimate that our models could have prevented the illegal distribution of hundreds of millions of pills. We note that the development of our machine learning algorithms to detect illicit drug transactions is timely, as recommended by a recent US Government Accountability Office report to the Department of Justice (US Government Accountability Office (2020)): "The DEA Administrator should develop and implement additional ways to use algorithms in analyzing ARCOS and other data to more proactively identify problematic drug transaction patterns."

Prior research has primarily been concerned with using machine learning to identify patient risk factors for opioid abuse. These studies integrate patient-level data, such as electronic health records, with machine learning techniques to predict opioid dependency (Hastings et al. (2020), Ellis et al. (2019), Lo-Ciganic et al. (2019), Prieto et al. (2020), Dufour et al. (2014), Hylan et al. (2015), Han et al. (2020)). On the supply side, Mackey et al. (2017) combines Twitter data with machine learning to identify the marketing of illicit opioids by online sellers. Unlike the papers mentioned above, we present machine learning algorithms to detect illegal diversion of controlled substances within the US prescription drug supply chain. To our knowledge, this is the first paper to develop machine learning models to predict and flag suspicious supply chain entities diverting prescription opioids into the market.

The paper proceeds as follows. Section 2 discusses the data sources used in the analysis. Section 3 describes our methodology. Section 4 presents the results of the study. Section 5 concludes the paper.

2. Data

ARCOS Dataset

A key data source that we utilize in our study is the Automated Reports and Consolidated Ordering System (ARCOS), a Drug Enforcement Agency (DEA) database that records transactions of all controlled substances in the US. The data includes transaction records between reporting entities, consisting of opioid sellers such as manufacturers and distributors, and buyers, predominantly pharmacies and doctors. Each record contains information about these parties and the specific medication being supplied, including the generic drug name (e.g., oxycodone) and dosage units (e.g., number of pills). The dataset consists of almost 380 million transactions in total, representing the flow of just over 76.5 billion pills from their point of manufacture to their point of sale between January 2006 and December 2012.[1]

For our analysis, we focus on the combined dosage units of oxycodone and hydrocodone, aggregated by buyer and month of transaction. For brevity, we refer to this dataset from hereon as the *buyer timeseries*. Summary statistics from the buyer timerseries are summarized in Table 1, which demonstrates a consistent year-on-year increase in the total dosage units per capita of oxycodone and hydrocodone procured by registered buyers, from 2007 to 2011. Aggregate statistics by state are also provided in Table 3 in the Appendix. To obtain high-level information about the nation-wide supply chain, we also combine the total number of transactions between every reporter-buyer pair, and use this as a proxy for the *adjacency matrix* of the network.

Supervisory Labels

As the ARCOS database exists for reporting purposes, information about transactions being flagged by the DEA as suspicious is not included. Accordingly, no ground-truth *labels*—binary indicators

[1] Data are available at: https://wpinvestigative.github.io/arcos/.

Year	Active Buyers	Total Dosage Units	Dosage Units Per Captia	Dosage Units Per Buyer
2006	90,092	8.56E+09	28.7	94,976
2007	89,024 (-1%)	9.63E+09 (13%)	32.0 (11%)	108,172 (14%)
2008	80,329 (-10%)	1.04E+10 (8%)	34.2 (7%)	129,474 (20%)
2009	85,653 (7%)	1.13E+10 (9%)	37.5 (10%)	131,942 (2%)
2010	86,986 (2%)	1.21E+10 (7%)	39.8 (6%)	139,176 (5%)
2011	87,093 (0%)	1.29E+10 (7%)	42.2 (6%)	148,520 (7%)
2012	89,459 (3%)	1.28E+10 (-1%)	41.5 (-2%)	143,403 (-3%)

Table 1 Yearly nationwide ARCOS transaction statistics for oxycodone and hydrocodone combined. Percentage year-to-year changes are shown in brackets.

of whether an opioid buyer engaged in improper opioid purchasing behavior—can be extracted directly from this data. To obtain such labels, we focus on Walgreens, a chain pharmacy that has faced considerable legal and media scrutiny as the largest retailer of opioids from 2006 to 2012. During this period, the company distributed approximately 12.9 billion pills across 7,437 unique buyer accounts in the ARCOS database. This represents a 30% increase in volume relative to Walgreens' largest competitor, CVS, which distributed roughly 9.9 billion pills over the same time frame.

Using various sources, ranging from legal proceedings to investigative reports (see, e.g., United States Department of Justice (2013), United States District Court (2020), and Abelson et al. (2018)) we are able to flag 308 Walgreens buyers listed in the ARCOS database as potential bad actors in the distribution of opioids to the public, which we use to produce our labels. As an example, in United States Department of Justice (2013), a Department of Justice communiqué, six Walgreens stores in Florida are explicitly named in a civil penalty investigation that results in an $80 million fine being leveled against the company. In a more nuanced example, a report by pharmaceuticals distributor Anda, obtained by the Washington Post (Abelson et al. (2018)), implicated numerous

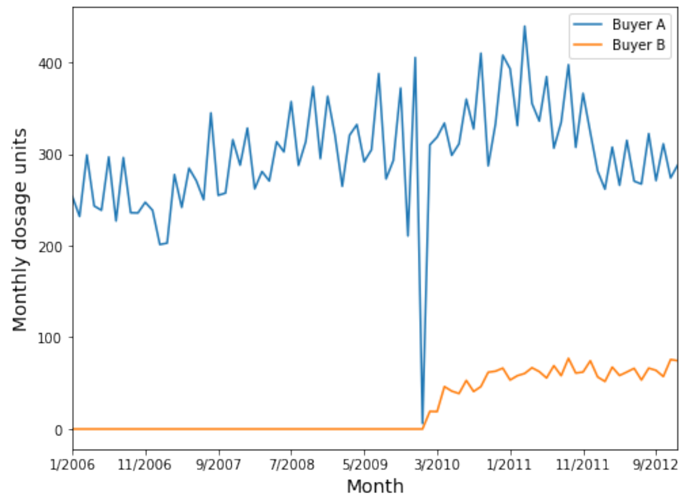

Figure 1 **Comparison of monthly dosage unit orders profiles for Walgreens buyers in Chico, CA**

Walgreens stores in Arizona, Oregon, and Maine. The list of flagged stores, and the sources used to identify these buyers, is detailed in the Appendix.[2]

Additional Data Processing

Exploratory analysis reveals clear interactions between geographically proximate Walgreens buyers. An instructive example of this behavior is shown in Figure 1, which compares the buyer time-series for two Walgreens buyer accounts from Chico, CA. Buyer A is flagged in the labels, and its registered location is approximately three minutes drive away from Buyer B. Immediately after a temporary disruption in supply to Buyer A, in January 2010, we observe that Buyer B begins ordering opioids. This likely represents an increase in over-procurement to this locale, but would not be detected if the buyers were considered in isolation. We attempt to account for such relationships by aggregating the data at the zip code level. This results in 206 zip codes with flagged buyers out of a total of 5441 zip codes in the data, or approximately 3.8%. Note that purchase amounts of Walgreens mail buyers are orders of magnitude larger than other Walgreens buyers.

[2] A limitation of our forensic approach for flagging stores is that some states, such as Arizona and Oregon, are over-represented in the sample of labels. We account for this limitation by comparing models that consider geographic information with those that do not rely on any geographically correlated data. The inter-agreements between the models suggest the robustness of our findings (see Section 4 for more details).

Due to the disparity in scale and the lack of label information for mail buyers, they are excluded from the analysis.

Throughout our analysis, we partition the data into training and test sets in an 80% to 20% ratio using stratified sampling based on the labels.

3. Methodology

When considering the appropriateness of machine learning models for early-detection, a central consideration is that the availability of labeled data can vary, and this, in turn, informs which methods are viable. Accordingly, we propose and evaluate a suite of detection methods that are applicable to three possible scenarios: (i) unsupervised, for when no labels are available; (ii) semi-supervised, for when only partial label information is available; and (iii) supervised, for when there is high confidence in many labels. The unsupervised and supervised settings represent the classical machine learning paradigms; thus, they can be viewed as sensible baselines. We remark that, both in this case and likely future applications of this work, the semi-supervised setting most closely resembles the true nature of the data. In this context, we propose a novel application of matrix completion for semi-supervised detection.

Clustering for Unsupervised Detection

In the unsupervised setting, we aim to cluster zip codes based on the similarity of their aggregated buyer timeseries. Human intuition can then be applied to determine which clusters warrant further investigation. To this end, a reasonable hypothesis is that smaller clusters represent anomalous behavior and should, therefore, be considered the most suspicious. Furthermore, if labels do become available, the probability of violation for a cluster can be estimated based on the proportion of flagged zip codes in it.

To quantify the similarity between timeseries, we utilize the Dynamic Time Warping (DTW) algorithm (Berndt and Clifford (1994)), an approach for matching temporal sequences that may vary in speed and phase. DTW finds the optimal match between two timeseries by non-linearly stretching or shrinking one sequence such that it most closely aligns with the other one. The matching can then be used to find the distance/similarity between the timeseries.

Formally, given aggregated buyer timeseries $\{x_i(t) : t \in [T]\}$ and $\{x_j(t) : t \in [T]\}$ for zipcodes i and j, respectively, the DTW distance is defined recursively as:

$$D_{ij}(t_1, t_2) = |x_i(t_1) - x_j(t_2)| + \min \left\{ \begin{array}{c} D_{ij}(t_1 + 1, t_2) \\ D_{ij}(t_1, t_2 + 1) \\ D_{ij}(t_1 + 1, t_2 + 1) \end{array} \right\} \tag{1}$$

with initial condition $D_{ij}(1, 1) = |x_i(1) - x_j(1)|$ and the total DTW distance given by $D_{ij}(T, T)$. In this manner, the sequences are warped non-linearly over time to most closely match each other. For efficient computation of this metric between millions of timeseries pairings, we employ the fast DTW algorithm proposed by Salvador and Chan (2007). We then apply agglomerative hierarchical clustering with the complete linkage metric. This is a greedy heuristic that builds clusters in a bottom-up fashion by iteratively linking samples such that the maximum DTW distance within each cluster is minimized. The number of clusters is then selected using the elbow heuristic, which aims to prevent over-fitting by reducing the number of clusters until there is significant deterioration in the objective.

Ensemble Model for Supervised Detection

We suggest two supervised methods that can be employed simultaneously and combined into an ensemble model. The first is k-nearest neighbor classification of buyer timeseries using the DTW distance metric. This method classifies a buyer timeseries of past opioid purchases by first identifying k buyers in the training set that are closest to it, where closeness is measured using the DTW distance. It then predicts whether the buyer is engaging in suspicious ordering behavior by using a voting scheme that aggregates the labels of the k closest observations. We determine the number of voting neighbors, k, through cross-validation.

The second supervised model we employ is the Optimal Classification Tree algorithm (OCT) (Bertsimas and Dunn (2017, 2019)) with the reporter-buyer adjacency matrix as input data. OCT is a hierarchically organized structure of nodes that make predictions by sequentially partitioning the data based on values of independent variables until it reaches a leaf node (a node holding a class label). OCT outperforms classical decision trees (Breiman et al. (1984)) as it leverages mixed-integer optimization to find the "optimal" tree with the highest accuracy. This algorithm formulates the problem of finding the globally optimal decision tree according to the loss function:

$$\min_{\mathbb{T}} \quad error(\mathbb{T}, T) + \alpha \times complexity(\mathbb{T}), \tag{2}$$

where $error(\mathbb{T}, T)$ measures the fit of the tree $\mathbb{T}$ on the training set T, $complexity(\mathbb{T})$ is the penalty function for the complexity of the tree, and α is the parameter that controls the bias-variance trade-off of the model. Comprehensive experiments show that OCT has accuracy comparable with the best classification methods while maintaining the interpretability of a single decision tree (Bertsimas and Dunn (2019)).

The ensemble prediction is given by the mean of the predicted probabilities of the two methods, which we refer to as the *combined supervised method*. Note that when training these models, we treat the labels as ground-truth, which is a necessary assumption given the data available.

Semi-Supervised Detection via Matrix Completion

The problem of detecting suspicious opioid purchasing behaviors can be formulated as a low-rank matrix completion problem. Such an approach attempts to recover a matrix from a partial sample of its entries (Cai et al. (2010), Jain et al. (2012)). Matrix completion has been successfully applied to a wide range of applications including, collaborative filtering (Candes and Plan (2010)), computer vision (Candes and Plan (2010)), recommender systems (Koren et al. (2009)), and product and content personalization (Farias and Li (2019)).

In our context, we seek to estimate the matrix whose rows correspond to the zip codes, columns correspond to timesteps (months), and entries contain the violation probabilities. Specifically, given N zip codes and T timesteps, we wish to create a binary matrix of labels $\mathbf{Y} \in \{0,1\}^{N \times T}$ such that $Y_{it} = 1$ if buyers in zip code i are flagged for suspicious behavior at timestep t, and zero, otherwise. As most Walgreens locations do not have a ground-truth label, most entries of matrix $\mathbf{Y}$ are missing. To create the initial incomplete labels matrix, we use the following procedure:

- If zip code i orders 0 doses during timestep t, set $Y_{it} = 0$;
- Else, if a buyer in zip code i is flagged in the labels and orders more than the median monthly amount at time t, set $Y_{it'} = 1, \ \forall t' \geq t$;
- Else, set Y_{it} as missing.

We wish to find a completed approximation of matrix $\mathbf{Y}$, which we denote as $\hat{\mathbf{Y}}$, whose entries can be interpreted as the predicted probabilities of violation. To achieve this, we utilize side information matrix $\mathbf{X}^{N \times P}$, which is a concatenation of the buyer timeseries and buyer-seller adjacency matrix. Formally, we consider the optimization problem:

$$\begin{aligned} \min_{\mathbf{U} \in \mathbb{R}^{T \times P}} \min_{\mathbf{S} \in \mathcal{S}_k^P} \ & \tfrac{1}{|\Omega|} \textstyle\sum_{(i,t) \in \Omega} \left(\hat{Y}_{it} - Y_{it} \right)^2 + \tfrac{1}{\gamma} \|\mathbf{U}\|_F^2 \\ \text{s.t. } & \hat{\mathbf{Y}}^\top = \mathbf{U}\mathbf{S}\mathbf{X}^\top \end{aligned} \tag{3}$$

where we have defined the set:

$$\mathcal{S}_k^P \triangleq \left\{ \operatorname{diag}(\mathbf{s}) : \mathbf{s} \in \{0,1\}^P, \ \sum_{j=1}^{P} s_i = k \right\}. \tag{4}$$

The first term in the objective function aims to minimize the mean squared error between known entries and their counterparts in the completed matrix. The second term in the objective is an L_2 regularization term, which is added for robustness. The strength of the regularization is controlled by the parameter γ. The constraint imposes that the completed matrix $\hat{\mathbf{Y}}$ is a linear response to side information $\mathbf{X}$, with coefficients $\mathbf{U}$. Moreover, given $\mathbf{S} \in \mathcal{S}_k^P$, the constraint also ensures that $\operatorname{rank}(\hat{\mathbf{Y}}) = k$. To solve (3), we utilize the OptComplete algorithm (Bertsimas and Li (2018)), as it is highly scalable and allows for the incorporation of side information into the predictions. The rank k and regularization parameter γ are tuned via a grid search. As the elements of $\hat{\mathbf{Y}}$ represent probabilities, we should also impose that $\hat{Y}_{it} \in [0,1]$. In practice, this is enforced as post-processing step by clipping the entries of the completed matrix. This approach has the additional benefit of inferring the time periods over which potential abuses took place. Nonetheless, the predictions for each month can be combined to give an estimate for the overall probability of violation for zip code i as follows:

$$\hat{p}_i = 1 - \prod_{t=1}^{T} \left(1 - \hat{Y}_{it} \right). \tag{5}$$

Year	Clustering	Matrix Completion	k-NN	OCT	Combined Supervised
2006	0.712	0.963	0.747	0.840	0.909
2007	0.694	0.937	0.750	0.959	0.994
2008	0.706	0.949	0.785	0.911	0.959
2009	0.716	0.964	0.783	0.925	0.953
2010	0.704	0.983	0.901	0.912	0.942
2011	0.702	0.955	0.892	0.934	0.963
2012	0.709	0.979	0.931	0.958	0.973

Table 2 Out-of-sample AUC scores by method, trained on the data available up to the end of each year.

Note that a more conservative and robust estimate, which remains monotonically increasing with T, can be obtained by taking the maximum imputed values across all months: $\hat{p}_i = \max_{t=1,\ldots,T} \hat{Y}_{it}$.

4. Results

To assess the capacities of the proposed methods for early detection, we trained separate versions of the models using the input data available up to the end of every calendar year from 2006 to 2012. Intuitively, models that can achieve strong predictive performance on fewer years of data are better suited for an early-detection system. We evaluated the performance of the trained models by measuring the out-of-sample area under the curve (AUC).

4.1. Predictive Performance

Table 2 shows the out-of-sample performance of the models, evaluated using the labels. The matrix completion and combined supervised methods are the strongest, achieving AUC scores of over 0.97 in 2012. The performance of these models is relatively consistent when fewer years of data are used, suggesting that these models are capable of early violation detection, potentially several years in advance. Moreover, the receiver operating characteristic (ROC) curves for 2010 and 2012, depicted in Figure 2, also highlight the superiority of the matrix completion and combined supervised methods.

While the unsupervised results are not as strong, they do offer empirical support for our hypothesis that small clusters are likely to indicate suspicious activity. Figure 3 depicts the relationship

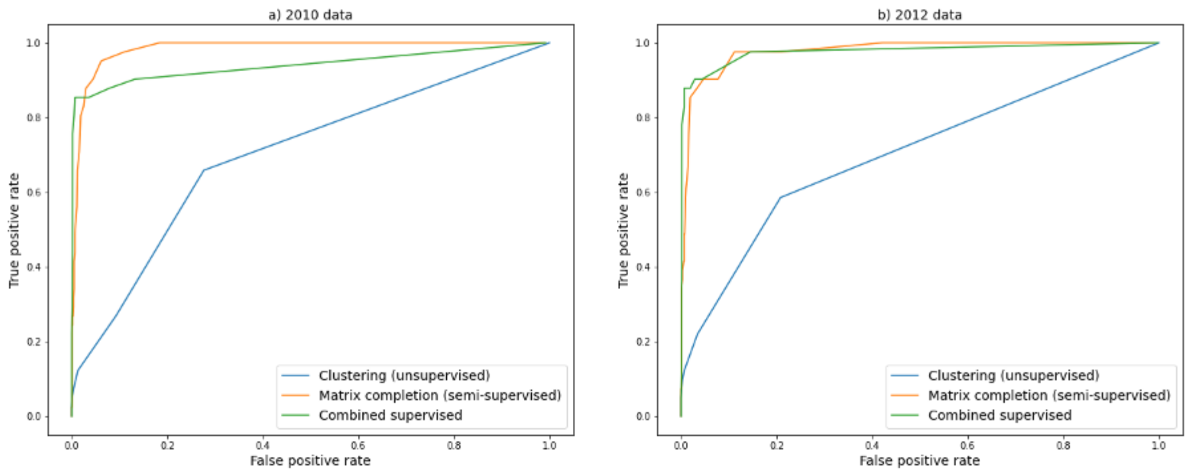

Figure 2 **Out-of-sample ROC curves for unsupervised, semi-supervised and supervised methods using data up to the end of 2010 and 2012, respectively.**

between cluster size and the percentage of cluster members flagged in the raw labels. As expected, there is a strong negative correlation, with 100% of the zip codes in the two smallest clusters being flagged. This finding is significant because it suggests a viable approach for anomaly detection in the absence of any supervisory labels.

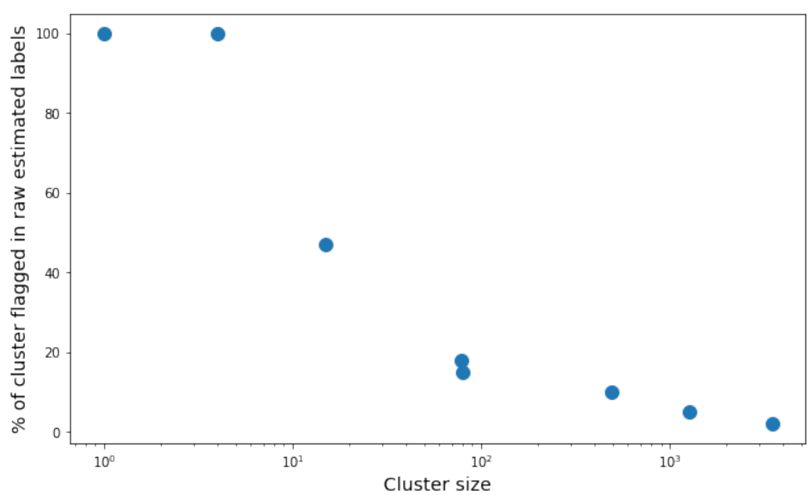

Figure 3 **Relationship between cluster size and percentage of flagged cluster members using data up to the end of 2012.**

Robustness

In order to provide further validation of the proposed models in detecting illegal activity, we examine the inter-model agreement on a case-by-case basis. The rationale behind this approach is that in the absence of authoritative labels, consistency between models is an indicator of predictive accuracy. It should be noted that the amount of agreement between models is dependent on the cut-off threshold used to flag zip codes based on their predicted violation probabilities. This effect is exacerbated by the imbalanced nature of the dataset. To obtain fair and reasonable comparison, we select the cut-off threshold for each case such that the minimum positive rate of the two methods being compared is equal to the proportion of flagged zip codes in the labels. This results in a positive rate of approximately 3.8% being used in the assessment. To indicate the sensitivity of the inter-model agreement to the cut-off threshold, we also perform the same analysis for a lower threshold that corresponds to a positive rate of 10%. The results, shown in Figure 4, demonstrate a high level of inter-model agreement, though this decreases as the positive rate increases. The agreement between the matrix completion and combined supervised methods is particularly strong: above 80% for all cases. Moreover, the degree of agreement is relatively independent of the number of years of data used. Altogether, the consistency between methods demonstrated by these robustness checks further supports their efficacy in the absence of authoritative labels.

Note that in October 2019, DEA launched the Suspicious Orders Report System Online, a database for collecting suspicious order reports from distributors of controlled substances. This data could provide more authoritative labels that may be incorporated into our machine learning framework to improve the quality and robustness of our predictions.

4.2. Impact of Early-Detection

A clear benefit of using an early detection system is a reduction in the number of illegally-obtained pills. To estimate this quantity for each of the models, we attempt to gauge the impact of an intervention at the conclusion of each calendar year. For each model and intervention time, we select the cut-off threshold that makes the positive rate equal to the proportion of flagged zip codes

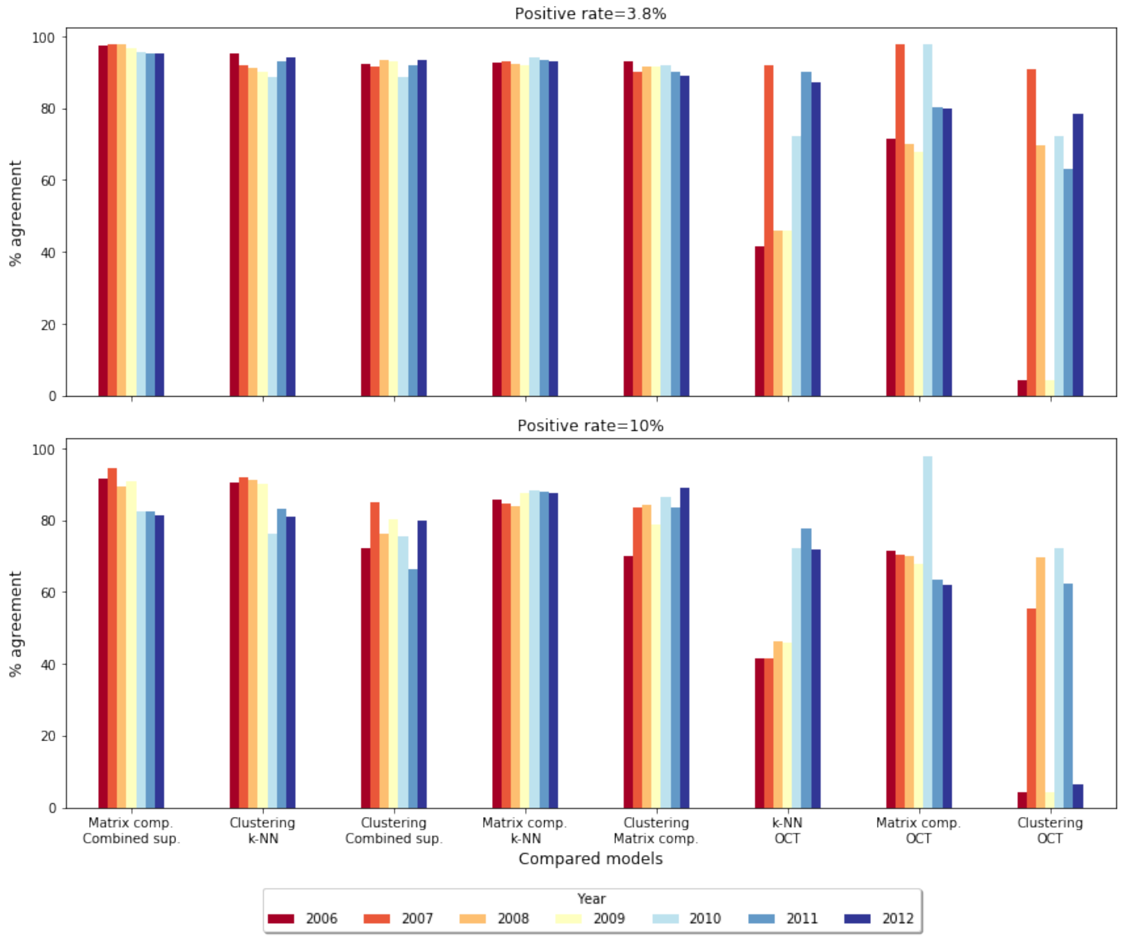

Figure 4 **Visualization of out-of-sample inter-model agreement, for models trained on the data available up to the end of each year.**

in the labels. This threshold is chosen to obtain a conservative but realistic number of flagged zip codes. We then calculate the number of pills, in excess of the monthly median across all zip codes, distributed to the zip codes flagged by the model from the intervention time to the end of 2012, when the data ceases. The outcomes of this procedure are summarized in Figure 5. We observe the semi-supervised and supervised models perform very similarly, with the semi-supervised method resulting in a slightly higher estimated reduction for earlier intervention times. Overall, we estimate that the illegal distribution of as many as 600 million pills could have been avoided if an early detection system was in place by the end of 2006. This amounts to 5% of the total pills distributed by Walgreens over the time period for which data is available.

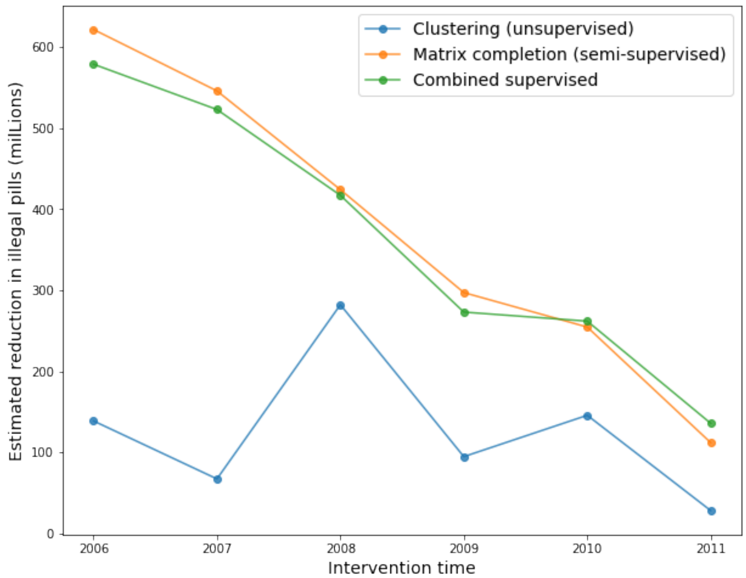

Figure 5 **Visualization of the estimated number of illegal pills prevented by early detection models, based on year of implementation.**

We also applied the combined supervised early detection model trained on Walgreens data to CVS stores, also aggregated by zip code. Using the cut-off threshold calibrated on the Walgreens data to achieve a positive rate matching the labels, we flagged 62 zip codes for CVS, corresponding to additional prevention of up to 160 million potential pills.

5. Conclusion

In this paper, we introduced machine learning models that leverage vast amounts of administrative data to reliably detect patterns of opioid over-procurement within the United States' prescription drug supply chain. Our results suggest that our models can provide substantial opportunities for early violation detection, up to several years in advance. We estimate that the models could have prevented the illegal distribution of hundreds of millions of pills.

The machine learning models presented in this paper can be readily utilized by the DEA to complement their current reactive policies (US Government Accountability Office (2020)), detecting violations only after they have occurred, with proactive data-driven alternatives. This would enhance DEA's efforts to prevent, detect, and investigate the diversion of illicit drugs in a more timely manner.

The opioid epidemic is a complex societal problem requiring policy interventions ranging from detection to treatment. Our work demonstrates the feasibility of machine learning models for proactively identifying illegal diversion of controlled substances. Such early detection models aid in preventing further cases of opioid use disorder, which can ultimately save lives.

References

Abelson J, Williams A, Ba Tran A, Kornfeild M (2018) At height of crisis, Walgreens handled nearly one in five of the most addictive opioids *The Washington Post*, 7 November 2019. https://www.washingtonpost.com/investigations/2019/11/07/height-crisis-walgreens-handled-nearly-one-five-most-addictive-opioids/?arc404=true.

Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, 359–370, AAAIWS'94 (AAAI Press).

Bertsimas D, Dunn J (2017) Optimal classification trees. *Machine Learning* 106(7):1039–1082.

Bertsimas D, Dunn J (2019) *Machine Learning Under Modern Optimization Lens* (Dynamic Ideas, Belmont, MA, 2019).

Bertsimas D, Li ML (2018) Interpretable matrix completion: A discrete optimization approach. *arXiv preprint arXiv:1812.06647* .

Breiman L, Friedman J, Olshen R, Stone C (1984) *Classification and Regression Trees* (Wadsworth, Monterey, CA, 1984).

Brownstein J, Green T, Cassidy T, Butler S (2010) Geographic informations systems and pharmacoepidemiology: Using spatial cluster detection to monitor local patterns of prescription opioid abuse. *Pharmacoepidemiology and drug safety* 19:627–37.

Cai JF, Candès E, Shen Z (2010) A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20:1956–1982.

Candes EJ, Plan Y (2010) Matrix completion with noise. *Proceedings of the IEEE* 98(6):925–936.

Dufour R, Mardekian J, Pasquale M, Schaaf D, Andrews G, Patel N (2014) Understanding predictors of opioid abuse: predictive model development and validation. *Am J Pharm Benefits* 6(5):208–16.

Ellis RJ, Wang Z, Genes N, Ma'ayan A (2019) Predicting opioid dependence from electronic health records with machine learning. *BioData mining* 12(1):3.

Farias VF, Li AA (2019) Learning preferences with side information. *Management Science* 65(7):3131–3149.

Han DH, Lee S, Seo DC (2020) Using machine learning to predict opioid misuse among us adolescents. *Preventive medicine* 130:105886.

Hastings JS, Howison M, Inman SE (2020) Predicting high-risk opioid prescriptions before they are given. *Proceedings of the National Academy of Sciences* 117(4):1917–1923.

Hylan TR, Von Korff M, Saunders K, Masters E, Palmer RE, Carrell D, Cronkite D, Mardekian J, Gross D (2015) Automated prediction of risk for problem opioid use in a primary care setting. *The Journal of Pain* 16(4):380–387.

Jain P, Netrapalli P, Sanghavi S (2012) Low-rank matrix completion using alternating minimization. *Proceedings of the Annual ACM Symposium on Theory of Computing* .

Kenan K, Mack K, Paulozzi L (2012) Trends in prescriptions for oxycodone and other commonly used opioids in the united states, 2000-2010. *Open medicine* 6:e41–7.

Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. *Computer* 42(8):30–37.

Lo-Ciganic WH, Huang JL, Zhang HH, Weiss JC, Wu Y, Kwoh CK, Donohue JM, Cochran G, Gordon AJ, Malone DC, et al. (2019) Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. *JAMA network open* 2(3):e190968–e190968.

Mackey TK, Kalyanam J, Katsuki T, Lanckriet G (2017) Twitter-based detection of illegal online sale of prescription opioid. *American journal of public health* 107(12):1910–1915.

Prieto JT, Scott K, McEwen D, Podewils LJ, Al-Tayyib A, Robinson J, Edwards D, Foldy S, Shlay JC, Davidson AJ (2020) The detection of opioid misuse and heroin use from paramedic response documentation: Machine learning for improved surveillance. *Journal of Medical Internet Research* 22(1):e15645.

Salvador S, Chan P (2007) Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis* 11(5):561–580.

Scholl L, Seth P, Kariisa M, Wilson N, Baldwin G (2018) Drug and opioid-involved overdose deaths in the united states from 2013 to 2017. *Morbidity and mortality weekly report* 67:1419–1427.

The Washington Post (2020) DEA database: where the pain pills went. *The Washington Post*, 17 January 2020. https://www.washingtonpost.com/graphics/2019/investigations/dea-pain-pill-database/.

United States Department of Justice (2013) Walgreens agrees to pay a record settlement of $80 million for civil penalties under the controlled substances act, 11 June 2013. https://www.justice.gov/usao-sdfl/pr/walgreens-agrees-pay-record-settlement-80-million-civil-penalties-under-controlled.

United States District Court (2020) National prescription opiate litigation. https://www.ohnd.uscourts.gov/mdl-2804.

US Government Accountability Office (2020) Actions needed to ensure usefulness of data on suspicious opioid orders.

Appendix
Additional statistics

State	Registered buyers	Mean dosage units per year	Mean dosage units per year per capita
Total	147,175	1.11E+10	36.6
WV	975	1.23E+08	67.1
KY	1,834	2.74E+08	64.1
SC	2,194	2.64E+08	58.8
TN	3,883	3.64E+08	58.6
OK	2,073	2.03E+08	55.3
NV	1,208	1.45E+08	55.2
AL	3,064	2.47E+08	52.5
OR	1,719	1.93E+08	51.4
IN	3,095	3.07E+08	47.9
DE	424	3.99E+07	45.4
AR	1,584	1.30E+08	45.4
KS	1,182	1.27E+08	45.1
LA	2,954	2.00E+08	45.1
FL	11,837	8.07E+08	43.7
ME	583	5.76E+07	43.4
OH	4,629	4.91E+08	42.7
MS	1,636	1.24E+08	42.2
MI	4,557	4.13E+08	41.5
WA	2,853	2.69E+08	41.1
NC	4,266	3.70E+08	40.0
AZ	2,820	2.49E+08	39.8
MO	2,365	2.32E+08	39.1
UT	1,004	1.01E+08	37.9
NM	658	7.54E+07	37.6
MT	573	3.57E+07	36.8
ID	737	5.59E+07	36.7
PA	5,950	4.49E+08	35.6
RI	443	3.69E+07	35.0
GA	5,588	3.31E+08	34.9
WY	285	1.84E+07	34.0
WI	2,284	1.86E+08	33.1
TX	12,040	7.89E+08	32.6
CA	18,329	1.17E+09	32.0
NH	466	4.08E+07	31.0
CO	2,222	1.49E+08	30.5
MD	2,826	1.70E+08	29.8
VA	3,218	2.32E+08	29.7
AK	292	2.02E+07	29.4
VT	285	1.80E+07	28.8
MA	2,030	1.86E+08	28.7
CT	1,524	9.83E+07	27.8
IA	1,272	8.22E+07	27.3
NJ	5,055	2.22E+08	25.5
NE	931	4.58E+07	25.5
NY	8,546	4.88E+08	25.3
HI	522	3.19E+07	24.0
MN	2,066	1.23E+08	23.5
IL	5,385	2.84E+08	22.3
SD	336	1.69E+07	21.2
ND	293	1.32E+07	20.1
DC	280	8.63E+06	14.7

Table 3 **Aggregate ARCOS transaction statistics by state.**

Label information

Buyer DEA No.	Source
AW1366877, AW8830247, BW4713992, BW5872494, BW6561270 BW8487438	US Department of Justice (United States Department of Justice (2013))
BW3614888, BW9373565, AW5430943, BW1229118, BW2101880 BW7288752, BW6407262, BW4963977, BW4986622, BW5315204 BW8421707, BW5483970, BW5737791, BW7249623, FW2277172 BW5837591	US District Court (United States District Court (2020))
BW4243426, BW8002759, AW9474723, BW1441865, BW6997906 BW3419567, BW7512595	The Washington Post (The Washington Post (2020))
AW2981315, BW1264415, BW2192223, BW7109069, BW7305370 BW9131727, BW9975422, AW0572481, AW0572493, AW0572619 AW1195076, AW1994400, AW3247699, AW4328375, AW5115755 AW5115779, AW5506968, AW6909103, BW0214849, BW0264971 BW0403078, BW0804042, BW0952956, BW0987327, BW1773008 BW2188159, BW2778946, BW3173197, BW3589314, BW3617620 BW3717254, BW3749225, BW3819781, BW4041098, BW4115677 BW4293192, BW4642232, BW5204627, BW5452228, BW6053502 BW6185474, BW6618461, BW6708359, BW6831944, BW7042411 BW7109071, BW7189194, BW7189207, BW7340778, BW7407516 BW7612244, BW7652274, BW7686011, BW7891460, BW7949069 BW8591047, BW8678748, BW9013107, AW3058662, AW3220869 AW5722396, AW5732119, BW0621905, BW1231846, BW1293769 BW4041086, BW4218055, BW4642256, BW6510956, BW6708347 BW7130533, BW7191644, BW7191668, BW7505019, BW7608637 BW7822631, BW8038437, BW8365733, BW8616685, BW9569483 FW0275506, BW7503382, BW9164574, BW9164586, BW9164598 BW3793191, AW0572506, AW0572520, AW0572532, AW0572568 AW0572582, AW0572594, AW1055284, AW1195052, AW2890627 AW2905466, AW2943618, AW3058674, AW3164136, AW5115767 AW5116404, AW5120237, AW5416361, AW5745255, AW6336867 AW6742616, AW7354753, AW8348852, AW8879883, AW9355517 AW9487768, AW9623770, AW9699818, BW0264957, BW0482531 BW0610837, BW0664121, BW0691938, BW0744929, BW0744931 BW0804030, BW0952944, BW1059991, BW1104859, BW1381134 BW1464596, BW1997317, BW2094441, BW2245012, BW2988751 BW3173212, BW3176220, BW3276068, BW3652193, BW3717228 BW3793177, BW3875830, BW4013823, BW4115653, BW4197871	Anda Pharmaceuticals (Abelson et al. (2018))

BW4382975, BW4392964, BW4456504, BW4456516, BW4487674
BW4612215, BW4612227, BW4721393, BW4764709, BW4846068
BW4908907, BW4940866, BW4961377, BW5335117, BW5335131
BW5335143, BW5524891, BW5638082, BW5648425, BW5939648
BW6197897, BW6197924, BW6213881, BW6283561, BW6368737
BW6368749, BW6448105, BW6448129, BW6593948, BW6658857
BW6993631, BW7109083, BW7141334, BW7143996, BW7165358
BW7189170, BW7189219, BW7207512, BW7236032, BW7236044
BW7308124, BW7326499, BW7412238, BW7452028, BW7505007
BW7612256, BW7621952, BW7686023, BW7905144, BW8002735
BW8106519, BW8157819, BW8223050, BW8303389, BW8382056
BW8438651, BW8457106, BW8490283, BW8505351, BW8567604
BW8888488, BW9388895, BW9485194, BW9519995, BW9537094
BW9642984, BW9684297, BW9696711, BW9740742, FW0041234
FW0178601, FW0434770, FW0625220, FW1015432, FW1053747
FW1224839, FW1398406, AW0572518, AW5120249, BW0492380
BW1377200, BW4566393, FW0549254, FW1398393, FW1851408
BW5775664, BW8650271, BW8926961, BW5018379, BW5139616
BW5139628, BW5234454, BW5234543, BW5667437, BW5730925
BW5932771, BW5993349, BW6002567, BW6302183, BW6792255
BW6979972, BW7013155, BW7147766, BW7193775, BW7198698
BW7326514, BW7472032, BW7860693, BW8050837, BW8178217
BW8226195, BW8330805, BW8346555, BW8410716, BW8628806
BW8628806, BW8756047, BW8756061, BW8892019, BW9485207
BW9565170, BW9710131, BW9710167, FW0020115, FW0395978
FW0467325, FW0600103, FW0757659, FW0757661, FW0974320
FW1236769, FW1690456, BW5756626, BW7440819, FW0757647
FW1094820,FW1144562, FW2431372, FW2818334

Table 4: Flagged Walgreens buyers, and the data sources used to extract the labels. As a preliminary sanity check, we verified that these buyers had at least one month of orders that exceeds the median monthly dosage units, across all Walgreens buyers.