Advances in Social Science, Education and Humanities Research, volume 638
Proceedings of the 2021 International Conference on Public Art and Human Development (ICPAHD 2021)

Predict Opioid-using Pharmacists from the Prescribing Practices Based on SVM Classification Model

Ximing Ran[1,*]

[1]*Shanghai University of Finance and Economics, School of Statistics and Management, Shanghai 200433, China*
**Corresponding author. Email: rxm@163.sufe.edu.cn*

ABSTRACT
The number of people dying from drug overdoses in the United States is increasing every year, and most of them are caused by opioids, so it becomes crucial to analyze the use of opioids. Since an important influence on opioid use is the physicians who use opioids, we studied physicians' medication habits to obtain an analysis of physicians' opioid use habits and predictions of physicians' propensity to use opioids. We first analyzed physicians' opioid use through the physician medication use dataset provided by CMS, and analyzed several opioids that are used more frequently to analyze the correlation of physicians' behavior toward different types of opioid use. After that, through doctors' medication habits, for whether doctors will use opioid drugs to make predictions, by constructing support vector machine models, comparing the classification effects of different kernel functions, and finally constructing a classification model with 85% prediction accuracy.

Keywords: *Opioids, Pharmacists, Prescription, Support vector machines*

1. INTRODUCTION

Prescription opioids can be used to treat moderate-to-severe pain and are often prescribed following surgery or injury, or for health conditions such as cancer. In recent years, there has been a dramatic increase in the acceptance and use of prescription opioids for the treatment of chronic, non-cancer pain, such as back pain or osteoarthritis, despite serious risks and the lack of evidence about their long-term effectiveness. In addition to being used as a powerful painkiller, opioids also produce pleasure, which makes them highly addictive and causes abuse. The same mechanism that inhibits pain also inhibits breathing and can lead to death by asphyxiation if used in excess. Therefore, long-term use can cause opioid addiction and endanger the health of the user. Opioid abuse is one of the greatest challenges facing public health in the 21st century. CDC statistics show that overdoses as a cause of death have killed more people in the United States than car accidents, guns and the flu virus. The majority of these casualty are opioid-induced deaths, with opioids causing 49,860 overdose deaths in 2019 (70.6% of all drug overdose deaths)[1]. The number of deaths caused by opioids in the United States is increasing year by year, and there is a nationwide proliferation of opioids uses, so the Food and Drug Administration, the Centers for Disease Control and Prevention, and the U.S. Drug Enforcement Administration have all placed restrictions on the use of opioids. Therefore, this paper will analyze the current use of opioids in the United States, analyze the medication habits of physicians, predict the frequency of opioid use by physicians, and prevent opioid abuse.

2. METHOD

2.1 Data Sources

To investigate the overall U.S. national picture, we obtained the U.S. CMS public dataset Medicare Part D Prescription Drug Dataset ("Part D Prescription Drug Dataset"), which contains information on prescription drug events (PDEs) occurring in the Part D prescription drug program for Medicare beneficiaries. The primary data source for these data is the CMS Chronic Conditions Data Warehouse, which contains Medicare Part D PDE records received through the claim's submission cut-off date. The submission cut-off date is June 30th following the end of the preceding calendar year. For instance, the 2019 Part D Prescribers Dataset includes PDEs received through June 30, 2020[2]. Drug brand names and generic names used in the summaries were obtained by linking the National Drug Codes (NDCs) in the PDE records to commercially available drug information databases. A

small percentage of PDE records with NDCs that did not match drug information databases were excluded from all summaries.

2.2 Sample Data Set

To examine the most recent data, we obtained the CMS public dataset Medicare Part D Prescription Drug Dataset ("Part D Prescription Drug Dataset"), which contains Medicare beneficiaries in the Part D prescription drug program who had a prescription drug events (PDEs) that occur in the Part D prescription drug program for Medicare beneficiaries. The primary data source for these data is the CMS Chronic Disease Data Warehouse, which contains records of Medicare Part D PDEs received through the claims submission deadline. The submission deadline is June 30 following the end of the previous calendar year. For example, the 2017 Part D prescriber data set includes PDEs received through June 30, 2018. This portion of the data totals 25,401,870 records, each of them represents the total use of a kind of drug by a pharmacist during the year. To facilitate the calculation, a random sample was taken from which 1% was processed, 254,018 medication use data, after which the data were organized by pharmacist to count the medication use habits of each pharmacist. Each piece of data represents the total use of various drugs used by this pharmacist during the year.

For the definition of opioids, the CMS criteria were used, and pharmacists who had used opioids in the past year were marked as 1 and those who had not were marked as 0. Opioid use was later analyzed separately.

At the same time, the Part D prescription data set has some limitations. First, the data contain information only for patients with Part D coverage, but physicians typically treat many patients with practical other coverage, so the data in this dataset may not be representative of all prescriptions used by a single physician. Because not all Part D plans have supplemental coverage for products that are excluded, the utilization and cost statistics shown in the data may underestimate the true use of these products in this population.

2.3 Data Set Description

To examine the most recent data, i.e., all Medicare enrollment between July 1, 2019 and June 30, 2020 (this data represents only Medicare-enrolled prescription records within the United States; actual opioid prescription drug use should be greater than this data). This portion of the data totals 25,401,870 records, each of which represents the total use of a particular drug by a pharmacist during the year. To facilitate the calculation, a random sample was taken from which 1% was processed, 254,018 medication use data, after which the data were organized by pharmacist to count the medication use habits of each pharmacist. Each piece of data represents the total use of various drugs used by this pharmacist during the year.

For the definition of opioids, the CMS criteria were used, and pharmacists who had used opioids in the past year were marked as 1 and those who had not were marked as 0. Opioid use was later analyzed separately.

To solve the sample imbalance problem, we first resample the sample by taking a sample of 1000 opioid-using physicians and 1000 non-opioid-using physicians for training the model.

2.4 Analysis of Prescription Opioid Use

The following table provides an analysis of opioid use by all opioid-using physicians. Hydrocodone/Acetaminophen is a combination of Hydrocodone and Acetaminophen, which are moderate to strong pain relievers; Acetaminophen is a non-narcotic analgesic. The purpose of combining these two different ingredients into a pain reliever is to maximize the efficacy of the drug and to minimize side effects and avoid the abuse of narcotic drugs through their different pharmacological mechanisms (principles). This drug was used by 23.63% of all physicians who use opioids, which shows that individual physicians are contributing to reducing patients' dependence on opioids.

Table 1: Opioid use by U.S. Physicians

drug name	mean	std	max	sum total	ratio of doctor
Hydrocodone/Acetaminophen	19.61	100.41	3728	209719	23.63%
Tramadol Hcl	11.87	38.71	958	126969	20.81%
Oxycodone Hcl/Acetaminophen	7.17	51.84	2002	76642	11.75%
Oxycodone Hcl	6.96	41.50	1507	74472	12.60%
Topiramate	2.69	16.10	603	28803	7.41%
Acetaminophen With Codeine	2.32	13.32	603	24774	7.09%
Morphine Sulfate	2.24	17.74	808	23942	5.91%
Fentanyl	1.46	14.09	861	15637	3.87%
Methadone Hcl	0.81	15.07	992	8662	1.87%
Hydromorphone Hcl	0.64	6.93	339	6846	2.13%

Tramadol Hcl/Acetaminophen	0.26	3.86	214	2746	0.94%
Buprenorphine Hcl	0.23	4.08	229	2469	0.68%
Buprenorphine	0.20	3.30	197	2183	0.78%
Oxycodone Myristate	0.13	2.82	175	1440	0.47%
Oxymorphone Hcl	0.10	2.55	192	1023	0.36%
Tapentadol Hcl	0.09	1.72	61	1013	0.41%
Morphine Sulfate/Naltrexone	0.06	2.55	211	631	0.15%
Hydrocodone Bitartrate	0.05	1.24	77	569	0.25%
Hydrocodone/Ibuprofen	0.05	1.31	104	552	0.28%
Butalbit/Acetamin/Caff/Codeine	0.02	0.60	25	237	0.15%
Codeine/Butalbital/Asa/Caffein	0.02	0.59	26	227	0.14%
Butorphanol Tartrate	0.02	0.67	43	202	0.10%
Pentazocine Hcl/Naloxone Hcl	0.01	0.42	21	121	0.07%
Levorphanol Tartrate	0.01	0.47	42	85	0.04%
Codeine Sulfate	0.01	0.30	22	59	0.04%
Acetaminophen/Caff/Dihydrocod	0.00	0.32	29	52	0.03%
Meperidine Hcl	0.00	0.22	18	32	0.02%
Fentanyl Citrate	0.00	0.18	14	26	0.02%
Oxycodone Hcl/Aspirin	0.00	0.14	14	14	0.01%

The top opioids used were: Tramadol Hcl, Oxycodone Hcl/Acetaminophen, Oxycodone Hcl, Topiramate, Acetaminophen With Codeine. All of these drugs are used by more than 5% of physicians, so the regulation of these drugs should be monitored with emphasis.

A total of 29 opioid drugs were used in the sample data, and by finding the correlation coefficient for the top ten drugs in use, the absolute value of the effect coefficient for any two drugs was less than 0.2, and it can be assumed that the correlation between individual opioid drugs is small, and each drug can be analyzed separately in the follow-up study

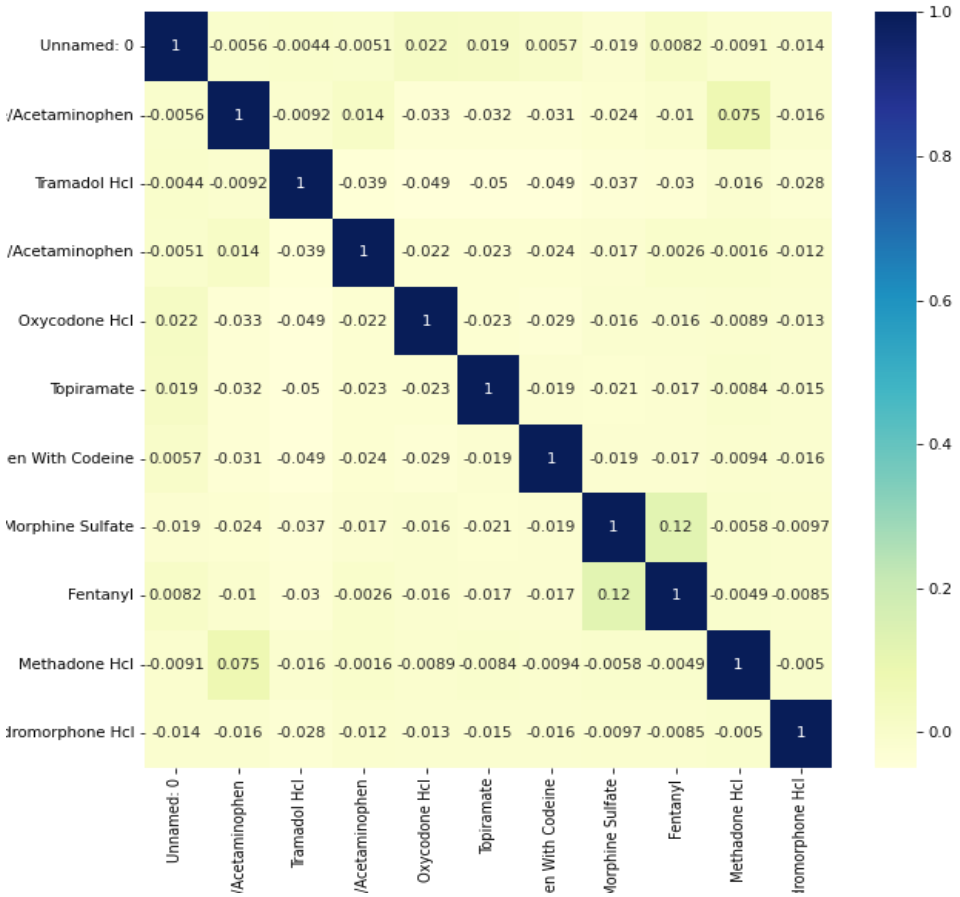

Figure 1 : Correlation analysis of opioid use among all opioid-using physicians

2.5 SVM classification model-based prediction of opioid use propensity by US physicians

We then analyzed the model using a SVM (support vector machine) classification model, a linear classifier that classifies data by constructing hyperplanes in vector space. The goal of our study: to predict physicians' propensity to use opioids is a binary classification problem that can be better solved by the SVM classification model[4].

We first screened for opioid use among the sample physicians. Of the 177,341 physicians in the sample, only 10,693 used opioids in the past year, accounting for 6.02%. And for the analysis of opioid use.

2.5.1 Model evaluation method: cross-validation method

By performing gradient descent optimization of the model parameters, the final SVM classification model had 80% accuracy on 10-fold cross-validation and better correctness on the test set, indicating that the model has good generalization ability to have better accuracy on new data. It is possible to classify a large number of usage cases with a smaller amount of data. On the re-sampled 1000 data, the correct rate is 78%. There is a high rate of correctness.

2.5.2 Comparison of SVM classification model optimization based on different kernel functions

In solving linear indistinguishable problems, the SVM classification model achieves the best classification by using kernel functions to dimensionalize the data so that the data can be partitioned by hyperplanes in the vector space[3]. Therefore, the choice of kernel function is intuitively important for the effectiveness of classification[6], so we first test several classes of kernel functions to test their accuracy.

Table 2. Comparison of different kernel functions for support vector machine models

Kernel functions	precision	recall	f1-score
Linear functions	0.635	0.64	0.63
Polynomial functions	0.615	0.62	0.59
Radial basis functions	0.85	0.85	0.85
Sigmoid functions	0.725	0.73	0.72

Since we are studying an unbalanced sample, i.e., a situation where doctors who tend to use opioids are in the minority, we judged that we need to consider both accuracy as well as recall, weighting the two and evaluating them using F1 scores, and comparing several kernel functions, and finally we found that the best results were obtained using the Radial basis functions[6].

3. CONCLUSION

The current prescription opioid abuse situation in the United States is serious and has a large impact on the survival of the general population. The classification model constructed by the support vector machine model dynamically predicts whether a physician has a tendency to favor opioid use based on his or her prescribing record. It reduces the abuse of opioids from the physician's perspective. It allows physicians to be more cautious in their use of medications. The results demonstrate that only about 2% of physicians tend to use opioids, so opioid addiction promotion is needed for this group of physicians. The most frequently used opioid among all opioids is Hydrocodone/Acetaminophen, an opioid that significantly reduces addiction, but we also need to pay attention to the number of physicians prescribing this drug to avoid addiction to it. In addition we found a weak correlation between the use of individual opioids, i.e. there was less mixed use of multiple opioids.

REFERENCES

[1] Mattson CL, Tanz LJ, Quinn K, Kariisa M, Patel P, Davis NL. Trends and Geographic Patterns in Drug and Synthetic Opioid Overdose Deaths — United States, 2013–2019. MMWR Morb Mortal Wkly Rep 2021;70:202–207. DOI: http://dx.doi.org/10.15585/mmwr.mm7006a4

[2] https://data.cms.gov/provider-summary-by-type-of-service/medicare-part-d-prescribers/medicare-part-d-prescribers-by-provider-and-drugzz

[3] Brewer, D. D., Catalano, R. F., Haggerty, K., Gainey, R. R., & Fleming, C. B. (1998). RESEARCH REPORT A meta-analysis of predictors of continued drug use during and after treatment for opiate addiction. Addiction, 93(1), 73-92.

[4] Huang, C., Zhang, R., Chen, Z., Jiang, Y., Shang, Z., Sun, P., ... & Li, X. (2010). Predict potential drug targets from the ion channel proteins based on SVM. Journal of theoretical biology, 262(4), 750-756.

[5] Björne, J., Kaewphan, S., & Salakoski, T. (2013, June). UTurku: drug named entity recognition and drug-drug interaction extraction using SVM classification and domain knowledge. In Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) (pp. 651-659).

[6] Joachims, T. (1998). Making large-scale SVM learning practical (No. 1998, 28). Technical report.