Predictive Analysis in Identifying Frequent Opioid Prescriber

Erjia Meng, Jiayi Hao, Shiyu Liu, Xinyi Liu, and Yulin Feng

I. INTRODUCTION

Opioid is a kind of drug that is derived from the opium poppy. It can be prescribed by doctors to relieve severe pain, but it is highly addictive and can have serious side effects while use. While being an effective prescribed medication for clinical uses, in recent years, an increasing number of people in the United States have been suffering from the side effects or the withdrawal effects caused by the misuse or overdose of opioid drugs. One source of the increasing misuse of opioids is that there are a certain amount of doctors who became the 'pill whale' who are over-prescribing opioids. As the drug is highly addictive, the doctors could thus benefit themselves more by prescribing more drugs.

Aimed at making some contributions to reduce the opioid overdose, we built a two-class classification model based on a dataset of 25,000 unique licensed medical professionals in the United States where their basic information and partial prescribing history are recorded. In this project, we first conducted detailed Exploratory Data Analysis on the data, then performed Dimension Reduction to visualize the data, and finally compared the performance of various Supervised Learning models in predicting 'frequent opioid prescribers' given the prescribers' information. In our project, we found our model, with Random Forest classifier, to be the most effective in predicting 'frequent opioid prescribers' with a precision over 95% and sensitivity over 90%. With our work, it is shown that Machine Learning can be a systematic and powerful method to locate the doctors who are frequently prescribing opioids so that further analysis can be done to identify the real 'pill whales'.

II. LITERATURE REVIEW

A. Opioids as Prescription Medications

Opioids are commonly prescribed for controlling pain. Many prescription opioids are used to block pain signals between the brain and the body and are typically prescribed to treat moderate to severe pain.[6] Common types are oxycodone (OxyContin), hydrocodone (Vicodin), morphine, and methadone.[2] However, although opioid medication can be effective in controlling pain, addiction could be a common consequence of prolonged use. As many as one in four patients receiving long-term opioid therapy in a primary care setting struggles with opioid addiction.[4] Specifically, as pain-relieving effect would lessen after prolonged use, there's a high risk for addicts to experience opioid overdose, which may lead to death. [8] As a result, it's important to find effective ways in stopping opioid addiction and overdose.

There are many causes behind opioid overdose, those we may think of including for one to take opioid to get high or for one to take extra doses of a prescription opioid. However, beside educating and helping patients by holding opioid overdose prevention programs, we believe that it's also important to consider the source, the opioid prescribers.

B. Predicting Opioid Prescribers

For starters, machine learning can be used to identify significant predictors. Robert C Schell and his teammates included LASSO and random forest to identify neighborhood-level predictor for opioid overdose. They also used K-fold cross-validation to assess predictive performance. The variables that are most influential included a range of dimensions of socioeconomic status, such as education, income, wealth etc. Besides these commonly perceived variables, other seemingly less relevant ones such as residential stability, race and social isolation also plays an important role.[9]

There are many scholars who used similar approach to investigate the related topic. One group of researchers utilized machine learning to predict opioid overdose risk among fee-for-service Medicare beneficiaries. They randomly split more than 56000 samples into training, testing and validations sets of equal size. For variable selection, there are ones similar to ours such as sociodemographics, patterns of opioid use, and practitioner-level, but these measures were recorded in 3-month windows so they are more detailed. In terms of methodology, they also used LASSO, logistic regression and random forest, but incorporated gradient boosting machine, and deep neural network. For outcomes, they reported C statistics of each model. Though DNN can classify each risk groups and identify the subgroups having the highest chance of overdose, the positive predicative value is only 0.18 percent, meaning that it is very poorly at predicting true overdoses.[7]

Another group of researchers used machine learning to predict opioid involvement in unclassified drug overdoses and the number of fatal opioid overdoses from 1999 to 2016. They established five models of different sets of predictors using logistic regression and random forest to compare their performances, and use the best one to estimate opioid-related mortality. The measures they used are total predictive accuracy, receiver operating characteristic and F1 score. They successfully classified different types of overdoses and discovered a significant rising trend of opioid ones, but logistic regression and random forests have similar level of predictive accuracy, which coincides with our finding. [5]

III. DATA

The dataset we're using is cleaned and compiled by "ALAN 'AJ' PRYOR" on kaggle[3], which is a small dataset

sourced from cmc.gov[1]. The full data contains about 24 million prescription instances, and the author limited 1 million total unique prescribers down to 25,000 ones. We decide to build our model based on the sampled dataset considering computational cost.

The dataset includes three parts in total:

- prescriber-info.csv: contains basic information of some medical professionals in 2014 in the United States and their prescription records for hundreds of common opioid and non-opioid drugs
- overdoses.csv: contains information on opioid related drug overdose fatalities, which is sorted by state abbreviations
- opioids.csv: contains names of all opioid drugs in the prescriber-info.csv dataset

There are a total of 257 variables from the first two datasets (prescriber-info.csv and overdoses.csv) as input variables, including:

- NPI – unique National Provider Identifier number
- Gender - male or female
- State - U.S. State by abbreviation
- Credentials - set of initials indicative of medical degree
- Specialty - description of type of medicinal practice
- A long list of drugs (see Appendix A for a full list of variables) with numeric values indicating the total number of prescriptions written for the year by that individual
- Death rate - ratio of death caused by overdose to population every state

And the target variable we are going to predict is:

- Opioid.Prescriber - a boolean label indicating whether or not that individual prescribed opiate drugs more than 10 times in the year

A. Data preprocessing

We first do further cleaning on the prescriber_info.csv dataset.

1) Gender: The "gender" feature consists of two categorical values "M" and "F". We apply binary encoding on it to make it more readable for our model.

2) State: The "State" feature consists of 57 categorical values. After cleaning, besides rows containing the 50 states values, we eliminate six rows that contain 8 of the categorical values, and preserve about 300 rows that contain values "DC" and "PR" standing for Washington D.C. and Puerto Rico. Then we apply apply one-hot encoding on this feature, resulting in a total of 52 distinct columns with binary values.

3) Credentials: The "Credentials" feature consists of 888 categorical values. After we apply further cleaning using regular expression operations, there are still 618 values, so we ultimately decide to drop this feature because of its computation cost and the possible noise it would cause to our model.

4) Specialty: The "Specialty" feature consists of 109 categorical values. After counting and graphing, we decide to combine those whose frequency is less than 50 and convert it to a separate label "Others". (See Fig 1.)

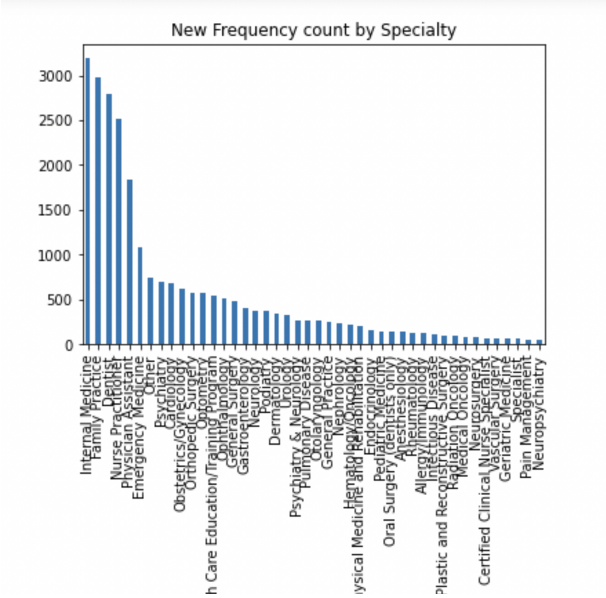

Fig. 1. Specialty Frequency plot

5) drug List: There are 250 features representing 250 different kinds of drugs, and each one is made up with numeric values. After checking, there is no missing values, and we did not do further processing on this.

B. Exploratory Data Analysis

1) Opioid Prescriber: We first look at the our target variables, which is Opioid.Prescriber feature in prescriber_info.csv. There are a total of 24992 observations, and after doing value count on Opioid.Prescriber column, we see that there are 58.82% Opioid Prescribers among all observations. We then further plot on the percentage of Opioid Prescribers for top 30 Specialities that have the most Opioid Prescribers. (See Fig. 2)

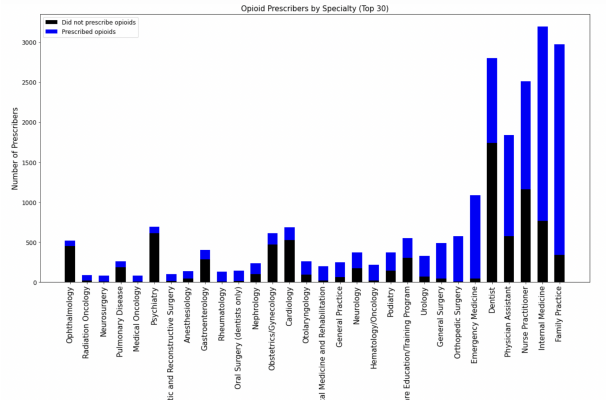

Fig. 2. Opioid Prescriber per Specialty plot

2) State: We first utilize the "population", "Death", and "State" columns in overdose.csv to plot a death (caused by opioid overdose) rate per state.(See fig. 3)

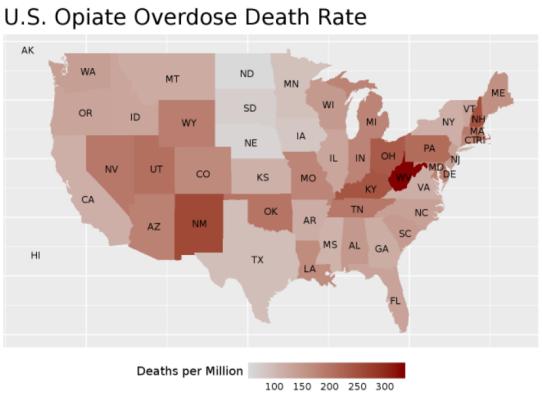

Fig. 3. Specialty Frequency plot

We connect the prescriber_info.csv with the overdose.csv through state abbreviations, so as to add "death_rate" feature into our feature list.

3) Opioids: We connect the prescriber_info.csv with opioids.csv to see how many drugs listed in our main dataset are opioids. The result is:

- HYDROMORPHONE.HCL
- TRAMADOL.HCL
- FENTANYL
- OXYCONTIN
- HYDROCODONE.ACETAMINOPHEN
- METHADONE.HCL
- ACETAMINOPHEN.CODEINE
- MORPHINE.SULFATE
- OXYCODONE.ACETAMINOPHEN
- OXYCODONE.HCL
- MORPHINE.SULFATE.ER

So the insight we may get from this is that, there are 11 opioids out of 250 types of drugs in our features. Then after counting opioids prescribed for each observations, we made a heatmap plotting the relationship between prescribing opioid amount and being Opioid Prescriber.(See Fig.4)

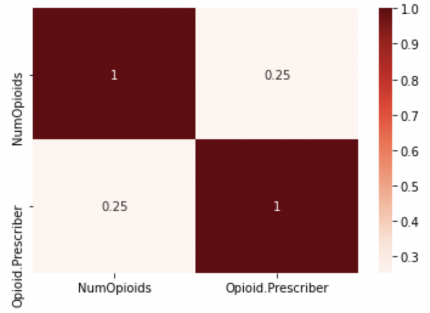

Fig. 4. Correlation between Opioid amount and Opioid Prescriber

IV. Goals

In this project, we are conducting a predicative analysis to identify the frequent prescribers of opioid drugs among millions of medical professionals. Our main goal is to develop a model that gives the highest possible prediction accuracy and sensitivity in predicting these prescribers. This goal is fundamental and very essential for research to identify the over-prescribers because it is a very complex to identify prove a doctor is over-prescribing where large amount of investigation are required. Thus, it is not possible for the authority to conduct such investigation on every prescribers in United States. However, a doctor could be a over-prescribe only if they are frequently prescribing opioid drugs; therefore, if we can first rapidly and accurately distinguish 'frequent opioid prescribers' from others - which is much easier than distinguishing over-prescribers - we can largely reduce the workload for experts by reducing the number of investigation needed.

As for the model performance, we expect our data cleaning and feature engineering will able improve the performance of our model. In particular, we expect we can improve the precision and sensitivity of our model by bringing in additional information of death rate by state from the overdoses.csv dataset. As for dimension reduction, we expect it can help in producing good visualization result for us to interpret the dataset, while it might not be able to improve our model performance largely as the dataset is already selected so the model is not likely to overfit. In terms of overall performence, we expect our model to achieve precision and sensitivity higher than 90%.

V. Results

A. Dimensionality Reduction

Since our data has over 300 attributes and is in high dimension, we want to know if it is possible to reduce some redundant features, and to visualize our data and have an idea of classification through this process. We first use lasso regression to do feature selection, and it turns out that only 7 out of 346 coefficients could be reduced to 0, which is far from what we expected.

Next, we implement PCA to do visualization and dimensionality reduction. First, we create an elbow plot (see in Fig. 5) to determine how many principal components we should keep to capture most of the variation in the data.

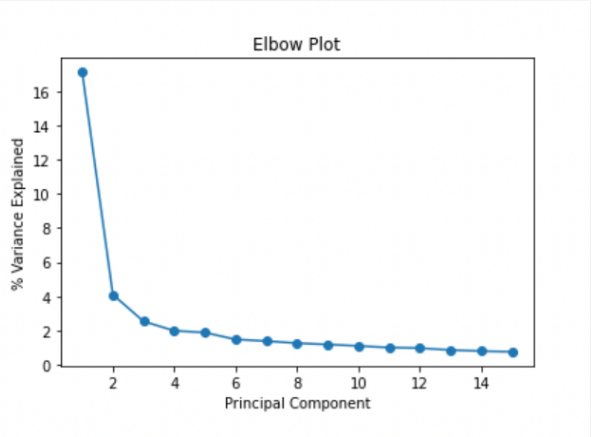

Fig. 5. Elbow Plot

In the graph above, we can observe that 2 and 3 is on the point of inflection on the curve, so we are going to choose 2-d and 3-d graphs to visualize our data.

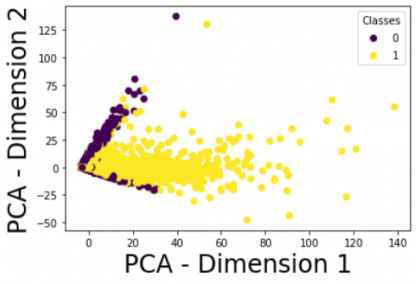

Fig. 6. 2D and 3D Map of Opioid Prescriber

From the 2-d map (see in Fig. 6) above, we can clearly find different patterns between the two classes distribution. The true class of opioid prescribers is mainly distributed in the central place, and the false class is scattered as two lines at both sides.

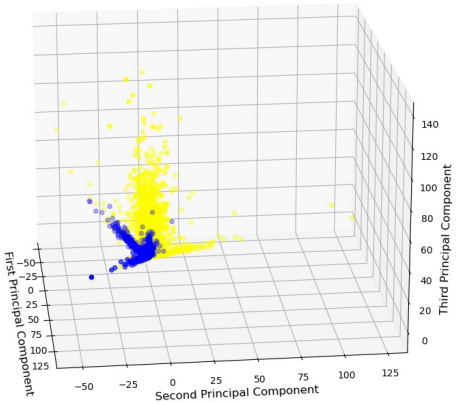

Fig. 7. 3D Map of Opioid Prescriber

A more obvious pattern can be observed when we change it to the 3-d visualization (see in Fig. 7). There is a clear boundary between the two classes. As a result, we can conclude that there is a good separation between the two classes from this visualization.

Using PCA, we also try to eliminate the number of attributes to 211, while still explaining 90% of the variance in the target variable. However, compared to the original number of attributes, reducing features from 346 to 211 is not a very significant improvement at a cost of losing 10% explaining ratio of variance. Thus, we will not use this reduction in our prediction model.

In conclusion, we do not find an effective way to reduce the dimension of features. It may also be a reflection that there is few redundant variables in our dataset. However, PCA does provide a good insight into data visualization, and it shows that there is boundary between the two target groups we will try to classify.

B. Predictive Models

Our team uses 6 basic machine learning models to predict whether the doctor is a Opioid-Prescriber and using the classification-report of confusion matrix to obtain the detailed conclusion of each model. For each of the parameters in the table, we first declare their definitions here:

- Target variable: Opioid.Prescriber (frequent (1) or not (0)).
- Overall precision: Percentage of correct prediction of both 1 and 0 .
- Precision of 1: Percentage of correct prediction of frequent opioid prescribers.
- Recall: Percentage of predicted frequent opioid prescribers out of all frequent opioid prescribers in real situation.
- F1 score: Harmonic mean of precision and recall.

With implementing Logistic Regression, Decision Tree, Random Forest, KNN Classifier, SVM Classifier and Gaussian Naive Bayes, we get all the detailed parameters and predictions. Concluding from the following table:

- Random Forest has the highest overall precision.
- Logistic Regression and SVM Classifier have the highest precision of 1s.
- Decision Tree and Random Forest have the highest Recall.
- Logistic and Random Forest have the highest F-1 score.
- Gaussian Naive Bayes has the highest run time.

TABLE I
PRECISION OF PREDICTIONS

Model's name	Overall precision	Precision of 1
Logistic Regression	0.921	0.98
Decision Tree	0.904	0.93
Random Forest	0.923	0.96
KNN Classifier	0.872	0.95
SVM Classifier	0.881	0.98
Gaussian Naive Bayes	0.91	0.95

After taking into full consideration what we want to achieve, the group agreed that the best model is Random Forrest. Although the running time is 3.9 seconds, from a combined perspective, it performs best recall and F1 scores among all models. These indicate that it has the ability to not only accurately predict frequent opioid users, but also has overall predictive accuracy.

If people want to take a look at each model's performance, we could using the ROC curve. The ROC (receiver operating characteristics) curve displays the true positive rate (sensitivity) agianst the false positive rate (1-specificity). The closer the curve follows the left hand border and then the top left border of the ROC space, the more accurate the model. In this figure we can see that Random Forest performs the best, closest to the top left corner of the figure, so we can also

TABLE II
CLASSIFICATION RESULT

Model's name	Recall	F-1 score	time(sec)
Logistic Regression	0.88	0.93	0.6
Decision Tree	0.91	0.92	0.8
Random Forest	0.91	0.93	3.9
KNN Classifier	0.82	0.88	5.4
SVM Classifier	0.82	0.89	48.6
Gaussian Naive Bayes	0.89	0.92	0.2

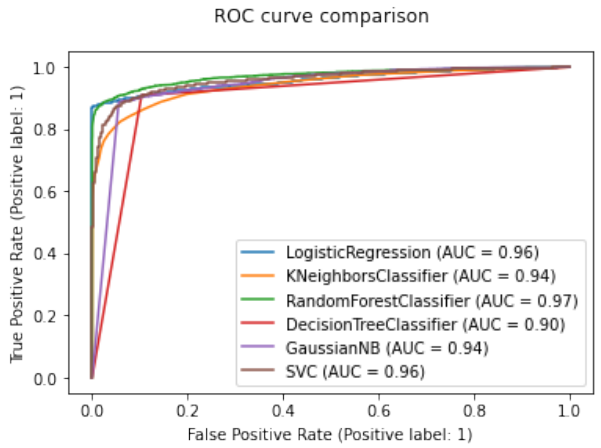

Fig. 8. Comparison of ROC Curve

conclude that Random Forest should be chosen as the model for predicting frequent opioid prescriber for this project.

VI. CONCLUSIONS

The aim of our project is to identify frequent opioid prescribers using machine learning. At the start, we did data cleaning and EDA to the original dataset. We dropped irrelevant columns and rows containing missing data, fixed typos, calculated the death rate per capita related to overdoses in each state then normalized it, and vectorized all of the categorical data.

Next, we tried to use PCA for dimension reduction. In both 2D and 3D graphs, we can see clear distinctions between the two clusters of out target variable. However, within a total of 346 variables, 211 of them are still needed to explain 90 percent of the variance. Therefore, we chose to use the original cleaned data for further analysis. In the prediction process, we established several models for performance comparisons: Logistic Regression, Decision Tree, Random Forrest, KNN Classifier, SVM Classifier and Gaussian Naïve Bayes. We adjusted the parameters of these models to get the best overall performance and did side-by-side comparisons to find the top 3 for each criterion: The models with highest overall precisions are: Random Forrest: 0.923, Logistic Regression: 0.921 and Gaussian Naïve Bayes: 0.910. The models with highest precision for predicting frequent opioid prescribers (1) are: Logistic Regression: 0.98, SVM: 0.98 and Random Forrest: 0.96. The models with highest recall for predicting frequent opioid prescribers (1) are: Decision Tree: 0.91, Random Forrest: 0.91, and Gaussian Naïve Bayes: 0.89. The models with highest F1 scores are: Random Forrest: 0.93, Logistic Regression: 0.93, Decision Tree: 0.92 and Gaussian Naïve Bayes: 0.92. The models with the fastest runtime are Gaussian Naïve Bayes: 0.2s, Logistic Regression: 0.6s, Decision Tree: 0.8s.

After thorough consideration with respect to the goal we are trying to achieve, we decided that the best model overall is Random Forrest. Although it might be a little time costly (3.9s), it produced the best recall and F1 score among all the models. These indicate not only its outstanding capability to accurately predict frequent opioid prescribers, but also a well-rounded predictive accuracy. More importantly, it provides the extra benefit of identifying the influential attributes, which marks its potential stability with all kinds of datasets.

Shedding light into the medical and data science community, our project is innovative and meaningful. Through a systematic procedure, our project guarantees more stability, better precision and higher efficiency than manual predictions by data experts. At last, predicting frequent opioid prescribers narrow down the pool of opioid overdose, creating a solid base for future research that focus on reducing overdose mortality rate.

APPENDIX A

, ACETAMINOPHEN.CODEINE, ACYCLOVIR, ADVAIR.DISKUS, AGGRENOX, ALENDRONATE.SODIUM, ALLOPURINOL, ALPRAZOLAM, AMIODARONE.HCL, AMITRIPTYLINE.HCL, AMLODIPINE.BESYLATE, AMLODIPINE.BESYLATE.BENAZEPRIL, AMOXICILLIN, AMOX.TR.POTASSIUM.CLAVULANATE, AMPHETAMINE.SALT.COMBO, ATENOLOL, ATORVASTATIN.CALCIUM, AVODART, AZITHROMYCIN, BACLOFEN, BD.ULTRA.FINE.PEN.NEEDLE, BENAZEPRIL.HCL, BENICAR, BENICAR.HCT, BENZTROPINE.MESYLATE, BISOPROLOL.HYDROCHLOROTHIAZIDE, BRIMONIDINE.TARTRATE, BUMETANIDE, BUPROPION.HCL.SR, BUPROPION.XL, BUSPIRONE.HCL, BYSTOLIC, CARBAMAZEPINE, CARBIDOPA.LEVODOPA, CARISOPRODOL, CARTIA.XT, CARVEDILOL, CEFUROXIME, CELEBREX, CEPHALEXIN, CHLORHEXIDINE.GLUCONATE, CHLORTHALIDONE, CILOSTAZOL, CIPROFLOXACIN.HCL, CITALOPRAM.HBR, CLINDAMYCIN.HCL, CLOBETASOL.PROPIONATE, CLONAZEPAM, CLONIDINE.HCL, CLOPIDOGREL, CLOTRIMAZOLE.BETAMETHASONE, COLCRYS, COMBIVENT.RESPIMAT, CRESTOR, CYCLOBENZAPRINE.HCL, DEXILANT, DIAZEPAM, DICLOFENAC.SODIUM, DICYCLOMINE.HCL, DIGOX, DIGOXIN, DILTIAZEM.24HR.CD, DILTIAZEM.24HR.ER, DILTIAZEM.ER, DILTIAZEM.HCL, DIOVAN, DIPHENOXYLATE.ATROPINE, DIVALPROEX.SODIUM, DIVALPROEX.SODIUM.ER, DONEPEZIL.HCL, DORZOLAMIDE.TIMOLOL, DOXAZOSIN.MESYLATE, DOXEPIN.HCL, DOXYCYCLINE.HYCLATE, DULOXETINE.HCL, ENALAPRIL.MALEATE, ESCITALOPRAM.OXALATE, ESTRADIOL, EXELON, FAMOTIDINE, FELODIPINE.ER, FENOFIBRATE, FENTANYL, FINASTERIDE, FLOVENT.HFA, FLUCONAZOLE, FLUOXETINE.HCL, FLUTICASONE.PROPIONATE, FUROSEMIDE, GABAPENTIN, GEMFIBROZIL, GLIMEPIRIDE, GLIPIZIDE, GLIPIZIDE.ER, GLIPIZIDE.XL, GLYBURIDE, HALOPERIDOL, HUMALOG, HYDRALAZINE.HCL, HYDROCHLOROTHIAZIDE, HYDROCODONE.ACETAMINOPHEN, HYDROCORTISONE, HYDROMORPHONE.HCL, HYDROXYZINE.HCL, IBANDRONATE.SODIUM, IBUPROFEN, INSULIN.SYRINGE, IPRATROPIUM.BROMIDE, IRBESARTAN, ISOSORBIDE.MONONITRATE.ER, JANTOVEN, JANUMET, JANUVIA, KETOCONAZOLE, KLOR.CON.10, KLOR.CON.M10, KLOR.CON.M20, LABETALOL.HCL, LACTULOSE, LAMOTRIGINE, LANSOPRAZOLE, LANTUS, LANTUS.SOLOSTAR, LATANOPROST, LEVEMIR, LEVEMIR.FLEXPEN, LEVETIRACETAM, LEVOFLOXACIN, LEVOTHYROXINE.SODIUM, LIDOCAINE, LISINOPRIL, LISINOPRIL.HYDROCHLOROTHIAZIDE, LITHIUM.CARBONATE, LORAZEPAM, LOSARTAN.HYDROCHLOROTHIAZIDE, LOSARTAN.POTASSIUM, LOVASTATIN, LOVAZA, LUMIGAN, LYRICA, MECLIZINE.HCL, MELOXICAM, METFORMIN.HCL, METFORMIN.HCL.ER, METHADONE.HCL, METHOCARBAMOL, METHOTREXATE, METHYLPREDNISOLONE, METOCLOPRAMIDE.HCL, METOLAZONE, METOPROLOL.SUCCINATE, METOPROLOL.TARTRATE, METRONIDAZOLE, MIRTAZAPINE, MONTELUKAST.SODIUM, MORPHINE.SULFATE, MORPHINE.SULFATE.ER, MUPIROCIN, NABUMETONE, NAMENDA, NAMENDA.XR, NAPROXEN, NASONEX, NEXIUM, NIACIN.ER, NIFEDICAL.XL, NIFEDIPINE.ER, NITROFURANTOIN.MONO.MACRO, NITROSTAT, NORTRIPTYLINE.HCL, NOVOLOG, NOVOLOG.FLEXPEN, NYSTATIN, OLANZAPINE, OMEPRAZOLE, ONDANSETRON.HCL, ONDANSETRON.ODT, ONGLYZA, OXCARBAZEPINE, OXYBUTYNIN.CHLORIDE, OXYBUTYNIN.CHLORIDE.ER, OXYCODONE.ACETAMINOPHEN, OXYCODONE.HCL, OXYCONTIN, PANTOPRAZOLE.SODIUM, PAROXETINE.HCL, PHENOBARBITAL, PHENYTOIN.SODIUM.EXTENDED, PIOGLITAZONE.HCL, POLYETHYLENE.GLYCOL.3350, POTASSIUM.CHLORIDE, PRADAXA, PRAMIPEXOLE.DIHYDROCHLORIDE, PRAVASTATIN.SODIUM, PREDNISONE, PREMARIN, PRIMIDONE, PROAIR.HFA, PROMETHAZINE.HCL, PROPRANOLOL.HCL, PROPRANOLOL.HCL.ER, QUETIAPINE.FUMARATE, QUINAPRIL.HCL, RALOXIFENE.HCL, RAMIPRIL, RANEXA, RANITIDINE.HCL, RESTASIS, RISPERIDONE, ROPINIROLE.HCL, SEROQUEL.XR, SERTRALINE.HCL, SIMVASTATIN, SOTALOL, SPIRIVA, SPIRONOLACTONE, SUCRALFATE, SULFAMETHOXAZOLE.TRIMETHOPRIM, SUMATRIPTAN.SUCCINATE, SYMBICORT, SYNTHROID, TAMSULOSIN.HCL, TEMAZEPAM, TERAZOSIN.HCL, TIMOLOL.MALEATE, TIZANIDINE.HCL, TOLTERODINE.TARTRATE.ER, TOPIRAMATE, TOPROL.XL, TORSEMIDE, TRAMADOL.HCL, TRAVATAN.Z, TRAZODONE.HCL, TRIAMCINOLONE.ACETONIDE, TRIAMTERENE.HYDROCHLOROTHIAZID, VALACYCLOVIR, VALSARTAN, VALSARTAN.HYDROCHLOROTHIAZIDE, VENLAFAXINE.HCL, VENLAFAXINE.HCL.ER, VENTOLIN.HFA, VERAPAMIL.ER, VESICARE, VOLTAREN, VYTORIN, WARFARIN.SODIUM, XARELTO, ZETIA, ZIPRASIDONE.HCL, ZOLPIDEM.TARTRATE

REFERENCES

[1] Medicare provider utilization and payment data: Part d prescriber. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Part-D-Prescriber.

[2] Prescription opioids. https://www.cdc.gov/opioids/basics/prescribed.html, Aug 2017.

[3] U.s. opiate prescriptions/overdoses. https://www.kaggle.com/datasets/apryor6/us-opiate-prescriptions, Nov 2019.

[4] Caleb J. Banta-Green, Joseph O. Merrill, Suzanne R. Doyle, Denise M. Boudreau, and Donald A. Calsyn. Opioid use behaviors, mental health and pain—development of a typology of chronic pain patients. *Drug and Alcohol Dependence*, 104(1-2):34–42, 2009.

[5] Andrew J. Boslett, Alina Denham, and Elaine L. Hill. Using contributing causes of death improves prediction of opioid involvement in unclassified drug overdoses in us death records. *Addiction*, 115(7):1308–1317, 2020.

[6] Benjamin Butanis. What are opioids? https://www.hopkinsmedicine.org/opioids/what-are-opioids.html, Apr 2018.

[7] Wei-Hsuan Lo-Ciganic, James L. Huang, Hao H. Zhang, Jeremy C. Weiss, Yonghui Wu, C. Kent Kwoh, Julie M. Donohue, Gerald Cochran, Adam J. Gordon, Daniel C. Malone, and et al. Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. *JAMA Network Open*, 2(3), 2019.

[8] Assistant Secretary of Public Affairs (ASPA). Hhs.gov/opioids: The prescription drug amp; heroin overdose epidemic. https://www.hhs.gov/opioids/.

[9] Robert C Schell, Bennett Allen, William C Goedel, Benjamin D Hallowell, Rachel Scagos, Yu Li, Maxwell S Krieger, Daniel B Neill, Brandon D Marshall, Magdalena Cerda, and et al. Identifying predictors of opioid overdose death at a neighborhood level with machine learning. *American Journal of Epidemiology*, 191(3):526–533, 2021.