Strategy on Early Intervention: Using Logistic Regression with Elastic Net and Random Forest Models to Predict Risks of Continued Long-term Opioid Therapy (LTOT)

Humana-Mays Healthcare Analytics
2019 Case Competition

October, 15, 2019

Executive Summary

Opioid therapy is an opioid-related treatment for chronic pain management. It is prescribed to patients with ongoing severe pain. The goal of opioid therapy is to reduce pain and to increase daily function. Since 1980, opioid therapy has been widely introduced and applied to cancer and non-cancer treatments. Because opioid itself is highly addictive, prescribing opioid therapy requires physicians to have adequate knowledge and to follow strict procedures. However, in real-world practices, these requirements were hard to fulfill due to many reasons, including the nature of chronic pain and the nature of opioid treatment. This caused wide incidents of opioid prescription abuse. In 2017, the United States government declared the opioid epidemic as a public emergency. The serious situation of drug control leads to many researches on risk prediction. Researchers are eager to find out ways to predict whether a patient who is currently receiving Long-term Opioid Therapy (LTOT) would be addictive in a given period of time. Many studies have been designed and implemented. This study is one of them.

This study focuses on predicting risks of patients who receive continued LTOT regarding opioid overdose and drug addiction so that early intervention strategies can be developed. It applies machine learning models and techniques to establish a ranking system. Then, it sorts target patients with various risk levels from high to low. The result shows that among 6,000 patients, one with the highest risk score would have a 97.5% probability of becoming addicted to opioid, which is to consume opioid more than 90% of the time within 6 months. The patient with the lowest score has a 17.5% probability of being addicted. The overall accuracy (AUC) of models we tested is 79.44% with a False Negative Rate (FNR) of 0.22.

During the process, we define several Key Performance Indicators (KPIs). The first group of KPI is about RX claims that are directly related to opioid prescriptions, including total times of opioid prescription received, total opioid dosage amount, and total opioid quantity amount. Meanwhile, some non-opioid drug usage also affect the target. These indicators measure how often and how many opioid/non-opioid prescriptions do patients receive during a certain period of time. The second group of KPI is about coverage, including fully claim paid amount and responsible amount (Co-pay). Since all patients in the data are fully covered, we believe that financial privileges may provide as incentives to encourage patients who do not need adequate opioid prescriptions to request extra claims.

Several recommendations are suggested based on our study. First, the dosage amount of opioid/non-opioid prescription should be reduced. Medical providers should increase

interactions with patients, including regular check-in tests. Second, terms of conditional reimbursement should be included in long-term opioid therapy coverage. Thrid, healthcare providers should conduct more research on risk prediction using machine learning techniques in order to target potential high-risk groups.

Introduction

Ever since John Bonica identified pain as the fifth vital sign and claimed that pain treatment is considered as a proper patient care, pain relief treatment was gradually introduced as part of the treatment cycle. Although there were guidelines and strict regulations about opioid pain treatment, various concerns regarding drug abuse have been discussed as pain treatment techniques develop. However, due to positive feedback from patients incorporating pain relieving drugs during their treatments and with the promotion of opioid usage from health care companies, opioid-related pain treatment plans were widely applied. In the 1980s, opioids were included in World Health Organization cancer pain treatment guidelines. This marks the history of opioid recognized as an important essence of pain treatment.

Fast forwarding to 2018, a study published under *National Center of Biotechnology Information* conducted a research of top 200 prescribed drugs in the US. The study showed that out of 200 most prescribed drugs, 10 were in the category of opioid and one opioid, Acetaminophen, was ranked in the top 10 most prescribed drugs. While Fentanyl was not ranked in the top 200, it is a substance that is 50 times stronger than Heroin and it is still prescribed to patients for the purpose of alleviating chronic pains. With such strong substance being prescribed to patients, there arises a serious concern in the general public that the misuse of opioid treatment contributes to drug addiction, and even to the worst case, death. Professor Judith Feinberg states that "Most insurance, especially for poor people, won't pay for anything but a pill." Some argue that although there are alternative pain relieving methods available, given the affordable pricing, most patients would choose to take opioid for pain relief. As a result, in 2018, more than 68,000 deaths were reported as a result of opioid overdose.

A series of effects are observed from opioid overdose, including oversedation, respiratory depression, and even death. CDC guidelines advise a daily dose of exceeding 90 mg morphine per day would increase the risk of overdose, while study has shown that 23% of patients are receiving doses above 100 mg of morphine equivalents per day. In addition, patients are exposed to side effects due to Long Term Opioid Therapy (LTOT), such as endocrinopathies, sleep disordered breathing, constipation, fractures, falls, and mental status changes. Therefore, It is important for us to take immediate action on identifying patients in LTOT who would have a higher risk of drug addiction due to opioid overdose.

The problem that we are facing is complex. The side effects, along with the worrisome amount of dosage prescribed to patients and number of death, pull an alert to us. These are the last

things we want to see to our family, friends and even to ourselves. As data scientists, it is our responsibility to participate and help in every shape and form we could to draw public awareness about opioid overdose and drug addiction. One action we can take is to study the patients that have prescribed to opioids in the past, with the help of data collection from Humana, and use machine learning techniques to construct a predictive model to see if patients who are prescribed to opioid would be later identified as LTOT, as someone who has consumed opioid more than 90% of the time within 6 months. The goal of the construction of this model is to successfully predict patients who would be identified as LTOT so that early intervention can be performed to those patients to avoid misuse and addiction.

During our study, we encountered several issues. One of the biggest issues we overcome in this study is data cleaning. Since we are given with a raw dataset, there is no clear target variable. There are same attributes under different rows have inconsistent observations due to the nature of healthcare data format. Moreover, The raw dataset is large. We are given 14,000 unique patient IDs in the training datasets, and each unique ID has multiple entries. Therefore, efficient data cleaning is a vital step in this task. To address that, we first extracted all features that show high target relevance for all events. Then, we quantitatively reduced or added features after evaluating the initial cleaned dataset. More details of data cleaning can be found in the following section under **Methods**.

Methods

Data Cleaning

We believe the quality of our analysis depends on how we processed the data. While there are obviously various combinations of features that can be extracted from the original dataset, we decided not to force the extraction of all potential features at once for the consideration of computation cost and feature interpretation. There are 16 types of events exist in the dataset, with different attributes associated with them. The first round of feature extraction was mainly based on the corresponding relations of attributes of an event with the target. In detail, for each event, we only took those attributes that have intuitively possible interpretations of continued LTOTs. Specifically, for RX-Claim related events, we further divided the event based on whether it is a claim for opioid (Whether there are additional attributes like *MME*).

Take the event **`RX Claim - Paid`** for instance. First, the event is categorized as **`RX Claim_Paid_Opioid`** as opioid claim and **`RX Claim_Paid_Other`** as if it is a claim of other drugs. Then, the original feature extraction of **`RX Claim_Paid_Opioid`** and **`RX Claim_Paid_Other`** is the total times of such event, total cost, total net paid amount, and total member responsible amount (**`total Pay Day Supply Count`**, **`total Payable Quantity`** and **`total MME`** is added if it is an opioid claim). However, attributes like **`Brand Name of Drugs`** was not extracted during the first round extraction as they make less intuitive sense on their impacts on long-term use of opioid therapies.

The target, whether the patient suffers from long-term opioid therapies, is labeled based on the given definition. From the anchor day, we extend 180 days and check whether this patient has received opioid for more than 162 days. The patient is labeled as 1 if the above-mentioned condition is satisfied. Otherwise, we start from 90 days after the anchor day and find the new anchor day based on the given *opioid naive* definition. Starting from the new anchor day, we repeat the previous method to check whether the patient has long-term opioid therapies. We iteratively complete this label check until the last day of the record for one patient. If the patient is not labeled as 1 in this process, then he will be labeled as 0, which means he does not suffer from long-term opioid therapies.

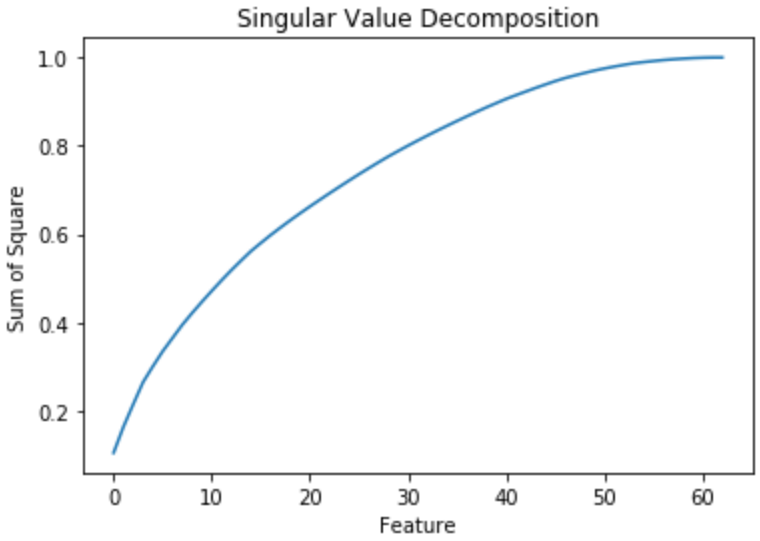
After initial data cleaning, the dataset is randomly split into training data and test data with a proportion of **8 to 2**.

Feature Engineering

Considering both model performance and computation cost, we want to only include features that can optimize the information they have for the target. The feature engineering process includes both feature reduction and feature addition.We experimented two different methods of feature reduction, **Singular Value Decomposition (SVD)** and **Recursive Feature Elimination (RFE)**, we also used SVD to add more features.

The first feature reduction method, SVD, is to factorize a matrix to its lower rank approximation. By decomposing the dataset, we extract after the data cleaning process. Throughout this process, we can evaluate percentages of features extracted and corresponding features have kept most variance within the original dataset. We expect that by incorporating partially of these attributes can represent the original dataset, while reducing noise, thus possibly yield better model performance. **Figure 1** shows the result of SVD of our extracted dataset. From the figure, **40 features** have covered **80 percent** of sum of square in the data, which are used later for experiment of model performance. The collection of all 40 features used after SVD is attached in **Appendix**.

Figure 1: Singular Value Decomposition



Another feature reduction method we have applied is RFE, which is a model-based algorithm. It tends to recursively use smaller sets of features to optimize the model performance. RFE refers to certain ranking criterion of features, namely feature importance of coefficients for models we have used, to select the best subset of features with the highest predictive power.

As we do not extract all possible features from original data at once, we add a couple of new features for further model examination based on SVD. Although SVD is mainly used for feature reduction, we can still identify top features from this process. Based on those features, we extract additional attributes as new features from related events. For example, based on the most related feature from SVD, **`total_RX_claim_paid_opioid`**, which is total times of receiving opioid drug before anchor day, we further add **`opioid GPI drug class description`** and **`drug group description`** as new feature for further model evaluation.

Models and Evaluation

The metric we use to evaluate our model is **Area Under the Curve (AUC)**. Since our best model has an AUC of 0.7944 , it represents that there is 79.44% of the chance that the model can correctly distinguish between LTOT patients and non-LTOT patients. In addition, **False Negative should be weighted more than False Positive**. False Negative would be true LTOT patients that the model failed to detect. Those patients are slipping away from the model and are likely to be addicted to opioids, while proactive steps could have been done if the model can successfully identify patients with LTOT in the first place. From Humana's perspective, obtaining a high **False Negative Rate (FNR)** would cause the company to pay for addiction that patients are developing; while from the perspective of patient, having a high FNR would fail to prevent patients becoming LTOT. Therefore, a False Negative harms both Humana and the patients and we take into consideration by maintaining a relatively low FNR while selecting the best model.

With the split of training and test data, we use a five fold cross validation methods to grid search the hyperparameters that can optimize the model AUC for each model. Meanwhile, we assign higher class weights on patients that have long-term opioid therapy in the cross validation in order to lower the FNR, aiming to reduce the risk of misclassifying patients who actually suffer from long-term opioid therapy. Finally, we use the test data to examine the model performance.

When we consider which models should we implement, we take three aspects into account: the predict power, the model interpretability, and the model's ability to work with high number of features, as shown in our cleaned dataset. Starting from these characteristics, we apply **Logistic Regression with Elastic Net** penalty as our baseline model. And we experiment **Random Forest, Gradient Boosting Machine and Multilayer Perceptron** and compare the test AUCs to determine the best model.

We choose Logistic Regression with Elastic Net as our baseline model because it is relatively simple to train as a linear model while Elastic Net combines Lasso property to work with many feature and Ridge property to avoid overfitting.

Random Forest and Gradient Boosting Machine are both tree-based algorithms with high model interpretability because they can calculate the feature importance based on information gain of the features. Meanwhile, Random Forest randomly samples subset of features to train to avoid overfitting and Gradient Boosting Machine tends to start from a weak model to yield better prediction result.

Multilayer Perceptron is a neural network model, that is always believed to have stronger predictive performance than traditional machine learning models. We build a four layer Multilayer Perceptron model to check whether it can give better results.

All these machine learning models are done in the fashion of five fold cross validation grid search, with **1:2 class weight** between non-LTOT target and LTOT target. Meanwhile, feature engineering that is discussed previously is implemented with model evaluation iteratively until we identify the best feature combinations and the best model based on the AUC and interpretability.

Key Performance Indicators

Given the serious situation of opioid abuse, there are ranges of problems we would like to solve in this study. Some assumptions have been raised, such as the main indicator that has led to patient addiction on opioid usage. Was the issue distributed evenly to the entire population, regardless of the patient's gender, income class and living area? Or, is it affected by the type of drugs they are prescribed to? Moreover, if patients that are identified as LTOT, is it due to truly in need of pain relieving treatment, or because they have already addicted to opioid.

Based on the above assumptions, we first analyzed the problem of patients with LTOT who are addicted to opioid with more details. Then, we redefined the problem with several Key Performance Indicators (KPIs) based on **feature importance** of the best Random Forest model (more details in **Results** section). The first group of KPI is about RX claims that are directly related to opioid prescriptions. With feature importance analysis, we examined the influence of different variables related to opioid prescription and determined three major KPIs: total times of opioid prescription received, total opioid dosage amount, and total opioid quantity amount. These indicators measure how often and how many opioid prescriptions do patients receive during a certain period of time. We believe LTOT patients can be prevented by controlling opioid supply pattern and amount.

The second group of KPI is about money. We apply fully claim paid amount and responsible amount (Co-pay) to our models as two key features. Since all patients in the data are fully covered with healthcare plans, we believe that financial privileges may provide as incentives to encourage patients who do not need adequate opioid prescriptions to request extra claims.

With the attributes given in this dataset, we are able to transform it into a data problem by identifying patients that will likely be addicted, who has a continuous use of an opioid medication with 90% days covered a 6-month period.

Results

Best results from each model can be found in the chart below. The results show that out of two feature reduction techniques we used, Random Forest using Recursive Feature Elimination (RFE) yields the highest AUC of **0.7944**. Taking advantage of random forest model, we identified top features that are the most useful for model construction. A detailed list of top features with their corresponding score is provided. The score ranges from 0 to 1. The higher the score, the more information the feature can provide to the model. The top 3 features are **`total_RX_claim_paid_supply_opioid`**, **`total_RX_claim_paid_quantity_opioid`** and **`total_RX_claim_paid_opioid`** (please refer to **Appendix** for feature explanations). **Figure 2** shows top 15 features with the highest feature importance extracted from the Random Forest model, while some of the features are trivial to us, such as **`total_RX_claim_paid_opioid`**, **`total_RX_claim_paid_supply_opioid`** and **`total_RX_claim_paid_quantity_opioid`**. There were some features, such as **`total_RX_claim_paid_other`**, which is the total time of a patient receiving non-opioid new drug. Further looking at the correlation between top attributes and our target LTOT in **Figure 3**, there is a positive correlation between **`total_RX_claim_paid_other`** and LTOT addiction, which we conclude the chance of LTOT addiction is higher when the total amount of drugs the patients receive increases. This sheds light to future study on severity of patients whose condition might correlates with LTOT addiction, since pain could accompany with severe health condition. In addition, the financial privilege, as stated in the previous section, is also shown within top features, such as **`total_RX_claim_paid_paid_opiod.`**

The confusion matrix of our best Random Forest model in **Figure 4** shows that the model could correctly identify **68%** of the patients to be identified as LTOT, while correctly identify **77%** of patients who are not LTOT. Since FNR is an indicator that patients who are truly LTOT however the model failed to identify by the model, causing a potential opportunity for Human healthcare to reach out to before patients are addicted to slip away. Therefore, as we evaluate the model, we are aiming to reduce FNR while maintaining a high TPR and FPR. With a FNR of 0.22, we should expect 22% of patients that will be identified as LTOT won't be classified by our model.

Table 1: Test AUC of Various Models

Test AUC	Singular Value Decomposition	Recursive Feature Elimination
Logistic Regression	0.7630	0.7758
Random Forest	**0.7944**	0.7941
Gradient Boosting Machine	0.7729	0.7676
Multilayer Perceptron	0.7724	N/A
NOTE: RFE not available for Multilayer Perceptron because of no coefficients or feature importance available in Python sklearn model		

Figure 2: Feature Importance from Random Forest Model

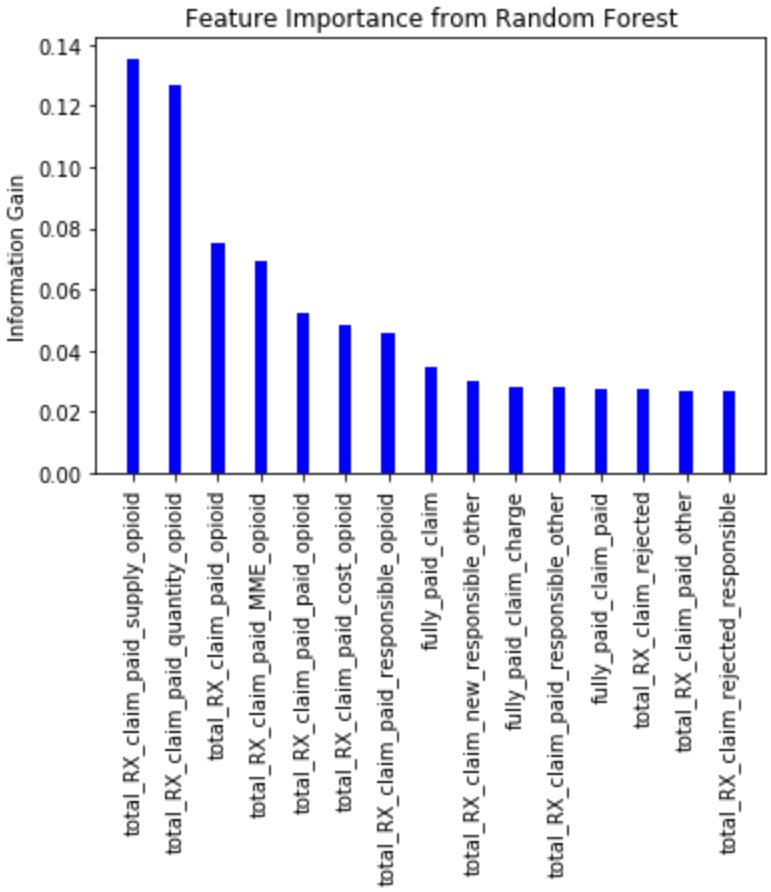


Figure 3: Feature Correlation regarding to Label

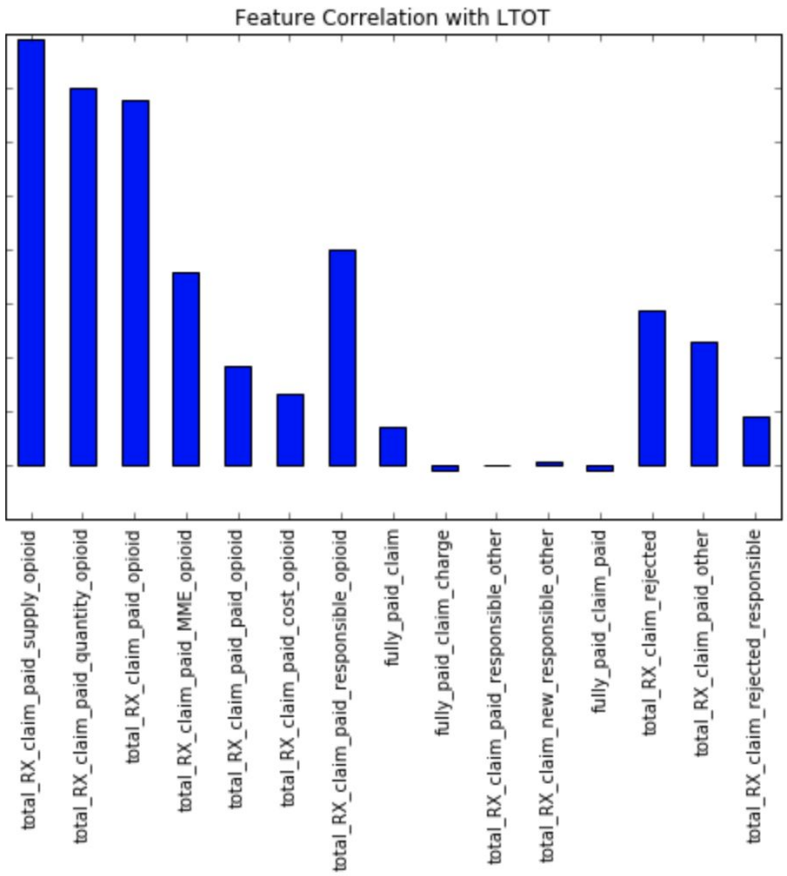
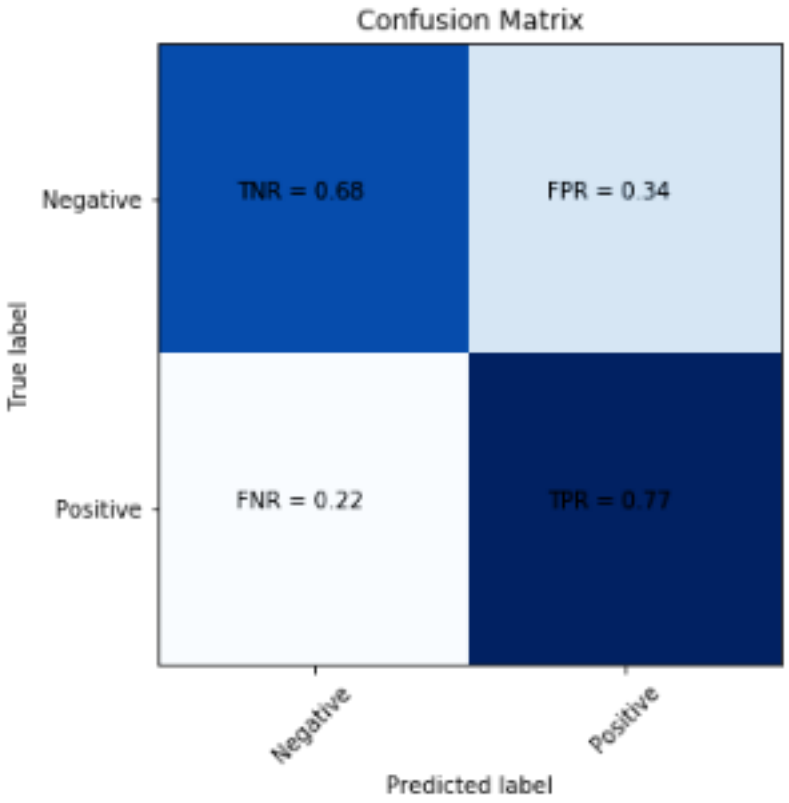


Figure 4: Confusion Matrix of the Best Model



Recommendations

Our results indicate that the risk of patients who are likely to be addicted to opioid is highly related to key performance indicators, including amount of opioid prescriptions received and amount of reimbursement claimed. Based on our findings, there are several recommendations we would like to address.

First, dosage is the key. Our study shows that the total opioid supply amount variable is amount the highest level of information gain. This means that the amount of opioid supply patients received is closely related to probability of opioid misuse. Meanwhile, the amount of other types of drugs also more or less affects the possibility of opioid misuse. Within a period of time, the more prescriptions a patient receives, the more likely he/she would get addicted to opioid. Therefore, several strategies should be implemented by healthcare and medical providers. First, the dosage amount of opioid/non-opioid prescription should be reduced. Regulations on opioid/non-opioid related prescriptions should be established to supervise physicians and to prevent them from over-prescribing drugs. Second, medical providers should increase follow up practices, including home visits, phone call follow-ups and regular check-in tests. With frequent follow-ups, medical providers can acquire first-hand data on patient's treatment, which can be referred for the adjustment of treatment plans. Through this process, opioid misuse due to over-prescribing can be reduced.

Second, coverage matters. Although full coverage gives significant financial and emotional alleviations to patients, it does potentially encourage patients to abuse it. With external financial incentives, such as illegal drug deals, patients may request additional prescriptions that they do not need by exaggerating their illnesses. To solve this, healthcare providers need to revise plans by adding terms of conditional reimbursement. For example, for patients who have received more than 90 days amount of opioid prescriptions, a check-in test is required before their claims can be processed. Patients are still fully covered. No additional financial burden is added. In this way, healthcare providers can contribute to the prevention of potential drug abuse by including the cost of follow-up tests, which are far lower than the extra social responsibility burden of new drug addicts.

Third, healthcare providers should conduct more research on risk prediction using machine learning techniques. Through modeling, healthcare providers can identify certain groups of patients with higher risks. Healthcare providers can benefit from prediction modeling by

concentrating on these patients and to reduce their chances of opioid addiction. It is a strategy to prevent healthcare providers from covering patient’s extra claims due to drug addiction.

Discussion

Our study applies machine learning techniques to predict the risk of patients who receive long-term opioid therapy. There are limitations in this study. First, the raw data we acquired are second-hand with adjustment to increase the contrast. This limits the study from accurately representing true positive results in real-world cases. The model representativeness in this study can be dramatically increased if more first-hand accurate data can be collected and added for training.

Second, previous studies on this topic frequently included patients' demographic information as key features. Demographic information, such as age, gender, and zip code, is highly related to model outcomes. For example, the age variable may have positive relationship since patients demand more pain treatments as age increases and as chronic disease develops. However, the raw dataset of this study does not include demographic information, causing major features to be missing.

Third, the model results of the study shows a False Positive Rate at 0.34. This is considered as relatively high because we lowered the False Negative Rate in our study. False Negative means that patients who have high risks are not identified by the model with similar influential level. We lowered False Negative Rate in order to reduce the chance of neglecting high risk individuals. Thus, the adjustment on model is a tradeoff to identify more LTOT addiction patients and to protect Humana's benefits.

Appendix

Table 2: Explanations of Feature Names

Feature Name	Explanation
`total_RX_claim_paid_opioid`	total times of receiving opioid
`total_RX_claim_paid_supply_opioid`	total opioid supply amount
`total_RX_claim_paid_quantity_opioid`	total opioid quantity amount
`fully_paid_claim`	fully paid claim event times
`total_RX_claim_new_other`	total times of receiving non-opioid new drug
`total_RX_claim_rejected`	total times of claim rejection
`total_RX_claim_paid_other`	total times of receiving non-opioid
`total_RX_claim_rejected_supply`	total rejected supply days of drugs
`inbound_call_by_prov`	total times of call by provider
`total_RX_claim_paid_responsible_opioid`	total opioid responsible payment
`total_RX_claim_paid_MME_opioid`	total opioid MME
`total_RX_claim_new_opioid`	total times of receiving opioid new drug
`total_RX_claim_new_quantity_opioid`	total opioid new drug quantity
`fully_paid_claim_paid`	total payment of fully paid claim
`total_RX_claim_new_supply_opioid`	total opioid new drug supply days
`fully_paid_claim_responsible`	total responsible amount of full paid claim
`total_RX_claim_paid_paid_opioid`	total opioid payment
`total_RX_claim_new_MME_opioid`	total opioid new drug MME

fully_paid_claim_charge	total charge of fully paid claim
inbound_call_by_mbr	total times of call by member
surgery	total times of surgery
total_RX_claim_paid_responsible_other	total non-opioid responsible payment
surgery_paid	total payment of surgery
surgery_charge	total charge of surgery
inbound_call_by_other	total call by others
total_RX_claim_new_responsible_other	total non-opioid new drug responsible payment
total_RX_claim_paid_cost_other	total non-opioid cost amount
total_RX_claim_new_responsible_opioid	total opioid new drug responsible amount
total_RX_claim_paid_paid_other	total non-opioid payment amount
total_RX_claim_rejected_responsible	total rejected responsible amount
total_RX_claim_paid_cost_opioid	total opioid cost amount
cpd	total times of cpd
chf	total times of chf
cad	total times of cad
diabetes	total times of diabetes
total_RX_claim_mail_other	total times of non-opioid mail order
hypertension	total times of hypertension
top_five	total times of top_five
diabetes_charge	total charge of diabetes
diabetes_paid	total payment of diabetes

Reference

Ellis, R., J., Wang, Z., Genes, N., & Ma'ayan, A. (2019). Predicting Opioid Dependence from Electronic Health Records with Machine Learning, BioData Mining, 12, Thompkins, D. A., Hobelmann, J. G., & Compton, P. (2017). Providing Chronic Pain Management in the "Fifth Vital Sign" Era: Historical and Treatment Perspectives on a Modern-Day Medical Dilemma. Drug and Alcohol Dependence, 173 Suppl 1(Suppl 1), S11-S22. Doi: 10.1016/j.drgalcdep.2016.12.002.

Guyon, I., Weston, J., & Barnhill, S. (2002). Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 46, 389-422.

Karhade, A., V., Schwab, J., H., & Bedir, H., S. (2019). Development of Machine Learning Algorithms for Prediction of Sustained Postoperative Opioid Prescriptions After Total Hip Arthroplasty. The Journal of Arthroplasty, 34, 2272-2277.

Lo-Ciganic, W., Huang, J., L., & Zhang, H., H. (2019). Evaluation of Machine-Learning Algorithms for Predicting Opioid Overdose Risk Among Medicare Beneficiaries With Opioid Prescriptions. JAMA New Open, 2.

Wader, A. (2019). Predicting Opioid Use Disorder (OUD) Using A Random Forest. COMPSAC, 960-961.

Wu, X., Xie, S., Wang, L., Fan, P., Ge, S., Xie, X., & Wu, W. (2018). A Computational Strategy for Finding Novel Targets and Therapeutic Compounds for Opioid Dependence. PloS One, 13.