Using Machine Learning to Predict the Risk of Opioid Use Disorder (OUD) in Patients with Chronic Opioid Prescriptions

Numaiya Hasan

Advisor: Dr. Nicoleta Serban

Supervisor: Guantao Zhao

Georgia Institute of Technology

ISYE PURA Fall 2024

INTRODUCTION

In the United States, opioids are the mainstay of pain management. Many leading trauma-related treatment plans still center around opioids, as we observe more prolonged opioid exposure within children and adolescents. In recent years, the misuse of prescription opioids has become an increasingly serious problem in the United States. Each year, over 10 million Americans report misusing prescription opioids, with approximately 20% diagnosed with OUD. From 1999 to 2017, the number of opioid overdose deaths increased fivefold, and it is currently estimated that 115 people die from opioid overdoses every day. The misuse or abuse of opioids results in an annual economic loss exceeding $78 billion, including healthcare costs, productivity losses, substance abuse treatment, and criminal justice expenses. Despite ongoing improvements to opioid prescription guidelines and repeated warnings from scholars about the severity of the issue, the problem remains unresolved. In this study we will aim to develop and validate machine learning algorithms to predict the risk of OUD in patients ages 18-65 with chronic opioid prescriptions. In doing so, we hope to predict and assess the risk of OUD and support clinical decision-making to enhance the effectiveness of interventions.

DATA SOURCE

The data is sourced from 2017-2020 Medicaid claims data of *Transformed Medicaid Statistical Information Systems Analytical Files (TAF) Other Services* from the Centers for Medicare and Medicaid Services (CMS), including individual-level claims for Medicaid, Medicaid expansion CHIP, and separate CHIP. Each claim is composed of patient and provider identifiers, demographics, the International Classification of Diseases (ICD-9, ICD-10) codes, National Drug Codes (NDC), and prescription quantities such as days of supply and dosage. According to the Data Use Agreement, all data must meet a minimum cell size of 11 patients.

The data is recorded on the virtual machine and cannot be shared or recorded. To gain access, I had to complete two CITI training courses to become IRB-certified and gain access to the Medicaid dataset.

APPROACH

The project is divided into two main parts: data extraction and model implementation. Unlike conventional machine learning (ML) projects, determining the study population in health analytics is particularly challenging and requires strict criteria. We reviewed various literature to learn how previous scientific studies have defined their study populations. I composed multiple reports summarizing articles based on the objectives of our research. Including the main research topic, any specific algorithms (whether ML or regression models), demographic characteristics, geographic scope (nation-, state-, or county-wide), and the observation window of their study design. In addition, there were inclusion and exclusion criteria that were important to emphasize.

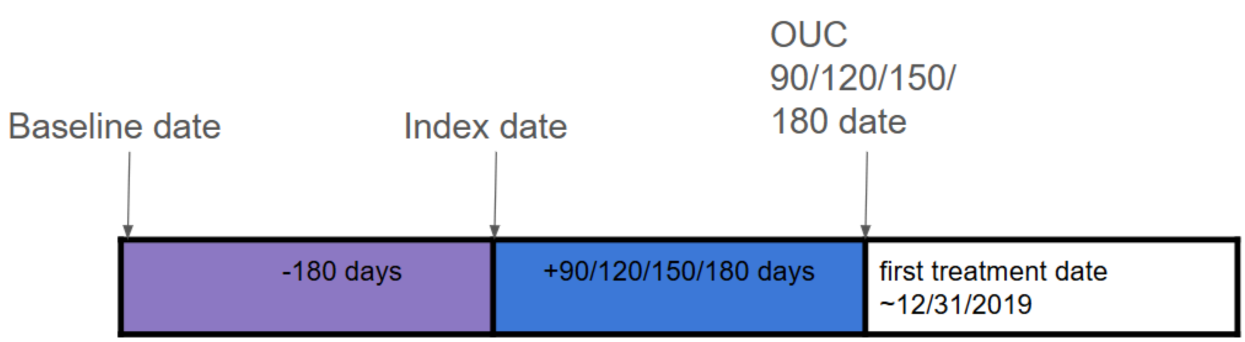
Once we had a collection of references, we extracted 196 NDC codes and 1140 ICD codes that were used to identify OUD patients. We filtered the data for codes that were consistent and overlapping across all articles, the drug codes were narrowed down to prescriptions for buprenorphine, naltrexone, or a combination of both. Then, we searched our SQL database using SQL queries to accurately identify target populations.

1. Find the earliest OUD treatment date for each patient
2. Check for any prior opioid prescription history before the earliest OUD treatment date
3. Calculate aggregate supply days for OUD prescription

4. Sum up the total dosage or average
5. Count the total number of claims of OUD

We need to check for any prior prescriptions of opioids to track the risk of overdose with a lapse in time depending on if they had overdosed before and if we would want to stratify the study groups by risk level.

We will follow the study timeline below, setting the index date to the patient's first opioid prescription after January 1, 2018. The baseline date is the 90-day period before the patient's first opioid prescription and cannot be earlier than October 3, 2017. It is so we can check and exclude patients who had prior OUD diagnoses or treatments. The OUC date is the patient's first treatment or diagnoses for OUD and must occur within 90/120/150/180 supply days after the patient's first opioid prescription and no later than December 31, 2019. We are still developing how to define the first treatment date, but this is the initial formulation.



METHODS

There are over 200 possible features we can analyze for how well they predict OUD, so our next step is to build a comprehensive feature list (see Table 1 in Appendix). Thus far we are considering demographic factors, inpatient and outpatient quantities such as length of stay, RX features for prescription data, and other features including mental health history.

For our machine learning model, we will implement and fine-tune four to five different algorithms, such as regression, boosting, bagging, deep learning, and reinforcement learning-based optimization algorithms. Utilizing different approaches will allow us to identify the optimal solution across various scenarios.

RESEARCH SIGNIFICANCE

While traditional statistical techniques have been applied in assessing opioid-related risks, the complexity and interactions within the data present significant challenges. Machine learning offers unique advantages in addressing this issue. By leveraging machine learning, we can identify hidden patterns in large datasets, uncover new risk factors, and generate more precise and personalized predictions. This will provide clinicians with a powerful tool to more effectively identify patients and design more targeted interventions for these patients, thereby mitigating the problem of opioid misuse to some extent.

Appendix

Table 1: Machine Learning OUD Features Table

1.Index_date	first opioid fill in date	11.# of treatment claim (3 different drugs)	between ouc_date(90/120/150/180) and 2019-12-31
2.Baseline_date	Index_date-180 days	12. # of quantity (3 different drugs)	
3.ouc90_date	Index_date+90days	13.# of supply days(3 different drugs)	treatment start date + supply days(overleap?)
4.ouc120_date	Index_date+120days	14.oud_treatment_index_date	first oud treatment(diagnosis) fill in date
5.ouc150_date	Index_date+150days	15.hybrid treatment indicator	
6.ouc180_date	Index_date+180days	16.oud_treatment_still_opioid	
7.oud_eligible	Indicates patients with OUD **treatment** or **daily MME ≥ 45** between the baseline and index dates.		
8.Longest number of consecutive days (opioid)(90/120/150/180)	between the index_date and ouc_date		
9.Type of opioid	Long/short/comb		
10.Average daily MME(by ouc date)	agg_daily MME/(90/120/150/180)		