Development and External Validation of ML Models for Identifying Patients at Risk of Postoperative Prolonged Opioid Use

A Network Study on OMOP Databases

Stanford MEDICINE | Center for Biomedical Informatics Research

Study Protocol

Document Contributors

Institution/Affiliation	Author	Email
Stanford University	Behzad Naderalvojoud	behzadn@stanford.edu
Stanford University	Catherine Curtin	boussard@stanford.edu
Stanford University	Tina Hernandez-Boussard	ccurtin@stanford.edu

Document History

Date	Version	Editors	Status
08/01/2022	0.1	BN	Table of Content
08/28/2022	0.8	BN	First draft
09/05/2022	0.9	BN, CC, THB	Review
09/10/2022	1.0	BN	Final draft

Table of Contents

List of Figures

List of Tables

Glossary

Acronym	Meaning
PORPOISE	POstopeRative Prolonged OpioId uSE
OMOP	Observational Medical Outcomes Partnership
CDM	Common Data model
NLM	National Library of Medicine
IRB	Institutional Review Board
ML	Machine Learning
PLP	Patient Level Prediction
PPV	Positive Predictive Value
NPV	Negative Predictive Value
ROC	Receiver Operating Characteristics
PRC	Precision Recall Curve
AUC	Area Under Curve
LLR	Lasso Logistic Regression
RF	Random Forest
AB	AdaBoost
GBM	Gradian Boosting Machine
NB	Naive Bayes

1 Project basics

1.1 Sponsor

The project is supported by the National Library of Medicine of the National Institutes of Health under Award Number R01LM013362 and has received approval from the Institutional Review Board (IRB) at Stanford University.

1.2 Description

Opioids are potent analgesics often used to manage pain, including postoperative pain. However, opioids can be highly addictive, even when prescribed correctly and taken as directed. The balance between pain management and opioid misuse is challenging1. To improve patient outcomes following surgery, it is crucial to identify patients at risk for prolonged opioid use prior to prescribing pain management regimes. Many studies have identified a limited number of features predictive of prolonged postoperative opioid use from real-world data, but these are isolated studies using non-standardized data from different sources, which limits their generalizability and reliability across populations2-4. In this study, we developed and validated five machine learning (ML) models to predict patients at risk of prolonged opioid use in a diverse, multisite cohort by evaluating not only the performance, but also their discrimination and calibration abilities. We address generalizability limitations by using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to identify postoperative patients at higher risk of prolonged opioid use based on preoperative risk factors using ML approaches.

1.3 Goals

The project main goals are summarized as follows:

- Improve pain management following surgery.
- Identify patients at risk for prolonged opioid use prior to prescribing pain management regimes.
- Develop and validate ML models in a diverse, multisite cohort by evaluating their generalizability, discrimination, and calibration abilities.
- Evaluate the generalizability and calibration of the ML models across multisite cohort subgroups: diabetes, depression, obesity.
- Evaluate the transportability of ML models based on population differences in the various CDM databases.
- Identify common preoperative risk factors predictive of postoperative opioid use over CDM databases.
- Incorporate ML models trained on different databases to increase generalizability.
- Incorporating developed ML models into an open-access web application so that researchers can evaluate the performance of models and compare them to their own models in a transparent setting.

2 Rationale and Background

Opioids are potent analgesics often used to manage pain, including postoperative pain. However, opioids can be highly addictive, even when prescribed correctly and taken as directed. Over the past 15 years, the number of opioid-related drug overdose deaths has tripled, with prescription

opioids accounting for more than half of all overdose deaths (Dowell et al, 2016). On the other hand, opioid use has been linked to serious complications such as dependence and abuse, which have resulted in significant morbidity and mortality (Lyden & Binswanger, 2019). Although postoperative opioid exposure is a major risk factor for prolonged use and abuse, prescription opioid medications continue to play important roles in the management of postoperative pain following surgery (Brummett et al, 2017). To reduce the morbidity associated with opioid use after surgery, it is crucial to identify patients at risk for prolonged opioid use prior to prescribing opioids. Several studies have used machine learning (ML) to predict patients at risk and to identify risk factors associated with this complication for specific surgeries, such as arthroscopic meniscectomy and spine surgery (Karhade et al, 2020, Lu et al, 2022). In this study, we will develop and validate ML models to predict patients at risk of prolonged opioid use for most surgeries. Brummett et al. found no difference in new persistent opioid use between patients who underwent minor and major surgical procedures, indicating that prolonged opioid use is not solely caused by surgical pain. Therefore, in developing ML models, we will consider patients' 6-month prior medications, diagnoses, lab measurements, and procedures.

Moreover, many patient-specific clinical parameters complicate the prediction of prolonged opioid use. Many studies have identified a limited number of features predictive of prolonged postoperative opioid use from real-world data, but these are isolated studies using non-standardized data from different sources, which limits their generalizability and reliability across population (Katakam et al, 2020; Dong et al, 2021; Ward et al, 2021). This study aims to address these limitations, by using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to identify postoperative patients at higher risk of prolonged opioid use based on preoperative risk factors using ML approaches. The OMOP CDM, developed by the Observational Health Data Sciences and Informatics (OHDSI) community, standardizes the format of observational healthcare data, allowing the code to be directly shared with other researchers working with different datasets. We will also investigate the impact of CDM features on the performance of several ML models using various feature evaluation metrics and provided transparent validation on model discrimination, calibration, and clinical utility. Unlike previous studies, which only used a limited number of features determined in the literature, using CDM features allows us to consider a wide range of covariates across domains, such as demographic, drug, condition, procedure, and measurement, in the development of ML models without the need for manual feature engineering.

3 Materials and methods

3.1 Data source

The study will rely on multisite observational data from electronic health records (EHRs) mapped to the OMOP CDM. All the data will be analyzed in a federated manner, where the data will remain with the data owners and only the analysis results will be shared and published.

3.2 Study design

The PORPOISE study is a retrospective analysis of observational health data that has received approval from the Institutional Review Board (IRB) at Stanford University. The analysis and prediction models were implemented using the OHDSI PatientLevelPrediction (PLP) package,[1] and all study materials are available in the project Git repository to be run on any OHDSI network

[1] https://github.com/OHDSI/PatientLevelPrediction

databases[2]. In the study, all observations about opioid medications are looked at in two different time periods: the preoperative period, which is a six-month window before surgery, and a six-month follow-up period, including the second 90 days of the six-month follow-up period. On the basis of these two time periods, the following analyses are performed:

- The target and outcome cohorts, as well as evaluation subgroups, are defined.
- The datasets included in the study are characterized based on the predefined features across three subgroups: Diabetes, Depression, and Obesity to determine the distribution differences of study datasets.
- Five ML algorithms are trained and validated internally over the partners local datasets.
- The trained models are shared across partners for external validation.
- All internal and external validation results are discussed and published.

3.2.1 Target cohort

The target cohort includes adult patients who underwent surgery during an inpatient visit between 2008 and 2019 with at least one opioid prescription 30 days before or after the surgery. Patients were included if they had at least two visits two years before surgery and two visits 30 days to two years after surgery. If a patient had multiple surgeries, only the first event was included. We also excluded patients who had any other surgery from two months to seven months after the index surgery. Figure 1 outlines the inclusion criteria for the target cohort.

In the target cohort, nine groups of RxNorm opioid drug ingredients, shown in Table 1, are used to determine opioid prescriptions, and ICD/CPT codes associated with the following seventeen key groups of surgeries are considered to identify target procedures. As a result of ICD/CPT code mapping, 5,182 CDM concepts were employed to identify target procedures.

1) Laminectomy, excision intervertebral disc
2) Spinal fusion
3) Cholecystectomy and common duct exploration
4) Partial excision bone
5) Hysterectomy, abdominal and vaginal
6) Colorectal resection
7) Excision, lysis peritoneal adhesions
8) Appendectomy
9) Treatment, fracture or dislocation of hip and femur
10) Oophorectomy, unilateral and bilateral
11) Coronary artery bypass graft (CABG)
12) Inguinal hernia repair
13) Distal radial Fracture
14) Thoracotomy
15) Mastectomy
16) Knee Replacement
17) Treatment, fracture or dislocation of lower extremity (other than hip or femur)

[2] https://github.com/ohdsi-studies/PORPOISE

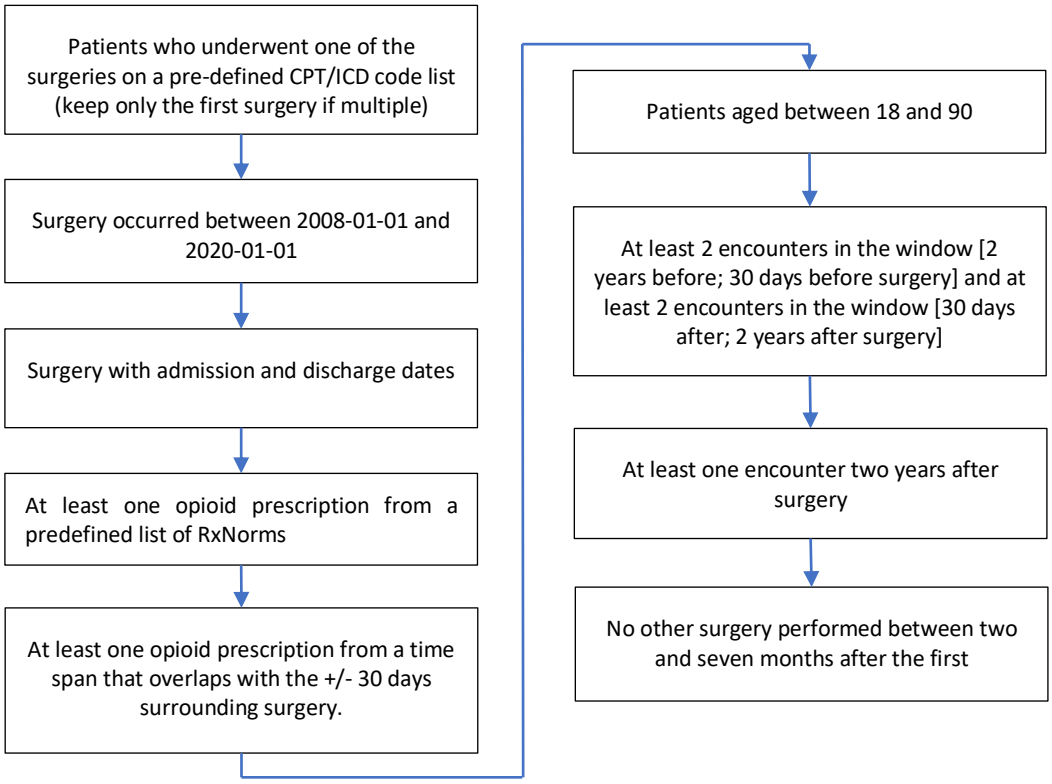

Figure 1. Target cohort inclusion criteria.

Table 1. RxNorms used to identify opioid prescriptions.

RxNorm	Drug ingredient
5489	HYDROcodone
4337	fentaNYL
2670	Codeine
3423	HYDROmorphone
6754	Meperidine
6813	Methadone
7052	Morphine
7804	oxyCODONE
10689	Tramadol

To characterize all the study datasets and evaluate the performance of predictions models three subgroups, including Diabetes, Depression, and Obesity, are defined based on the target cohort.

3.2.1.1 Diabetes subgroup

Occurring any of the following criteria is considered to determine diabetes subgroup:

1) At least one diagnosis of Type 1 or 2 diabetes mellitus all days before and 180 days after the index surgery, excluding diabetes mellitus during pregnancy.
2) At least one drug exposure, BLOOD GLUCOSE LOWERING DRUGS, EXCL. INSULINS, all days before and 180 days after the index surgery.
3) At least a measurement of Hemoglobin A1C with value between 6.5 and 20 percent.

3.2.1.2 Depression subgroup

To determine the depression subgroup, the following criterion is considered:

1) At least one occurrence of depressive disorder diagnoses 365 days before and 180 days after the index surgery, excluding postpartum depression.

3.2.1.3 Obesity subgroup

To determine the obesity subgroup, the following criterion is considered:

1) At least one occurrence of obesity diagnoses 180 days before and 180 days after the index surgery, excluding particular diagnoses such as Maternal Obesity Syndrome, Intellectual Disability, Seizures, Macrocephaly, Obesity Syndrome, Choroideremia With Deafness And Obesity Syndrome.
2) At least one occurrence of obesity observation 180 days before and 180 days after the index surgery, such as:
    a. Body mass index 30+ - obesity
    b. Body mass index 40+ - severely obese
    c. Obese class I (body mass index 30.0 - 34.9)
    d. Obese class II (body mass index 35.0 - 39.9)

3.2.2 Outcome cohort

The primary outcome is prolonged opioid use, defined as a new opioid prescription within three to six months after surgery. Based on this definition, two cohorts, namely prolonged and non-prolonged opioid users, are extracted from the target cohort to be used in characterization.

3.2.3 Cohort characterization

The four types of feature analysis, shown in Table 2, are used in the cohort characterization. The objective is to observe the distribution of features between prolonged and non-prolonged opioid users and understand the distribution differences between study datasets. Furthermore, three evaluation subgroups are used in the characterization to determine the distribution of these subgroups over individual features.

Table 2. Features analyses used in characterization.

Feature Type	Feature name
Demographics	Age
	Age group
	Ethnicity
	Race
	Gender
Distinct count	Distinct condition counts medium term
	Distinct procedure counts medium term
	Distinct measurement counts medium term
	Distinct ingredients count medium term
Clinical binary	Condition occurrence medium term
	Procedure occurrence medium term
	Measurement medium term
	Drug exposure medium term
Subgroup count	Diabetes count
	Depression count
	Obesity count

3.2.4 Patient Level Prediction

Figure 2 depicts the prediction problem based on a 180-day observation window and 90-day time-at-risk, as well as the study's outcome. As a result, the prediction problem is defined as estimating the risk of any opioid drug exposure at the time of risk based on patients' clinical values six months prior to surgery.

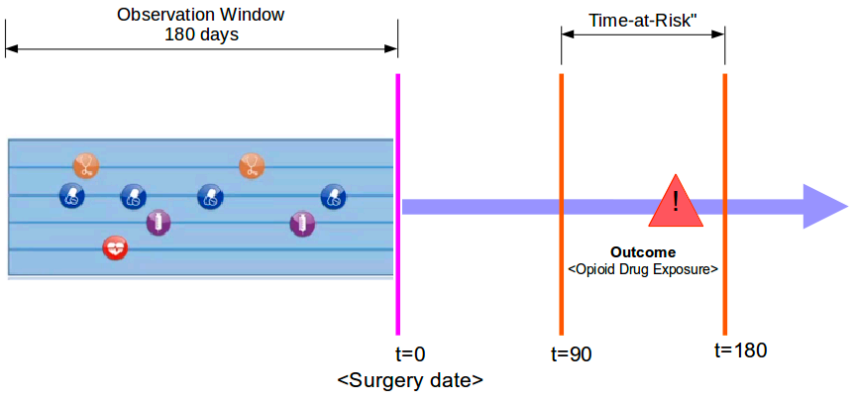

Figure 2. Prediction definition.

3.2.4.1 Covariate setting

All prediction features are defined based on CDM concepts shown in Table 3, and all infrequent covariates are removed with less than 0.001 frequency from the feature set.

Table 3. Prediction features.

Feature Type	Feature name
Demographics	Gender
	Age group
	Ethnicity
	Race
Distinct count	Distinct condition count
	Distinct procedure count
	Distinct measurement count
	Distinct ingredient count
Clinical binary	Condition occurrence
	Procedure occurrence
	Measurement
	Drug exposure
Group-based	Condition group
	Drug group

3.2.4.2 Machine leaning models

Five ML algorithms are used to develop our models: Lasso Logistic Regression (LR), Random Forest (RF), AdaBoost (AB), Gradian Boosting Machine (GBM), and Naive Bayes (NB). The models are developed using the PatientLevelPrediction (PLP) package from the OHDSI community, version 5.4.5. A stratified random sampling approach is used to split the dataset into train (80%) and test (20%) sets. To label patients in the target cohort, a 90-day time-at-risk (TAR) is defined between three and six months after surgery. For the ML models, any opioid drug exposure during this period is considered a positive case.

3.2.4.3 Internal evaluation

To evaluate models, standard metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are used. We also assess model discrimination using receiver operating characteristics (ROC) and precision recall curve (PRC) with their area under curve (AUC) values. The calibration curves are developed to evaluate the model outputs in terms of their ability to generate calibrated probabilities for identifying patients at risk of prolonged opioid use. To select the best hyperparameters and prediction threshold, 5-fold cross validation on the train set is used with a grid search strategy.

3.2.4.4 External validation

After training and evaluating the models on various CDM databases, the PLP results (the output of the prediction module) are shared with study partners to be validated externally on CDM databases where the models have not been trained. To that end, all models are validated on the entire target cohort as well as the three subgroup cohorts on the target CDM data base. All metrics used in the internal validation are adopted to evaluate external results. The distribution differences between the trained and validated databases will be used to discuss all external evaluations.

4 Dissemination plan

The study's findings will be presented as abstracts at Stanford AI + Health (2022) conference. The final results will be published in international peer-reviewed journals as full-text papers.

References

1. Dowell, D., Haegerich, T. M., & Chou, R. (2016). CDC guideline for prescribing opioids for chronic pain—United States, 2016. Jama, 315(15), 1624-1645.
2. Lyden, J., & Binswanger, I. A. (2019, April). The United States opioid epidemic. In Seminars in perinatology (Vol. 43, No. 3, pp. 123-131). WB Saunders.
3. Brummett, C. M., Waljee, J. F., Goesling, J., Moser, S., Lin, P., Englesbe, M. J., ... & Nallamothu, B. K. (2017). New persistent opioid use after minor and major surgical procedures in US adults. JAMA surgery, 152(6), e170504-e170504.
4. Karhade, A. V., Cha, T. D., Fogel, H. A., Hershman, S. H., Tobert, D. G., Schoenfeld, A. J., ... & Schwab, J. H. (2020). Predicting prolonged opioid prescriptions in opioid-naïve lumbar spine surgery patients. The Spine Journal, 20(6), 888-895.
5. Lu, Y., Forlenza, E., Wilbur, R. R., Lavoie-Gagne, O., Fu, M. C., Yanke, A. B., ... & Forsythe, B. (2022). Machine-learning model successfully predicts patients at risk for prolonged postoperative opioid use following elective knee arthroscopy. Knee Surgery, Sports Traumatology, Arthroscopy, 30(3), 762-772.
6. Dong, X., Deng, J., Rashidian, S., Abell-Hart, K., Hou, W., Rosenthal, R. N., ... & Wang, F. (2021). Identifying risk of opioid use disorder for patients taking opioid medications with deep learning. Journal of the American Medical Informatics Association, 28(8), 1683-1693.
7. Katakam, A., Karhade, A. V., Schwab, J. H., Chen, A. F., & Bedair, H. S. (2020). Development and validation of machine learning algorithms for postoperative opioid prescriptions after TKA. Journal of orthopaedics, 22, 95-99.
8. Ward, A., Jani, T., De Souza, E., Scheinker, D., Bambos, N., & Anderson, T. A. (2021). Prediction of prolonged opioid use after surgery in adolescents: insights from machine learning. Anesthesia & Analgesia, 133(2), 304-313.
9. Weng, Y., Tian, L., Tedesco, D., Desai, K., Asch, S. M., Carroll, I., ... & Hernandez-Boussard, T. (2020). Trajectory analysis for postoperative pain using electronic health records: A nonparametric method with robust linear regression and K-medians cluster analysis. Health informatics journal, 26(2), 1404-1418.
10. Hah, J., & Hernandez-Boussard, T. (2018). Defining Postoperative Opioid Needs Among Preoperative Opioid Users. JAMA surgery, 153(7), 689-690.
11. Desai, K., Carroll, I., Asch, S. M., Seto, T., McDonald, K. M., Curtin, C., & Hernandez-Boussard, T. (2018). Utilization and effectiveness of multimodal discharge analgesia for postoperative pain management. Journal of Surgical Research, 228, 160-169.
12. Tedesco, D., Asch, S. M., Curtin, C., Hah, J., McDonald, K. M., Fantini, M. P., & Hernandez-Boussard, T. (2017). Opioid abuse and poisoning: trends in inpatient and emergency department discharges. Health Affairs, 36(10), 1748-1753.
13. Yim, W. W., Tedesco, D., Curtin, C., & Hernandez-Boussard, T. (2017, August). Annotation of pain and anesthesia events for surgery-related processes and outcomes extraction. In BioNLP 2017 (pp. 200-205).
14. Tedesco, D., Gori, D., Desai, K. R., Asch, S., Carroll, I. R., Curtin, C., ... & Hernandez-Boussard, T. (2017). Drug-free interventions to reduce pain or opioid consumption after total knee arthroplasty: a systematic review and meta-analysis. JAMA surgery, 152(10), e172872-e172872.
15. Chin, K. K., Desai, K., Asch, S. M., Carroll, I., Curtin, C., McDonald, K., & Hernandez-Boussard, T. (2017, June). Rapid Practice Changes in Response to the Opioid Epidemic: Integrating Adjuvant Analgesics into Practice. In 2017 Annual Research Meeting. AcademyHealth.