

Original Paper

# Developing and Validating a Machine Learning Algorithm to Predict the Risk of Incident Opioid Use Disorder Among OneFlorida+ Patients: Prognostic Modeling Study

Jabed Al Faysal<sup>1,2</sup>, BSc, MSc; Weihsuan Lo-Ciganic<sup>3,4,5</sup>, PhD; Walid F Gellad<sup>3,4</sup>, MD; Yonghui Wu<sup>6</sup>, PhD; Christopher A Harle<sup>7,8</sup>, PhD; Khoa Nguyen<sup>9</sup>, PharmD; James L Huang<sup>1</sup>, PhD; Gerald Cochran<sup>10</sup>, PhD; Debbie L Wilson<sup>1</sup>, PhD; Stephanie AS Staras<sup>6</sup>, PhD; Siegfried OF Schmidt<sup>11</sup>, MD, PhD; Eric I Rosenberg<sup>1,12</sup>, MD, MSPH; Danielle Nelson<sup>11</sup>, MD; Shunhua Yan<sup>1</sup>, MED; Gary M Reisfield<sup>1</sup>, MD; William M Greene<sup>13</sup>, MD; Courtney Kuza<sup>4</sup>, PhD; Md Mahmudul Hasan<sup>1,14</sup>, PhD

<sup>1</sup>Department of Pharmaceutical Outcomes & Policy, University of Florida, Gainesville, FL, United States

<sup>2</sup>Computer Science and Engineering Discipline, Khulna University, Khulna, Bangladesh

<sup>3</sup>Division of General Internal Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

<sup>4</sup>Center for Pharmaceutical Policy and Prescribing, University of Pittsburgh, Pittsburgh, PA, United States

<sup>5</sup>Geriatric Research Education and Clinical Center, North Florida/South Georgia Veterans Health System, Gainesville, FL, United States

<sup>6</sup>Department of Health Outcomes & Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, United States

<sup>7</sup>Department of Health Policy and Management, School of Public Health, Indiana University, Indianapolis, IN, United States

<sup>8</sup>Regenstrief Institute, Indianapolis, IN, United States

<sup>9</sup>Department of Pharmacotherapy and Translational Research, College of Pharmacy, University of Florida, Gainesville, FL, United States

<sup>10</sup>Department of Internal Medicine, Division of Epidemiology, University of Utah, Salt Lake City, UT, United States

<sup>11</sup>Department of Community Health and Family Medicine, College of Medicine, University of Florida, Gainesville, FL, United States

<sup>12</sup>Department of Internal Medicine, College of Medicine, University of Florida, Gainesville, FL, United States

<sup>13</sup>Department of Psychiatry, College of Medicine, University of Florida, Gainesville, FL, United States

<sup>14</sup>Department of Information Systems and Operations Management, Warrington College of Business, University of Florida, Gainesville, FL, United States

## Corresponding Author:

Md Mahmudul Hasan, PhD

Department of Pharmaceutical Outcomes & Policy

University of Florida

1889 Museum Road, Malachowsky Hall, Suite 6300

Gainesville, FL 32611

United States

Phone: 12566946603

Email: [hasan.mdmahmudul@ufl.edu](mailto:hasan.mdmahmudul@ufl.edu)

## Abstract

**Background:** Opioid use disorder (OUD) remains a critical public health crisis in the United States. Despite widespread policy and clinical interventions, early identification of individuals at risk for developing OUD remains challenging due to limitations in traditional screening approaches and a lack of individualized risk stratification methods. Machine learning (ML) methods offer an opportunity to develop timely, high-performing, and explainable predictive models that can enhance OUD prevention strategies in clinical settings.

**Objective:** This study aims to develop and validate an ML model using electronic health record (EHR) data to predict the 3-month risk of incident OUD among adults initiating opioid therapy and to stratify patients into clinically actionable risk groups.

**Methods:** This prognostic modeling study used 2017-2022 OneFlorida+ EHR data to develop and validate ML algorithms predicting 3-month incident OUD risk. We included 182,083 adults ( $\geq 18$  y) without cancer, overdose, or OUD or hospice history who received  $\geq 1$  outpatient, noninjectable opioid prescription. Using 183 predictors measured in sequential 3-month intervals, we developed an elastic net, least absolute shrinkage and selection operator, gradient boosting machine (GBM), and random forest models on randomly split training, testing, and validation sets. Model performance was assessed using

C-statistics, predictive values, and number needed to evaluate, with patients stratified into risk deciles for clinical applicability. Model explainability was assessed using Shapley additive explanations, and fairness was evaluated using standard metrics. We externally validated the best-performing model using an independent cohort from the 2018-2020 UPMC (formerly University of Pittsburgh Medical Center) health system.

**Results:** In the validation sample (n=60,694), GBM (C-statistics=0.879, 95% CI 0.874-0.884) and elastic net (C-statistics=0.872, 95% CI 0.867-0.877) outperformed least absolute shrinkage and selection operator (C-statistics=0.846, 95% CI 0.840-0.851) and random forest (C-statistics=0.798, 95% CI 0.792-0.804), with GBM model requiring the fewest predictors (n=75) for predicting 3-month incident OUD. Using the GBM algorithm to predict the subsequent 3-month OUD risk, the top decile subgroup had a positive predictive value of 3.26%, a negative predictive value of 99.8%, and a number needed to evaluate of 31. The top decile (n=6696) captured ~68% of patients with OUD. Shapley additive explanations analysis identified age, number of outpatient visits, history of back and other pain conditions, comorbidity burden, and opioid prescribing patterns as the strongest predictors of incident OUD. Fairness assessment showed an acceptable false negative rate parity across race, age, and sex. In external validation on the UPMC cohort, the GBM model maintained good discrimination (C-statistics=0.756, 95% CI 0.750-0.762) and effective risk stratification.

**Conclusions:** An ML algorithm predicting incident OUD derived from OneFlorida+ EHR data performed well in external validation with data using UPMC. The algorithm might be valuable for incident OUD risk prediction and stratification across health systems, with potential to inform early intervention.

*J Med Internet Res* 2026;28:e79482; doi: [10.2196/79482](https://doi.org/10.2196/79482)

**Keywords:** opioid use disorder; machine learning; OneFlorida+; risk stratification; external validation

## Introduction

The United States continues to face a persistent and evolving opioid epidemic. Opioid use disorder (OUD) and overdose affect millions of Americans, leading to increased morbidity, mortality, and health care costs [1,2]. In 2022, more than 6 million individuals experienced OUD, imposing significant societal and economic burdens exceeding US \$78 billion annually [1,3,4]. Opioid-related deaths increased tenfold from 1999 to 2022 (>82,000 in 2022) [5-8]. Prescription opioids were responsible for approximately 280,000 overdose deaths over the same period [9]. The prevalence of OUD continues to rise despite efforts to curb opioid misuse [10,11]. In response, health systems, payers, and policymakers have implemented various interventions aimed at reducing unsafe prescribing and patient risk [12].

Early identification of OUD can prevent severe addiction, high-risk behaviors, overdose, and death, while enabling timely access to evidence-based treatments such as buprenorphine or referral to recovery services. Existing methods for identifying high-risk opioid users, such as high-dose opioid prescribing thresholds and use of multiple pharmacies, are often based on simple or single criteria [13,14]. These rule-based strategies leave many high-risk individuals undetected and may misclassify others, resulting in unintended consequences like delayed care, unnecessary monitoring, or stigma. For example, a study found that nearly 43% of individuals who developed OUD were missed by such approaches [15]. Furthermore, traditional statistical models emphasize population-level risk factors and assume linear relationships between predictors and outcomes, limiting their ability to capture complex interactions among patient demographics, comorbidities, health care use, and prescriber patterns. Although tools such as the Screener and Opioid Assessment for Patients with Pain-Revised and the Opioid Risk Tool are commonly used in clinical settings, studies

report inconsistent estimates of test accuracy and no evidence that they reduce OUD or overdose [16-18]. The 2022 Centers for Disease Control and Prevention Clinical Practice Guideline similarly concluded that current opioid risk screening tools lack proven clinical use. Proprietary platforms such as NarxCare leverage prescription and health record data but lack transparency, public validation, and demonstrated generalizability [19]. These limitations underscore the need for more robust, transparent, and generalizable models to support real-time clinical decision-making across diverse populations. Studies have also highlighted these shortcomings and called for more advanced, data-driven models to improve identification of individuals at risk (or low risk) of OUD [20-24].

Machine learning (ML) offers a promising alternative to traditional statistical methods in improving risk prediction. ML models can analyze large health care data to identify complex nonlinear patterns in patient data. Recent studies have demonstrated the potential of ML to predict opioid-related outcomes, including OUD and overdose, with higher accuracy than that of traditional methods [25-28]. Our previous work showed that ML approaches can improve risk prediction and stratification for incident OUD and subsequent overdose in Medicare and Medicaid beneficiaries. For example, we developed ML models to predict incident OUD among Medicare beneficiaries, achieving high discrimination (C-statistic >0.86) and effectively stratifying patients into clinically meaningful risk subgroups (eg, low, moderate, and high risk for developing OUD) [29]. However, many existing models are developed using Medicare or Medicaid data, which primarily include older adults or low-income populations.

In this study, we extend our work to develop and validate an ML algorithm to predict incident OUD among patients who received care from institutions within the OneFlorida+ network—a large, diverse patient population across Florida,

Georgia, Alabama, and Arkansas, including individuals with commercial insurance, Medicaid, or Medicare, and those with no insurance [30]. This study advances our prior Medicare-based work in several important ways. First, although claims-based models can be incorporated into some clinical workflows, electronic health records (EHRs) often lack the full use and medication detail available in claims and may vary in data quality. As such, adapting the model for EHR-based use required methodological refinement to support real-time, point-of-care integration. Second, the broader and more heterogeneous cohort enhances the model's generalizability beyond the older, fee-for-service Medicare population used previously. Third, we externally validated the model using an independent cohort from the UPMC health system, assessing cross-system generalizability and transportability—an important step toward real-world implementation. Fourth, we incorporated fairness and bias analyses across demographic subgroups to address equity considerations often overlooked in prior OUD risk prediction studies. Finally, we articulate a translational vision for integrating this model into clinical workflows. By organizing predictors to enable risk-based stratification, the model is designed to function as an EHR-embedded decision support tool in primary care settings, alerting providers (eg, physicians) when a patient is flagged as high-risk to prompt timely, preventive actions. We emphasize model interpretability to ensure that our findings can be seamlessly integrated into the clinical environment.

## Methods

### Study Design and Data Source

This is a prognostic modeling study with a retrospective cohort design. We used the TRIPOD+AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis + Artificial Intelligence) guidelines for reporting our work (Checklist 1) [31]. This study used data from OneFlorida+, a secure and centralized repository of patient-level data from both health system partners and insurers, and managed by the OneFlorida Clinical Research Consortium [32]. Currently, OneFlorida+ includes EHRs and claims data from 26 million individuals across Florida, Georgia, Alabama, and Arkansas, covering a wide range of populations and health care settings [30,33]. All data sources are mapped to the Patient-Centered Outcomes Research network common data model (version 7.0) to ensure standardization of data elements across sources. Major data elements in the common data model include demographics, enrollment, encounters, diagnoses, procedures, dispensed medications, and deaths. Additionally, while OneFlorida data includes Florida Medicaid recipients with claims data only, our study cohort requires having at least 1 EHR encounter. The study population included adult patients (aged  $\geq 18$  y) who had at least 1 opioid prescription filled between 2017 and 2022 (Figure S1 in Multimedia Appendix 1). The index date was defined as the date of a patient's first opioid prescription between October 1, 2017, and September 30, 2022. To construct a cohort appropriate for assessing incident OUD risk, we excluded patients who (1) had cancer diagnoses

(Table S1 in Multimedia Appendix 2), (2) received hospice care, (3) had a diagnosis of OUD, opioid overdose, other substance use disorders, drug abuse, or received methadone or buprenorphine for OUD before initiating opioids, or (4) did not have at least a 3-month observation window following the first opioid prescription to allow for predictor measurement. We also excluded patients with a diagnosis of other substance use disorders to minimize confounding, as some physicians may have used this diagnosis for patients with co-occurring OUD and other substance use conditions. Once eligible, patients remained in the cohort until they experienced an outcome of interest (ie, incident of OUD diagnosis) or were censored due to death, regardless of continued opioid use (Figure S2 in Multimedia Appendix 1).

For external validation, we applied the algorithm to 2018-2020 EHR data from the UPMC health system in Pennsylvania, a region with demographic and opioid prescribing characteristics distinct from the OneFlorida+ network. Cohort construction and predictor generation followed the same procedures used for the OneFlorida+ dataset.

### Outcome Variable

Similar to other claims-based analyses [29,34-36], our primary outcome was incident OUD (Table S2 in Multimedia Appendix 2), defined as the first recorded diagnosis of OUD from all settings in each 3-month window after the index prescription date (Figure S2 in Multimedia Appendix 1). The *ICD-10 (International Statistical Classification of Diseases and Related Health Problems 10th Revision)* codes for incident OUD diagnosis include F11.1X and F11.2X but exclude F11.11 (opioid-related disorders in remission) and F11.21 (opioid dependence in remission).

### Predictor Candidates

We identified 183 candidate predictors of OUD from the prior literature (Table S3 in Multimedia Appendix 2) [37-48]. These predictors included sociodemographic factors, patient health status, and patterns of opioid use and other nonopioid prescriptions measured at baseline (during the 3 month period before the first opioid fill) and in 3-month windows after initiating prescription opioids. This 3-month window was selected for both clinical and operational relevance. Evidence shows that early indicators of problematic opioid use (eg, dose escalation, early refills, or polypharmacy) typically emerge within the first 3 months of therapy [49-52]. Pharmacoepidemiologic studies demonstrate that transitions to sustained or high-risk opioid use most commonly occur within 3 months after treatment initiation [53, 54]. This timeframe also aligns with the quarterly monitoring cycles used by many prescription drug monitoring programs and health plans [14,38]. We updated the predictors measured in each 3-month period to account for changes over time for predicting incident OUD risks in each subsequent period (Figure S2 in Multimedia Appendix 1). This time-updating approach mimics active surveillance that a health system might conduct in real time [55] to provide clinicians with timely, actionable opportunities (eg, closer follow-up, patient counseling, or medication adjustments) before the

progression to OUD. We examined missingness across candidate predictors and applied prespecified imputation rules before model training. We used simple imputation techniques where missing values for all categorical predictors (including race, ethnicity, and provider sex) were imputed using the modal category. Continuous predictors (eg, use counts, comorbidity scores, and medication counts) with missing values were imputed using the median calculated from the full analytic dataset. Binary and count predictors representing the presence or frequency of diagnoses, procedures, or medications were imputed as zero when missing, consistent with the interpretation that the absence of a code reflects no recorded event during the baseline window.

## **ML Approaches and Prediction Performance Evaluation**

Our machine learning analysis using OneFlorida+ data had two objectives: (1) developing a prediction model to generate individuals' incident OUD risk scores and (2) stratifying individuals into subgroups based on similar OUD risk levels. We developed and tested 4 ML algorithms to predict incident OUD: elastic net (EN), least absolute shrinkage and selection operator (LASSO), gradient boosting machine (GBM), random forest (RF). These algorithms have consistently demonstrated strong predictive performance in prior studies [12,46,56,57]. Details of each algorithm are provided in [Multimedia Appendix 3](#). We randomly divided the cohort into 3 subsets: a training sample for algorithm development, a testing sample for algorithm refinement, and a validation sample for evaluating predictive performance. This split was conducted strictly by patient ID, such that all 3-month episodes from a given patient were assigned to a single dataset (training, testing, or validation), and no patient contributed episodes to more than one set, thereby avoiding information leakage across datasets. Because patients could contribute multiple 3-month episodes, episode-level observations within patients were correlated by design. Models were trained and evaluated at the episode level to reflect the intended real-world use case of repeated, longitudinal risk assessment over time and to maximize learning from rare outcomes. Tree-based ensemble methods such as GBMs are generally robust to correlated observations because they do not rely on independence assumptions and learn predictive structure through recursive partitioning rather than parametric estimation. To evaluate the potential impact of within-patient correlation on model performance estimates, we conducted sensitivity analyses using patient-level random subsets. We used one 3-month period with predictor measurements to forecast risk in the subsequent 3 months within the validation set.

To evaluate discrimination performance, we assessed whether individuals predicted to be at high risk had higher OUD incidence than those predicted to be at low risk. We compared C-statistics across different methods in the validation sample using the DeLong test [58]. A C-statistic between 0.7 and 0.8 indicates good discrimination, while values above 0.8 indicate very good discrimination. Precision-recall curves were also examined [42]. Given that OUD

events are rare outcomes and C-statistics do not incorporate information about outcome incidence, we reported other metrics, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), number needed to evaluate (NNE) to identify 1 OUD episode, and estimated rate of alerts, to thoroughly assess the algorithms' predictive ability (Figure S3 in [Multimedia Appendix 1](#)) [43, 59].

There is no universally applicable prediction probability threshold, so we evaluated performance across multiple sensitivity and specificity levels (eg, selecting 90% sensitivity). To determine an optimized threshold that balances sensitivity and specificity, we applied the Youden index in the training sample [60]. In the validation sample, patients were stratified into risk subgroups based on deciles of predicted incident OUD probability. The highest decile was further divided into 3 strata: the top first percentile, the second to fifth percentiles, and the sixth to tenth percentiles. This approach allowed for a more detailed assessment of individuals at the highest risk of developing OUD.

We conducted additional analyses to enhance clinical use. The primary goal of our ML algorithm was to generate risk scores for incident OUD. For comprehensive model explanation and further enhancement of model interpretability, we conducted a Shapley additive explanations (SHAP) analysis [61]. SHAP quantifies the contribution of each predictor to the model's output while accounting for interactions with other features. The SHAP plots indicate whether the mentioned features have a positive correlation (red color) or a negative correlation (blue color) with OUD. To evaluate bias across demographic factors (eg, race, age, and sex), we examined risk score distributions and false positive rates (FPR) and false negative rates (FNR) in the best performing model. We defined significant disparity as a parity ratio between subgroups falling outside the range of 0.80 to 1.25, consistent with the standards used in algorithmic fairness audits [62].

Additionally, we evaluated model calibration to assess the reliability of predicted risk scores. Given the potential for miscalibration in rare-outcome ML models, we applied isotonic regression to recalibrate predicted risk scores [63]. The overall accuracy of probabilistic predictions was assessed using the Brier score metric [64]. To determine clinical use beyond discrimination, we performed decision curve analysis. We calculated the net benefit [65] across a range of clinically relevant risk thresholds, comparing the recalibrated model against default strategies of "treat all" and "treat none". This approach quantifies the trade-off between the benefit of identifying true OUD cases and the potential harm of false-positive alerts.

## **Statistical Analysis**

All analyses were performed using SAS 9.4 (SAS Institute Inc) and Python 3.6 (Python Software Foundation). Continuous variables were summarized using mean (SD) values, while categorical variables were summarized using frequencies and percentages. We compared patients' characteristics in training, testing, and validation datasets using 2-tailed

Student *t* tests, chi-square tests, and ANOVA, or corresponding nonparametric tests.

## Ethical Considerations

This study was reviewed and approved by the University of Florida (UF) Institutional Review Board (IRB202101897). The research involved secondary analysis of limited electronic health record data from the OneFlorida+ Data Trust. In accordance with federal regulations (45 CFR 46.104[d] [4]) and institutional policies, the requirement for informed consent was waived because the study involved no direct participant contact and posed minimal risk to individuals. All analyses were conducted within the secure UF server environment; no identifiable private information was accessed or transmitted outside the system. The study adhered to institutional data-use agreements and complied with all relevant guidelines for human subjects research and data privacy. The external validation using UPMC data was reviewed by the University of Pittsburgh Human Research Protection Office and determined not to constitute human subjects research, as only deidentified data were used.

Therefore, institutional review board review and informed consent were not required.

## Results

### Patient Characteristics

OneFlorida+ patients in the training ( $n=60,694$ ), testing ( $n=60,695$ ), and validation ( $n=60,694$ ) samples had similar characteristics and outcome distributions (mean age 53.02, SD 7.60 y; 111,576/182,083, 61.28% female; 125,208/182,083, 68.76% White; 144,316/182,083, 79.26% non-Hispanic; 4328/182,083, 2.38% had incident OUD; [Table 1](#)). The external validation cohort from the UPMC included 129,215 patients (mean age 58.91, SD 16.75 y; 78,574/129,215, 60.81% female; 115,907/129,215, 89.7% White; 123,083/129,215, 95.25% non-Hispanic; 2537/129,215, 1.96% had incident OUD). While sex distributions were similar between the 2 health systems, the UPMC cohort was older and less diverse, with more White and non-Hispanic patients.

**Table 1.** Selected characteristics of OneFlorida+ and UPMC<sup>a</sup> health system patients.

Characteristics	2017-22 OneFlorida+ data			2018-20 UPMC <sup>a</sup> data
	Training ( $n=60,694$ )	Testing ( $n=60,695$ )	Internal validation ( $n=60,694$ )	External validation ( $n=129,215$ )
Development of incident opioid use disorder, n (%)	1443 (2.38)	1433 (2.36)	1452 (2.39)	2537 (1.96)
Age (years), mean (SD)	53.11 (17.57)	52.98 (17.64)	52.98 (17.57)	58.91 (16.75)
Age group (years), n (%)				
18-34	10,200 (16.81)	10,286 (16.95)	10,285 (16.95)	13,852 (10.73)
35-50	15,326 (25.25)	15,501 (25.54)	15,477 (25.5)	23,931 (18.52)
51-64	19,121 (31.5)	19,067 (31.41)	19,096 (31.46)	44,047 (34.10)
$\geq 65$	16,047 (26.44)	15,841 (26.1)	15,836 (26.09)	47,385 (36.65)
Sex, n (%)				
Male	23,534 (38.77)	23,489 (38.70)	23,481 (38.69)	50,641 (39.19)
Female	37,160 (61.23)	37,206 (61.30)	37,210 (61.31)	78,574 (60.81)
Race, n (%)				
White	41,737 (68.77)	41,751 (68.79)	41,720 (68.74)	115,907 (89.70)
Black	13,403 (22.08)	13,458 (22.17)	13,548 (22.32)	10,833 (8.38)
Other or unknown	5554 (9.15)	5486 (9.04)	5426 (8.94)	2475 (1.92)
Ethnicity, n (%)				
Non-Hispanic	47,951 (79)	48,242 (79.48)	48,123 (79.29)	123,083 (95.25)
Hispanic	10,756 (17.72)	10,455 (17.23)	10,540 (17.37)	1013 (0.78)
Other or unknown	1987 (3.27)	1998 (3.29)	2031 (3.35)	5123 (3.96)

<sup>a</sup>UPMC: Formerly known as University of Pittsburgh Medical Center.

## Prediction Performance Across ML Methods

[Figure 1](#) summarizes the 4 prediction performance measures of each model in the internal validation sample. GBM (C-statistics=0.879, 95% CI 0.874-0.884) and EN (C-statistics=0.872, 95% CI 0.867-0.877) models outperformed the LASSO (C-statistics=0.846, 95% CI 0.840-0.851) and RF (C-statistics=0.798, 95% CI 0.792-0.804) models, with the GBM model requiring the fewest predictors ( $n=75$ ) for

predicting 3-month incident OUD ( $P<.001$ ; [Figure 1A](#)). The GBM and EN models had similar prediction performance, and the GBM model had the best precision-recall performance ([Figure 1B](#)). Sensitivity analyses using patient-level data yielded similar results ([Figure S4](#) in [Multimedia Appendix 1](#)). To evaluate model robustness across key clinical subgroups, we stratified the GBM model's performance by age, sex, and ethnicity. The model demonstrated consistent discriminative ability across populations, with C-statistics ranging from 0.833 to 0.909 across age groups ([Figure S5](#) in [Multimedia](#)

Appendix 1), 0.865 to 0.887 between sexes (Figure S6 in Multimedia Appendix 1), and 0.864 to 0.914 by ethnicity (Figure S7 in Multimedia Appendix 1).

**Figure 1.** Performance matrix across machine learning models for predicting incident opioid use disorder in OneFlorida+ patients. This figure shows 4 prediction performance matrices in the validation sample. (A) The areas under receiver operating characteristic curves (or C-statistics). (B) The precision-recall curves (precision=positive predictive value and recall=sensitivity): precision-recall curves that are closer to the upper right corner or are above another method have improved performance. (C) The number needed to evaluate by different cutoffs of sensitivity. (D) Alerts per 100 patients by different cutoffs of sensitivity. AUC: area under the curve; EN: elastic net; GBM: gradient boosting machine; LASSO: least absolute shrinkage and selection operator; PPV: positive predictive value; RF: random forest; ROC: receiver operating characteristics.

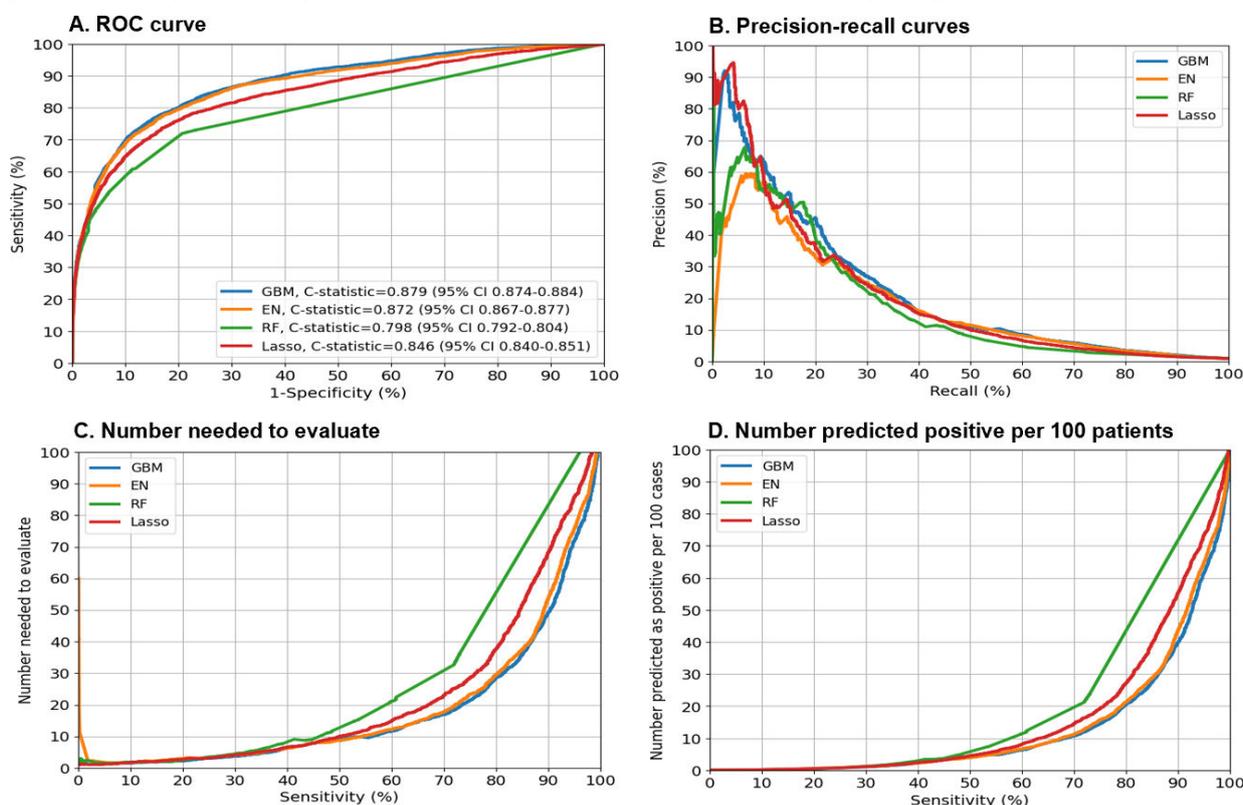


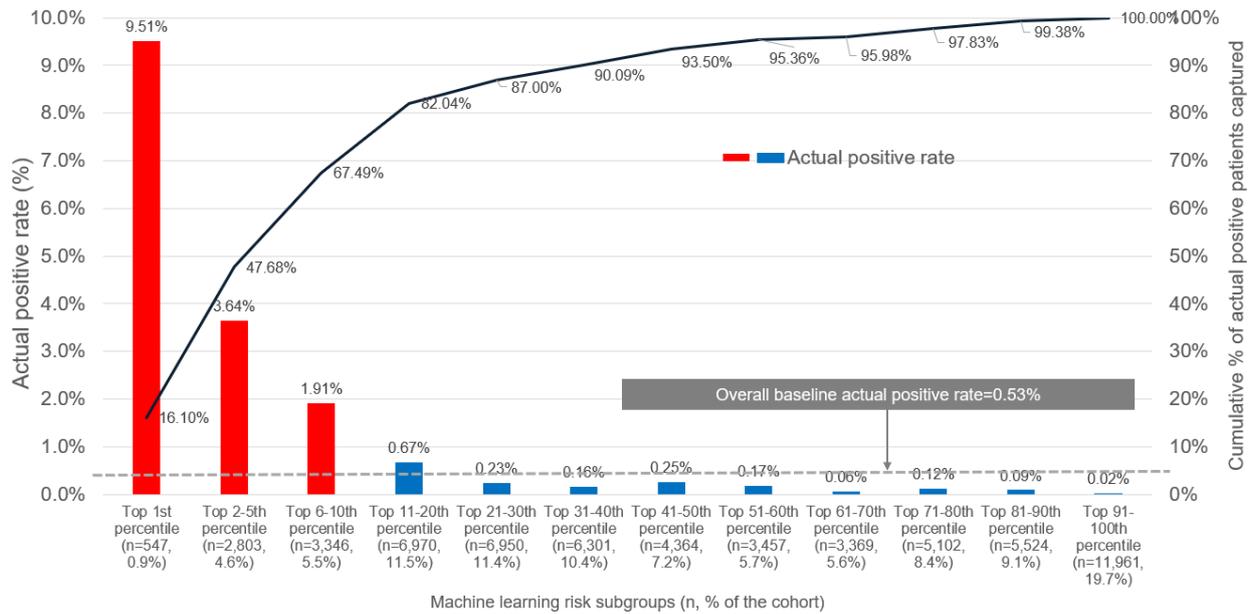
Table S4 in Multimedia Appendix 2 shows the performance measures for predicting incident OUD across different levels (90%-100%) of sensitivity and specificity for GBM and EN. When set at the optimized sensitivity and specificity as measured by the Youden index, the GBM model had a sensitivity of 76.59%, specificity of 84.75%, PPV of 4.36%, NPV of 99.75%, and NNE of 23; and the EN model had a sensitivity of 77.65%, specificity of 82.74%, PPV of 3.93%, NPV of 99.76%, and NNE of 25 (Figures 1C and 1D; Table S4 in Multimedia Appendix 2).

### Risk Stratification, Bias and Fairness Analysis, and Model Explainability

Figure 2 represents the actual OUD rate for individuals in each decile subgroup using GBM. The high-risk subgroup

(with risk scores in the top decile; 11% [n=6696] of the validation cohort) had a PPV of 3.26%, a NPV of 99.8%, and NNE of 31. In a hypothetical clinical application involving 1000 patients, applying the top decile threshold would generate alerts for approximately 100 individuals. Based on the observed PPV at this threshold, approximately 4 of the 100 flagged patients would be expected to develop OUD, demonstrating the model's ability to meaningfully concentrate risk among a small subgroup. Conversely, the high NPV (99.8%) indicates that very few incident OUD cases would arise among the 900 unflagged patients, consistent with the model's low false negative rate. Taken together, these estimates illustrate how the model could support prioritization of patients most likely to benefit from closer monitoring or preventive intervention.

**Figure 2.** Incident opioid use disorder was identified by the gradient boosting machine’s decile risk subgroup in the validation sample. Based on the individual’s predicted probability of an opioid use disorder event, we classified patients in the internal validation sample into decile risk subgroups, with the highest decile further split into 3 additional strata based on the top first, second to fifth, and sixth to tenth percentiles to allow closer examination of patients at the highest risk of developing opioid use disorder.



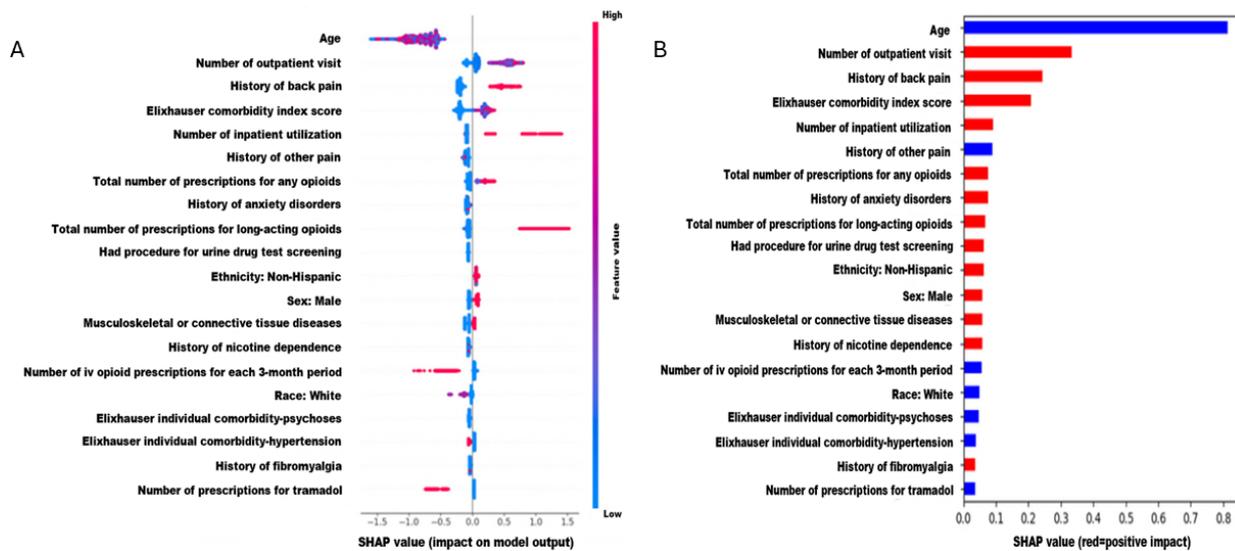
Among all 323 individuals with an incident OUD, 265 (82.04%) occurred in the top two decile subgroups (Decile 1=67.5 % and Decile 2=14.6%). Those in the first decile subgroup had at least a 14-fold higher OUD rate compared to the lower-risk groups (eg, observed OUD rate: Decile 1=9.51%, Decile 2=0.67%, Decile 10=0.02%). The third through tenth decile subgroups had minimal OUD incidence rate (2 to 16 per 10,000).

In our evaluation of racial fairness, we found that among the true positive cases identified in the top decile risk subgroup by the GBM model, 78.44% (171/218) were White individuals and 18.8% (41/218) were Black individuals. In the top first percentile, the distribution was 67.3% (35/52) White individuals and 25% (13/52) Black individuals. The detailed racial distribution of true positives across all risk subgroups is provided in Table S5 in Multimedia Appendix 2. We also compared FNR and FPR by race in the GBM model with and without race as a predictor. FNRs were consistently higher for Black individuals compared to White individuals across most percentiles, while FPR remained similar across racial groups (Figure S8 and Figure S9 in Multimedia Appendix 1). We compared the FNR and FPR between racial groups at the 90th percentile and found that the FNR parity ratio (1.23) and FPR parity ratio (1.03) were within the accepted threshold (0.8-1.25) [62], suggesting equitable model

performance across racial groups when race was included as a predictor. We also assessed fairness across age and sex. The true positive cases in the top decile were concentrated among the 35- to 50-year age group (82/218, 37.61%) and females (146/218, 66.97%). The detailed distribution of true positives by age and sex across all risk subgroups is provided in Tables S6 and S7 in Multimedia Appendix 2. FNRs were slightly higher for older individuals (≥65 y) across most percentiles compared to the young adults (18-34 y), while FNRs for males and females were closely aligned. FPRs remained consistent across both age groups and sex (Figures S10 and S11 in Multimedia Appendix 1). Parity ratios for sex (FNR=0.91, FPR=1.18) and age-based FPR (0.91; comparing 18-34 vs ≥65 y) remained within the acceptable range (0.8-1.25), whereas the age-based FNR parity ratio (0.74) was below the threshold.

Figure 3 presents variable importance plots using SHAP values derived from the GBM model. The plots on the right-hand side present the average impact of various features on incident OUD, listed in descending order of their significance. Variables such as age and history of other pain for each 3-month period were negatively correlated with incident OUD, while factors having a positive correlation with incident OUD included the number of outpatient visits, history of back pain, and Elixhauser comorbidity index score.

**Figure 3.** Variable importance plots using Shapley additive explanations values from the gradient boosting machine model. (A) Shapley additive explanations value computed from individual features' values and their impact (both positive and negative) on incident opioid use disorder. (B) Average Shapley additive explanations values of features showing average impact on and correlation with incident opioid use disorder. SHAP: Shapley additive explanations.



### Calibration and Decision Curve Analysis

Calibration assessment showed that the GBM model initially overestimated risk (Brier score=0.09), a known phenomenon in rare-event prediction. After isotonic recalibration, model reliability improved significantly, with a calibration slope of 0.98 (ideal=1), an intercept of -0.03 (ideal=0), and a Brier score of 0.01 (lower is better; Figure S12 in [Multimedia Appendix 1](#)). Decision-curve analysis showed that a “treat-all” strategy (intervene on everyone) yields negative net benefit at risk thresholds above 0.5% (Figure S13 in [Multimedia Appendix 1](#)), whereas the GBM model maintained positive net benefit across a broader range of clinically relevant thresholds. For example, at the top decile cutoff (probability threshold=0.009), the model achieved a net benefit of 0.0026. This corresponds to identifying approximately 2.6 additional true-positive OUD cases per 1000 patients (after accounting for false-positive alerts) compared with a strategy of no intervention. Relative to the OUD outcome incidence (0.53% in the prediction window), this net benefit reflects ~50% of the maximum achievable use for a perfect classifier in this setting, indicating substantial clinical value despite the low-prevalence context.

### External Validation Performance

When applied to the UPMC cohort, the best-performing GBM model achieved a C-statistic of 0.756 (95% CI 0.750-0.762; Figure S14 in [Multimedia Appendix 1](#)). The high-risk decile subgroup demonstrated a PPV of 2.52%, an NPV of 99.8%, and NNE of 40 (Figure S15 in [Multimedia Appendix 1](#)). The top first percentile had an OUD incidence rate 14-fold higher than the overall baseline (8.15% vs 0.58%). Fairness analyses in the UPMC cohort revealed variations in FNR and FPR across age groups (Figure S16 in [Multimedia Appendix 1](#)), while error rates remained similar across race (Figure S17 in [Multimedia Appendix 1](#)) and between males and females (Figure S18 in [Multimedia Appendix 1](#)). For example, using the 90th percentile risk score as the elevated risk cutoff,

younger adults (18-34 y) had a lower FNR (fewer missed cases) compared to older adults (≥65 y).

## Discussion

### Principal Findings

This study expanded our previous work using ML approaches to improve the accuracy of predicting incident OUD in the subsequent 3 months of prescription opioid initiation among fee-for-service Medicare beneficiaries and broaden the applicability of these models across a diverse population in the OneFlorida+ network [29]. Our GBM and EN models demonstrated strong predictive performance, achieving a high C-statistic (>0.87) and outperforming LASSO and RF. Our best-performing GBM model offers several advantages, including its ability to eliminate the need for a separate feature selection process and its flexibility in hyperparameter tuning to capture complex interactions between predictors and outcomes. We acknowledge that this flexibility in model tuning can be computationally intensive and time-consuming. The algorithm effectively stratified the population into distinct risk groups based on predicted risk scores. Notably, approximately 80% of the cohort had minimal OUD risk, while the highest-risk decile alone accounted for ~68% (218/323) of all individuals who developed OUD. Given the low incidence of OUD within a 3-month period, the PPV was expectedly low. However, in the context of OUD prevention, a lower PPV may be ethically and clinically acceptable when associated interventions are supportive (eg, closer follow-up, motivational interviewing, review of opioid therapy, or other high-risk concurrent medication use) rather than punitive actions (eg, refusal to prescribe or abrupt discontinuation). For these low-risk, supportive interventions, the potential harm of a false positive (providing additional support to a patient who would not have developed OUD) is minimal compared with the harm of missing a true high-risk case

(false negative). It is therefore important to frame the model as augmenting, not replacing, clinician judgment to ensure that low PPV does not lead to overreaction to false positives or unintended consequences.

A key strength of this study is the successful external validation of our algorithm using an independent UPMC cohort. Our model maintained good performance (C-statistic: 0.76) without retraining, demonstrating transportability across two distinct health care systems. This replication supports the stability of the key predictors across settings and indicates that the model can effectively identify high-risk individuals in diverse clinical environments.

Prior studies, targeting different aspects of OUD risk, varied in their prediction windows and data sources, including 6-month OUD risk based on private insurance claims [66]; 12-month risk of aberrant opioid use behaviors following an initial pain clinic visit [67]; and 12-month OUD risk using private insurance claims [38,68] or pharmacy benefit manager claims data [36]. Other models focused on longer-term risk, such as 2-year problematic opioid use documented in EHR from primary care [69] and 5-year OUD risk using electronic medical record from a medical center [34] or Rhode Island Medicaid data [35]. Despite their contributions, these models had key limitations. Most measured predictors are only at baseline, lacking temporal updates over time. Several relied on case-control designs, which may not generalize well to population-level OUD incidence rates. Additionally, even in non-case-control designs, the highest reported C-statistic was 0.85, indicating room for improvement in predictive performance [35,36,67,69]. Expanding on our prior Medicare claims-based study, this work explored whether predictive performance can be refined using EHR data. Our study addressed the abovementioned limitations by using a population-based cohort and a more immediate prediction window, estimating OUD risk within the next 3 months rather than over a year or longer. This short-term predictive approach, combined with innovative risk stratification and enhanced clinical interpretability, made our method highly applicable for timely intervention.

## Limitations

Although the findings of the study are promising, some limitations need to be acknowledged. First, incident OUD was identified solely through *ICD-10* diagnosis codes, which may undercapture true OUD cases due to known under-coding, variability in diagnostic practices, and reluctance to document substance use disorders in clinical settings. As a result, some individuals with OUD may have been misclassified as noncases. Although prior work has similarly relied on diagnosis-based definitions [29,34-36], broader definitions incorporating buprenorphine treatment or combinations of high-risk prescribing behavior indicators may improve case ascertainment. Second, unmeasured predictors, such as socioeconomic determinants and illicit opioid use, could influence risk trajectories. Third, while the study uniquely assessed OUD risk within 3 months of opioid initiation, its reliance on older EHR data may limit real-time applicability due to potential delays in data availability and processing.

## Clinical Utility

From a clinical use perspective, the model's ability to stratify patients by risk supports early intervention and more efficient resource allocation in primary care settings. Our analysis confirms that the model's risk estimates are not only accurate in distinguishing higher- from lower-risk patients but also well calibrated after post hoc adjustment. Decision-curve analysis further demonstrated that the model offers meaningful clinical utility across thresholds where default "treat-all" strategies offer little value in the context of a low outcome incidence. At the top decile threshold, the model achieved a positive net benefit equivalent to identifying approximately 2.6 additional true OUD cases per 1000 patients—representing about 50% of the maximum achievable use given the cohort's baseline incidence. These findings align with established decision-analytic frameworks [70] and highlight the model's potential to effectively prioritize individuals most likely to benefit from preventive intervention. Integrating such models into clinical decision support (CDS) tools could enable providers to proactively flag and manage patients at elevated risk for OUD before symptoms escalate. However, ethical considerations must also be addressed, including the potential for unintentionally reinforcing existing health disparities. Although the overall FNR and FPR parity ratio was within acceptable bounds, Black patients exhibited higher false negative rates across percentiles, suggesting that true OUD cases may be underidentified in this group. To address this disparity, we outline potential mitigation strategies for future model refinement. These include group-specific threshold adjustment to harmonize error rates, reweighting or cost-sensitive learning to shift model attention toward underidentified subgroups, and postprocessing approaches such as calibration-by-group or equalized-odds adjustment. Importantly, any fairness intervention should balance reducing disparities with avoiding unintended undertreatment or new forms of inequity. These considerations underscore the need for iterative fairness evaluation as part of a responsible deployment framework.

## Clinical Implementation and Future Directions

To support real-world deployment, our translational roadmap includes a structured, multiphase implementation strategy grounded in principles of human-centered design and clinical workflow alignment. In the retrospective phase, we will operationalize the OUD risk model within the UF Health Integrated Data Repository, a comprehensive enterprise clinical data warehouse. This step will enable validation using real-time patient data and ensure backend compatibility with the Epic EHR system. The model will generate individual-level risk scores for patients with recent opioid prescriptions, using secure internal infrastructure and consistent EHR data formatting. In the next phase, we will begin with a feasibility test and pilot integration of the model into the EHR system (Epic) within UF Health, gathering real-world feedback from clinicians before scaling up. Notably, UF Health's Epic EHR infrastructure has already implemented a system-wide

approach for an artificial intelligence–driven CDS tool that predicts patients with a high risk of opioid overdose in primary care workflows [71,72].

In the near term, we plan to deploy the OUD risk prediction model through EHR-based CDS tools that seamlessly integrate into routine care. The model's risk score will trigger automated alerts during opioid prescribing encounters, prompting timely interventions for those identified as high risk. For example, in a primary care clinic with approximately 1000 patients initiating opioid therapy annually, applying a top decile (10%) risk threshold would flag approximately 100 individuals as high risk, of whom ~4 would be expected to develop OUD within 3 months. Although the PPV remains low due to the rarity of incident OUD, the alert volume is manageable in routine practice, particularly when recommended interventions are supportive and low-intensity (eg, increased follow-up, medication review, and motivational interviewing). The remaining ~90% low-risk patients would have an extremely high NPV, allowing providers to continue appropriate pain management for those individuals without unnecessary alarm. To further address alert fatigue, we propose specific operational strategies such as (1) embedding alerts within existing opioid or pain management workflows (eg, triggering only when signing a new medication order) to minimize disruption, (2) using tiered alerting triggered by repeated high-risk predictions, and (3) combining risk scores with other clinically meaningful indicators (eg, concurrent benzodiazepine prescribing) to enhance alert specificity. To minimize potential harms such as stigma and inappropriate opioid restriction, alerts should be framed using universal-precaution

language that encourages supportive monitoring rather than implying misuse.

Although our findings indicate that the model has characteristics that could support future integration into clinical workflows, real-world feasibility and impact have not yet been evaluated. Additional implementation studies, including usability testing, workflow integration, and prospective evaluation, are needed before broad deployment in routine practice. While our separate pilot trial, Developing and Evaluating a Machine Learning Opioid Prediction & Risk-Stratification E-Platform, is currently assessing the feasibility of integrating an EHR-based overdose alert into clinical workflows within UF Health primary care clinics [71,72], analogous testing would be required for the model developed in this study.

## Conclusions

This study demonstrates that an ML algorithm derived from the OneFlorida+ network can accurately predict incident OUD and effectively stratify risk across diverse populations and health systems, as evidenced by successful external validation in UPMC. The next step is to prospectively validate this algorithm within the UF Health clinical workflow to leverage existing decision support infrastructure and facilitate early intervention. Future research should expand the model to incorporate unstructured clinical notes, prescription drug monitoring program data, and social drivers of health (eg, housing instability, employment status, and incarceration history) to further improve prediction accuracy and equity.

## Funding

This project is funded by the US National Institute on Drug Abuse (R01DA050676 and R01DA044985). WLC and WFG are named as inventors in one preliminary patent (U1195.70174US00) filing from the University of Florida and University of Pittsburgh for use of a machine-learning algorithm for opioid risk prediction in Medicare. The views presented here are those of the authors alone and do not necessarily represent the views of the Department of Veterans Affairs or the United States Government.

## Data Availability

The full modeling pipeline, including code for data preprocessing, training, validation, and fairness analysis, is publicly available in the GitHub repository [73]. The datasets generated or analyzed during this study are not publicly available due to patient privacy regulations, Health Insurance Portability and Accountability Act, but are available from the OneFlorida+ network upon reasonable request and institutional review board approval.

## Authors' Contributions

Conceptualization: JAF, WL-C, WFG, YW, CAH, KN, JLH, GC, DLW, SASS, SOFS, EIR, DN, SY, GMR, WMG, CK, MMH

Data curation: JAF, JLH, WLC, and MMH

Formal analysis: JAF, JLH, WLC, and MMH

Funding acquisition: WLC

Investigation: All authors

Methodology: All authors

Project administration: WLC

Resources: JAF, JLH, WLC, and MMH

Software: JAF, WLC, and MMH

Supervision: WLC and MMH

Validation: JAF, JLH, WLC, and MMH

Visualization: JAF, WLC, and MMH

Writing – original draft: JAF and MMH

Writing – review and editing: All authors

### Conflicts of Interest

DLW and WLC have received grant funding from Merck, Sharp & Dohme and Bristol Myers Squibb for unrelated research. WLC was compensated for consulting services by Teva Pharmaceuticals.

### Multimedia Appendix 1

Study design, model performance (internal and external validation), calibration and decision curves, and risk stratification. [[DOCX File \(Microsoft Word File\), 2663 KB-Multimedia Appendix 1](#)]

### Multimedia Appendix 2

Diagnosis codes, predictor definitions, and model performance. [[DOCX File \(Microsoft Word File\), 85 KB-Multimedia Appendix 2](#)]

### Multimedia Appendix 3

Methods for machine learning model development and evaluation. [[DOCX File \(Microsoft Word File\), 58 KB-Multimedia Appendix 3](#)]

### Checklist 1

TRIPOD+AI checklist. [[DOCX File \(Microsoft Word File\), 47 KB-Checklist 1](#)]

### References

1. Florence C, Luo F, Rice K. The economic burden of opioid use disorder and fatal opioid overdose in the United States, 2017. *Drug Alcohol Depend*. Jan 1, 2021;218:108350. [doi: [10.1016/j.drugalcdep.2020.108350](https://doi.org/10.1016/j.drugalcdep.2020.108350)] [Medline: [33121867](https://pubmed.ncbi.nlm.nih.gov/33121867/)]
2. Hébert AH, Hill AL. Impact of opioid overdoses on US life expectancy and years of life lost, by demographic group and stimulant co-involvement: a mortality data analysis from 2019 to 2022. *Lancet Reg Health Am*. Aug 2024;36:100813. [doi: [10.1016/j.lana.2024.100813](https://doi.org/10.1016/j.lana.2024.100813)] [Medline: [38978785](https://pubmed.ncbi.nlm.nih.gov/38978785/)]
3. Leslie DL, Ba DM, Agbese E, Xing X, Liu G. The economic burden of the opioid epidemic on states: the case of Medicaid. *Am J Manag Care*. Jul 2019;25(13 Suppl):S243-S249. [Medline: [31361426](https://pubmed.ncbi.nlm.nih.gov/31361426/)]
4. Keyes KM, Rutherford C, Hamilton A, et al. What is the prevalence of and trend in opioid use disorder in the United States from 2010 to 2019? Using multiplier approaches to estimate prevalence for an unknown population size. *Drug Alcohol Depend Rep*. Jun 2022;3:100052. [doi: [10.1016/j.dadr.2022.100052](https://doi.org/10.1016/j.dadr.2022.100052)] [Medline: [35783994](https://pubmed.ncbi.nlm.nih.gov/35783994/)]
5. About overdose prevention. Centers for Disease Control and Prevention (CDC). URL: <https://www.cdc.gov/overdose-prevention/about/index.html> [Accessed 2024-05-04]
6. Kline D, Hepler SA, Krawczyk N, Rivera-Aguirre A, Waller LA, Cerdá M. A state-level history of opioid overdose deaths in the United States: 1999-2021. *PLoS ONE*. 2024;19(9):e0309938. [doi: [10.1371/journal.pone.0309938](https://doi.org/10.1371/journal.pone.0309938)] [Medline: [39240938](https://pubmed.ncbi.nlm.nih.gov/39240938/)]
7. Chen LH, Hedegaard H, Warner M. Drug-poisoning deaths involving opioid analgesics: United States, 1999-2011. *NCHS Data Brief*. Sep 2014(166):1-8. [Medline: [25228059](https://pubmed.ncbi.nlm.nih.gov/25228059/)]
8. Ruhm CJ. Corrected US opioid-involved drug poisoning deaths and mortality rates, 1999–2015. *Addiction*. Jul 2018;113(7):1339-1344. [doi: [10.1111/add.14144](https://doi.org/10.1111/add.14144)] [Medline: [29430760](https://pubmed.ncbi.nlm.nih.gov/29430760/)]
9. Overdose prevention: about prescription opioids. Centers for Disease Control and Prevention (CDC). URL: <https://www.cdc.gov/overdose-prevention/about/prescription-opioids.html> [Accessed 2024-10-14]
10. National Institute on Drug Abuse, NIDA. Only 1 in 5 US adults with opioid use disorder received medications to treat it in 2021. URL: <https://nida.nih.gov/news-events/news-releases/2023/08/only-1-in-5-us-adults-with-opioid-use-disorder-received-medications-to-treat-it-in-2021> [Accessed 2024-10-14]
11. Volkow ND, Blanco C. The changing opioid crisis: development, challenges and opportunities. *Mol Psychiatry*. Jan 2021;26(1):218-233. [doi: [10.1038/s41380-020-0661-4](https://doi.org/10.1038/s41380-020-0661-4)] [Medline: [32020048](https://pubmed.ncbi.nlm.nih.gov/32020048/)]
12. Lo-Ciganic WH, Donohue JM, Yang Q, et al. Developing and validating a machine-learning algorithm to predict opioid overdose in medicaid beneficiaries in two US states: a prognostic modelling study. *Lancet Digit Health*. Jun 2022;4(6):e455-e465. [doi: [10.1016/S2589-7500\(22\)00062-0](https://doi.org/10.1016/S2589-7500(22)00062-0)] [Medline: [35623798](https://pubmed.ncbi.nlm.nih.gov/35623798/)]
13. Edlund MJ, Martin BC, Russo JE, DeVries A, Braden JB, Sullivan MD. The role of opioid prescription in incident opioid abuse and dependence among individuals with chronic noncancer pain: the role of opioid prescription. *Clin J Pain*. Jul 2014;30(7):557-564. [doi: [10.1097/AJP.0000000000000021](https://doi.org/10.1097/AJP.0000000000000021)] [Medline: [24281273](https://pubmed.ncbi.nlm.nih.gov/24281273/)]

14. Yang Z, Wilsey B, Bohm M, et al. Defining risk of prescription opioid overdose: pharmacy shopping and overlapping prescriptions among long-term opioid users in medicaid. *J Pain*. May 2015;16(5):445-453. [doi: [10.1016/j.jpain.2015.01.475](https://doi.org/10.1016/j.jpain.2015.01.475)] [Medline: [25681095](https://pubmed.ncbi.nlm.nih.gov/25681095/)]
15. Egualé T, Bastardot F, Song W, et al. A machine learning application to classify patients at differing levels of risk of opioid use disorder: clinician-based validation study. *JMIR Med Inform*. Jun 4, 2024;12:e53625. [doi: [10.2196/53625](https://doi.org/10.2196/53625)] [Medline: [38842167](https://pubmed.ncbi.nlm.nih.gov/38842167/)]
16. Screener and opioid assessment for patients with pain-revised (SOAPP®-r). Inflexxion, Inc; 2008. URL: <https://ddph-materials.s3.amazonaws.com/HelpsHere/SOAPP-Tool.pdf> [Accessed 2026-01-22]
17. Cheattle MD, Compton PA, Dhingra L, Wasser TE, O'Brien CP. Development of the revised opioid risk tool to predict opioid use disorder in patients with chronic nonmalignant pain. *J Pain*. Jul 2019;20(7):842-851. [doi: [10.1016/j.jpain.2019.01.011](https://doi.org/10.1016/j.jpain.2019.01.011)] [Medline: [30690168](https://pubmed.ncbi.nlm.nih.gov/30690168/)]
18. Chou R, Hartung D, Turner J, et al. Opioid Treatments for Chronic Pain. Agency for Healthcare Research and Quality (US); 2020. URL: <https://www.ncbi.nlm.nih.gov/sites/books/NBK556253/> [Accessed 2026-02-13]
19. Bamboo health. NarxCare. URL: <https://bamboohealth.com/solutions/narxcare/> [Accessed 2025-11-03]
20. Wei YJJ, Chen C, Sarayani A, Winterstein AG. Performance of the centers for medicare & medicaid services' opioid overutilization criteria for classifying opioid use disorder or overdose. *JAMA*. Feb 12, 2019;321(6):609-611. [doi: [10.1001/jama.2018.20404](https://doi.org/10.1001/jama.2018.20404)] [Medline: [30747958](https://pubmed.ncbi.nlm.nih.gov/30747958/)]
21. Canan C, Polinski JM, Alexander GC, Kowal MK, Brennan TA, Shrank WH. Automatable algorithms to identify nonmedical opioid use using electronic data: a systematic review. *J Am Med Inform Assoc*. Nov 1, 2017;24(6):1204-1210. [doi: [10.1093/jamia/ocx066](https://doi.org/10.1093/jamia/ocx066)] [Medline: [29016967](https://pubmed.ncbi.nlm.nih.gov/29016967/)]
22. Rough K, Huybrechts KF, Hernandez-Diaz S, Desai RJ, Patorno E, Bateman BT. Using prescription claims to detect aberrant behaviors with opioids: comparison and validation of 5 algorithms. *Pharmacoepidemiol Drug Saf*. Jan 2019;28(1):62-69. [doi: [10.1002/pds.4443](https://doi.org/10.1002/pds.4443)] [Medline: [29687539](https://pubmed.ncbi.nlm.nih.gov/29687539/)]
23. Goyal H, Singla U, Grimsley EW. Identification of opioid abuse or dependence: no tool is perfect. *Am J Med*. Mar 2017;130(3):e113. [doi: [10.1016/j.amjmed.2016.09.022](https://doi.org/10.1016/j.amjmed.2016.09.022)] [Medline: [28215952](https://pubmed.ncbi.nlm.nih.gov/28215952/)]
24. Wood E, Simel DL, Klimas J. Pain management with opioids in 2019-2020. *JAMA*. Nov 19, 2019;322(19):1912-1913. [doi: [10.1001/jama.2019.15802](https://doi.org/10.1001/jama.2019.15802)] [Medline: [31600370](https://pubmed.ncbi.nlm.nih.gov/31600370/)]
25. Dong X, Deng J, Rashidian S, et al. Identifying risk of opioid use disorder for patients taking opioid medications with deep learning. *J Am Med Inform Assoc*. Jul 30, 2021;28(8):1683-1693. [doi: [10.1093/jamia/ocab043](https://doi.org/10.1093/jamia/ocab043)] [Medline: [33930132](https://pubmed.ncbi.nlm.nih.gov/33930132/)]
26. Gao W, Leighton C, Chen Y, Jones J, Mistry P. Predicting opioid use disorder and associated risk factors in a Medicaid managed care population. *Am J Manag Care*. Apr 2021;27(4):148-154. [doi: [10.37765/ajmc.2021.88617](https://doi.org/10.37765/ajmc.2021.88617)] [Medline: [33877773](https://pubmed.ncbi.nlm.nih.gov/33877773/)]
27. Hasan MM, Young GJ, Patel MR, Modestino AS, Sanchez LD, Noor-E-Alam M. A machine learning framework to predict the risk of opioid use disorder. *Machine Learning with Applications*. Dec 2021;6:100144. [doi: [10.1016/j.mlwa.2021.100144](https://doi.org/10.1016/j.mlwa.2021.100144)]
28. Segal Z, Radinsky K, Elad G, et al. Development of a machine learning algorithm for early detection of opioid use disorder. *Pharmacol Res Perspect*. Dec 2020;8(6):e00669. [doi: [10.1002/prp2.669](https://doi.org/10.1002/prp2.669)] [Medline: [33200572](https://pubmed.ncbi.nlm.nih.gov/33200572/)]
29. Lo-Ciganic WH, Huang JL, Zhang HH, et al. Using machine learning to predict risk of incident opioid use disorder among fee-for-service medicare beneficiaries: a prognostic study. *PLoS ONE*. 2020;15(7):e0235981. [doi: [10.1371/journal.pone.0235981](https://doi.org/10.1371/journal.pone.0235981)] [Medline: [32678860](https://pubmed.ncbi.nlm.nih.gov/32678860/)]
30. Hogan WR, Shenkman EA, Robinson T, et al. The OneFlorida data trust: a centralized, translational research data infrastructure of statewide scope. *J Am Med Inform Assoc*. Mar 15, 2022;29(4):686-693. [doi: [10.1093/jamia/ocab221](https://doi.org/10.1093/jamia/ocab221)] [Medline: [34664656](https://pubmed.ncbi.nlm.nih.gov/34664656/)]
31. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. Apr 16, 2024;385:e078378. [doi: [10.1136/bmj-2023-078378](https://doi.org/10.1136/bmj-2023-078378)] [Medline: [38626948](https://pubmed.ncbi.nlm.nih.gov/38626948/)]
32. Smith SM, Winterstein AG, Gurka MJ, et al. Initial antihypertensive regimens in newly treated patients: real world evidence from the OneFlorida+ clinical research network. *J Am Heart Assoc*. Jan 3, 2023;12(1):e026652. [doi: [10.1161/JAHA.122.026652](https://doi.org/10.1161/JAHA.122.026652)] [Medline: [36565195](https://pubmed.ncbi.nlm.nih.gov/36565195/)]
33. OneFlorida+ clinical research network. The OneFlorida+ Data Trust. URL: <https://onefl.net/data/> [Accessed 2025-01-07]
34. Ellis RJ, Wang Z, Genes N, Ma'ayan A. Predicting opioid dependence from electronic health records with machine learning. *BioData Min*. 2019;12(1):3. [doi: [10.1186/s13040-019-0193-0](https://doi.org/10.1186/s13040-019-0193-0)] [Medline: [30728857](https://pubmed.ncbi.nlm.nih.gov/30728857/)]
35. Hastings JS, Howison M, Inman SE. Predicting high-risk opioid prescriptions before they are given. *Proc Natl Acad Sci U S A*. Jan 28, 2020;117(4):1917-1923. [doi: [10.1073/pnas.1905355117](https://doi.org/10.1073/pnas.1905355117)] [Medline: [31937665](https://pubmed.ncbi.nlm.nih.gov/31937665/)]

36. Ciesielski T, Iyengar R, Bothra A, Tomala D, Cislo G, Gage BF. A tool to assess risk of de novo opioid abuse or dependence. *Am J Med*. Jul 2016;129(7):699-705. [doi: [10.1016/j.amjmed.2016.02.014](https://doi.org/10.1016/j.amjmed.2016.02.014)] [Medline: [26968469](https://pubmed.ncbi.nlm.nih.gov/26968469/)]
37. Hall AJ, Logan JE, Toblin RL, et al. Patterns of abuse among unintentional pharmaceutical overdose fatalities. *JAMA*. Dec 10, 2008;300(22):2613-2620. [doi: [10.1001/jama.2008.802](https://doi.org/10.1001/jama.2008.802)] [Medline: [19066381](https://pubmed.ncbi.nlm.nih.gov/19066381/)]
38. White AG, Birnbaum HG, Schiller M, Tang J, Katz NP. Analytic models to identify patients at risk for prescription opioid abuse. *Am J Manag Care*. Dec 2009;15(12):897-906. [Medline: [20001171](https://pubmed.ncbi.nlm.nih.gov/20001171/)]
39. Sullivan MD, Edlund MJ, Fan MY, DeVries A, Braden JB, Martin BC. Risks for possible and probable opioid misuse among recipients of chronic opioid therapy in commercial and medicaid insurance plans: The TROUP Study. *Pain*. Aug 2010;150(2):332-339. [doi: [10.1016/j.pain.2010.05.020](https://doi.org/10.1016/j.pain.2010.05.020)] [Medline: [20554392](https://pubmed.ncbi.nlm.nih.gov/20554392/)]
40. Cepeda MS, Fife D, Chow W, Mastrogiovanni G, Henderson SC. Assessing opioid shopping behaviour: a large cohort study from a medication dispensing database in the US. *Drug Saf*. Apr 1, 2012;35(4):325-334. [doi: [10.2165/11596600-000000000-00000](https://doi.org/10.2165/11596600-000000000-00000)] [Medline: [22339505](https://pubmed.ncbi.nlm.nih.gov/22339505/)]
41. Cochran G, Gordon AJ, Lo-Ciganic WH, et al. An examination of claims-based predictors of overdose from a large medicaid program. *Med Care*. Mar 2017;55(3):291-298. [doi: [10.1097/MLR.0000000000000676](https://doi.org/10.1097/MLR.0000000000000676)] [Medline: [27984346](https://pubmed.ncbi.nlm.nih.gov/27984346/)]
42. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*. 2015;10(3):e0118432. [doi: [10.1371/journal.pone.0118432](https://doi.org/10.1371/journal.pone.0118432)] [Medline: [25738806](https://pubmed.ncbi.nlm.nih.gov/25738806/)]
43. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care*. Aug 13, 2015;19(1):285. [doi: [10.1186/s13054-015-0999-1](https://doi.org/10.1186/s13054-015-0999-1)] [Medline: [26268570](https://pubmed.ncbi.nlm.nih.gov/26268570/)]
44. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B*. Apr 1, 2005;67(2):301-320. [doi: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x)]
45. Larochelle MR, Zhang F, Ross-Degnan D, Wharam JF. Rates of opioid dispensing and overdose after introduction of abuse-deterrent extended-release oxycodone and withdrawal of propoxyphene. *JAMA Intern Med*. Jun 2015;175(6):978-987. [doi: [10.1001/jamainternmed.2015.0914](https://doi.org/10.1001/jamainternmed.2015.0914)] [Medline: [25895077](https://pubmed.ncbi.nlm.nih.gov/25895077/)]
46. Chu A, Ahn H, Halwan B, et al. A decision support system to facilitate management of patients with acute gastrointestinal bleeding. *Artif Intell Med*. Mar 2008;42(3):247-259. [doi: [10.1016/j.artmed.2007.10.003](https://doi.org/10.1016/j.artmed.2007.10.003)] [Medline: [18063351](https://pubmed.ncbi.nlm.nih.gov/18063351/)]
47. Ives TJ, Chelminski PR, Hammett-Stabler CA, et al. Predictors of opioid misuse in patients with chronic pain: a prospective cohort study. *BMC Health Serv Res*. Apr 4, 2006;6(1):46. [doi: [10.1186/1472-6963-6-46](https://doi.org/10.1186/1472-6963-6-46)] [Medline: [16595013](https://pubmed.ncbi.nlm.nih.gov/16595013/)]
48. Becker WC, Sullivan LE, Tetrault JM, Desai RA, Fiellin DA. Non-medical use, abuse and dependence on prescription opioids among U.S. adults: psychiatric, medical and substance use correlates. *Drug Alcohol Depend*. Apr 1, 2008;94(1-3):38-47. [doi: [10.1016/j.drugalcdep.2007.09.018](https://doi.org/10.1016/j.drugalcdep.2007.09.018)] [Medline: [18063321](https://pubmed.ncbi.nlm.nih.gov/18063321/)]
49. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain — United States, 2016. *MMWR Recomm Rep*. 2023;65(1):1-49. [doi: [10.15585/mmwr.rr6501e1](https://doi.org/10.15585/mmwr.rr6501e1)] [Medline: [26987082](https://pubmed.ncbi.nlm.nih.gov/26987082/)]
50. UK NGC, Others. Evidence review: risk factors for dependence. National Institute for Health and Care Excellence (NICE); 2022. URL: <https://www.ncbi.nlm.nih.gov/books/NBK580678/> [Accessed 2026-02-13]
51. How opioid use disorder occurs. Mayo Clinic. URL: <https://www.mayoclinic.org/diseases-conditions/prescription-drug-abuse/in-depth/how-opioid-addiction-occurs/art-20360372> [Accessed 2025-10-31]
52. Guidelines for prescribing controlled substances for pain. Medical Board of California. URL: <https://www.mbc.ca.gov/Resources/Medical-Resources/controlled-substance.aspx> [Accessed 2025-11-03]
53. Ling W, Mooney L, Hillhouse M. Prescription opioid abuse, pain and addiction: clinical issues and implications. *Drug Alcohol Rev*. May 2011;30(3):300-305. [doi: [10.1111/j.1465-3362.2010.00271.x](https://doi.org/10.1111/j.1465-3362.2010.00271.x)] [Medline: [21545561](https://pubmed.ncbi.nlm.nih.gov/21545561/)]
54. Committee on Pain Management and Regulatory Strategies to Address Prescription Opioid Abuse. *Pain Management and the Opioid Epidemic: Balancing Societal and Individual Benefits and Risks of Prescription Opioid Use*. National Academies Press; 2017. [doi: [10.17226/24781](https://doi.org/10.17226/24781)]
55. Lo-Ciganic WH, Huang JL, Zhang HH, et al. Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. *JAMA Netw Open*. Mar 1, 2019;2(3):e190968. [doi: [10.1001/jamanetworkopen.2019.0968](https://doi.org/10.1001/jamanetworkopen.2019.0968)] [Medline: [30901048](https://pubmed.ncbi.nlm.nih.gov/30901048/)]
56. Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2009. [doi: [10.1198/jasa.2004.s339](https://doi.org/10.1198/jasa.2004.s339)]
57. Al Faysal J, Noor-E-Alam M, Young GJ, et al. An explainable machine learning framework for predicting the risk of buprenorphine treatment discontinuation for opioid use disorder among commercially insured individuals. *Comput Biol Med*. Jul 2024;177:108493. [doi: [10.1016/j.combiomed.2024.108493](https://doi.org/10.1016/j.combiomed.2024.108493)] [Medline: [38833799](https://pubmed.ncbi.nlm.nih.gov/38833799/)]

58. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. Sep 1988;44(3):837-845. [doi: [10.2307/2531595](https://doi.org/10.2307/2531595)] [Medline: [3203132](https://pubmed.ncbi.nlm.nih.gov/3203132/)]
59. *Data Mining and Statistics for Decision Making*. Wiley; 2011. [doi: [10.1002/9780470979174](https://doi.org/10.1002/9780470979174)]
60. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J*. Aug 2005;47(4):458-472. [doi: [10.1002/bimj.200410135](https://doi.org/10.1002/bimj.200410135)] [Medline: [16161804](https://pubmed.ncbi.nlm.nih.gov/16161804/)]
61. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using improved shapley additive explanation. Presented at: BCB '19: Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; Sep 7-10, 2019:546; Niagara Falls NY USA. Sep 4, 2019. [doi: [10.1145/3307339.3343255](https://doi.org/10.1145/3307339.3343255)]
62. Saleiro P, Kuester B, Hinkson L, et al. Aequitas: a bias and fairness audit toolkit. *arXiv*. Preprint posted online on Nov 14, 2018. [doi: [10.48550/arXiv.1811.05577](https://doi.org/10.48550/arXiv.1811.05577)]
63. Jiang X, Osl M, Kim J, Ohno-Machado L. Smooth isotonic regression: a new method to calibrate predictive models. *AMIA Jt Summits Transl Sci Proc*. 2011:16-20. [Medline: [22211175](https://pubmed.ncbi.nlm.nih.gov/22211175/)]
64. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol*. Aug 2010;63(8):938-939. [doi: [10.1016/j.jclinepi.2009.11.009](https://doi.org/10.1016/j.jclinepi.2009.11.009)] [Medline: [20189763](https://pubmed.ncbi.nlm.nih.gov/20189763/)]
65. Peirce CS. The numerical measure of the success of predictions. *Science*. Nov 14, 1884;4(93):453-454. [doi: [10.1126/science.ns-4.93.453-a](https://doi.org/10.1126/science.ns-4.93.453-a)] [Medline: [17795531](https://pubmed.ncbi.nlm.nih.gov/17795531/)]
66. Dufour R, Mardekian J, Pasquale MK, Schaaf D, Andrews GA, Patel NC. Understanding predictors of opioid abuse: predictive model development and validation. *Am J Pharm Benefits*. 2014;6(5):208-216. URL: [https://ajmc.s3.amazonaws.com/media/pdf/AJPB\\_09to10\\_Dufour\\_has\\_eApx\\_208to216.pdf](https://ajmc.s3.amazonaws.com/media/pdf/AJPB_09to10_Dufour_has_eApx_208to216.pdf) [Accessed 2026-02-13]
67. Webster LR, Webster RM. Predicting aberrant behaviors in opioid-treated patients: preliminary validation of the opioid risk tool. *Pain Med*. 2005;6(6):432-442. [doi: [10.1111/j.1526-4637.2005.00072.x](https://doi.org/10.1111/j.1526-4637.2005.00072.x)] [Medline: [16336480](https://pubmed.ncbi.nlm.nih.gov/16336480/)]
68. Rice JB, White AG, Birnbaum HG, Schiller M, Brown DA, Roland CL. A model to identify patients at risk for prescription opioid abuse, dependence, and misuse. *Pain Med*. Sep 2012;13(9):1162-1173. [doi: [10.1111/j.1526-4637.2012.01450.x](https://doi.org/10.1111/j.1526-4637.2012.01450.x)] [Medline: [22845054](https://pubmed.ncbi.nlm.nih.gov/22845054/)]
69. Hylan TR, Von Korff M, Saunders K, et al. Automated prediction of risk for problem opioid use in a primary care setting. *J Pain*. Apr 2015;16(4):380-387. [doi: [10.1016/j.jpain.2015.01.011](https://doi.org/10.1016/j.jpain.2015.01.011)] [Medline: [25640294](https://pubmed.ncbi.nlm.nih.gov/25640294/)]
70. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol*. Jul 20, 2016;34(21):2534-2540. [doi: [10.1200/JCO.2015.65.5654](https://doi.org/10.1200/JCO.2015.65.5654)] [Medline: [27247223](https://pubmed.ncbi.nlm.nih.gov/27247223/)]
71. Nguyen K, Wilson DL, Diiulio J, et al. Design and development of a machine-learning-driven opioid overdose risk prediction tool integrated in electronic health records in primary care settings. *Bioelectron Med*. Oct 18, 2024;10(1):24. [doi: [10.1186/s42234-024-00156-3](https://doi.org/10.1186/s42234-024-00156-3)] [Medline: [39420438](https://pubmed.ncbi.nlm.nih.gov/39420438/)]
72. Hong JWJ, Wilson DL, Nguyen K, et al. Protocol for a single-arm pilot clinical trial: developing and evaluating a machine learning opioid prediction & risk-stratification E-platform (DEMONSTRATE). *J Clin Med*. Dec 1, 2025;14(23):8522. [doi: [10.3390/jcm14238522](https://doi.org/10.3390/jcm14238522)] [Medline: [41375825](https://pubmed.ncbi.nlm.nih.gov/41375825/)]
73. Incident-OUD-project. GitHub. URL: <https://github.com/faysal22/Incident-OUD-Project> [Accessed 2026-01-24]

## Abbreviations

**CDS:** clinical decision support

**EHR:** electronic health record

**EN:** elastic net

**FNR:** false negative rate

**FPR:** false positive rate

**GBM:** gradient boosting machine

**ICD-10:** *International Statistical Classification of Diseases and Related Health Problems 10th Revision*

**LASSO:** least absolute shrinkage and selection operator

**ML:** machine learning

**NNE:** number needed to evaluate

**NPV:** negative predictive value

**OD:** opioid use disorder

**PPV:** positive predictive value

**RF:** random forest

**SHAP:** Shapley additive explanations

**TRIPOD+AI:** Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis+Artificial Intelligence

**UF:** University of Florida

**UPMC:** University of Pittsburgh Medical Center

*Edited by Javad Sarvestan; peer-reviewed by Jiafeng Song, Mohammad Amin Roshani; submitted 22.Jun.2025; final revised version received 13.Dec.2025; accepted 15.Dec.2025; published 05.Mar.2026*

*Please cite as:*

*Faysal JA, Lo-Ciganic W, Gellad WF, Wu Y, Harle CA, Nguyen K, Huang JL, Cochran G, Wilson DL, Staras SAS, Schmidt SOF, Rosenberg EI, Nelson D, Yan S, Reisfield GM, Greene WM, Kuza C, Hasan MM*

*Developing and Validating a Machine Learning Algorithm to Predict the Risk of Incident Opioid Use Disorder Among OneFlorida+ Patients: Prognostic Modeling Study*

*J Med Internet Res 2026;28:e79482*

URL: <https://www.jmir.org/2026/1/e79482>

doi: [10.2196/79482](https://doi.org/10.2196/79482)

© Jabed Al Faysal, Weihsuan Lo-Ciganic, Walid F Gellad, Yonghui Wu, Christopher A Harle, Khoa Nguyen, James L Huang, Gerald Cochran, Debbie L Wilson, Stephanie AS Staras, Siegfried OF Schmidt, Eric I Rosenberg, Danielle Nelson, Shunhua Yan, Gary M Reisfield, William M Greene, Courtney Kuza, Md Mahmudul Hasan. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 05.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.