

Developing and validating a machine-learning algorithm to predict opioid overdose in Medicaid beneficiaries in two US states: a prognostic modelling study

Wei-Hsuan Lo-Ciganic, Julie M Donohue, Qingnan Yang, James L Huang, Ching-Yuan Chang, Jeremy C Weiss, Jingchuan Guo, Hao H Zhang, Gerald Cochran, Adam J Gordon, Daniel C Malone, Chian K Kwoh, Debbie L Wilson, Courtney C Kuza, Walid F Gellad



Summary

Background Little is known about whether machine-learning algorithms developed to predict opioid overdose using earlier years and from a single state will perform as well when applied to other populations. We aimed to develop a machine-learning algorithm to predict 3-month risk of opioid overdose using Pennsylvania Medicaid data and externally validated it in two data sources (ie, later years of Pennsylvania Medicaid data and data from a different state).

Methods This prognostic modelling study developed and validated a machine-learning algorithm to predict overdose in Medicaid beneficiaries with one or more opioid prescription in Pennsylvania and Arizona, USA. To predict risk of hospital or emergency department visits for overdose in the subsequent 3 months, we measured 284 potential predictors from pharmaceutical and health-care encounter claims data in 3-month periods, starting 3 months before the first opioid prescription and continuing until loss to follow-up or study end. We developed and internally validated a gradient-boosting machine algorithm to predict overdose using 2013–16 Pennsylvania Medicaid data (n=639 693). We externally validated the model using (1) 2017–18 Pennsylvania Medicaid data (n=318 585) and (2) 2015–17 Arizona Medicaid data (n=391 959). We reported several prediction performance metrics (eg, C-statistic, positive predictive value). Beneficiaries were stratified into risk-score subgroups to support clinical use.

Findings A total of 8641 (1·35%) 2013–16 Pennsylvania Medicaid beneficiaries, 2705 (0·85%) 2017–18 Pennsylvania Medicaid beneficiaries, and 2410 (0·61%) 2015–17 Arizona beneficiaries had one or more overdose during the study period. C-statistics for the algorithm predicting 3-month overdoses developed from the 2013–16 Pennsylvania training dataset and validated on the 2013–16 Pennsylvania internal validation dataset, 2017–18 Pennsylvania external validation dataset, and 2015–17 Arizona external validation dataset were 0·841 (95% CI 0·835–0·847), 0·828 (0·822–0·834), and 0·817 (0·807–0·826), respectively. In external validation datasets, 71 361 (22·4%) of 318 585 2017–18 Pennsylvania beneficiaries were in high-risk subgroups (positive predictive value of 0·38–4·08%; capturing 73% of overdoses in the subsequent 3 months) and 40 041 (10%) of 391 959 2015–17 Arizona beneficiaries were in high-risk subgroups (positive predictive value of 0·19–1·97%; capturing 55% of overdoses). Lower risk subgroups in both validation datasets had few individuals ($\leq 0\cdot 2\%$) with an overdose.

Interpretation A machine-learning algorithm predicting opioid overdose derived from Pennsylvania Medicaid data performed well in external validation with more recent Pennsylvania data and with Arizona Medicaid data. The algorithm might be valuable for overdose risk prediction and stratification in Medicaid beneficiaries.

Funding National Institute of Health, National Institute on Drug Abuse, National Institute on Aging.

Copyright 2022 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Introduction

The USA continues to grapple with an opioid epidemic, with an estimated 75 673 opioid overdose deaths in the 12-month period ending in April 2021.¹ Health systems, payers, and policy makers have implemented various policies and programmes to mitigate the crisis. The President's Commission on Combating Drug Addiction and the Opioid Crisis in 2017 recommended applying advanced data analytics to improve identification of individuals at high risk of opioid overdose (hereafter overdose).² Within the last 5 years, studies identified

shortcomings of current opioid risk prediction tools and called for the development of more advanced models to improve identification of individuals at risk (or at no risk) of overdose.²

Our previous work showed that machine-learning approaches can improve risk prediction and stratification for incident opioid use disorder and subsequent overdose in Medicare beneficiaries.^{3,4} Medicaid is one of the largest US health-care payers. Medicaid beneficiaries have low incomes, substantial physical and mental comorbidities, and more often suffer from substance use disorders,

Lancet Digit Health 2022; 4: e455–65

See [Comment](#) page e403

Department of Pharmaceutical Outcomes and Policy (W-H Lo-Ciganic PhD, J L Huang PhD, C-Y Chang MS, J Guo PhD, D L Wilson PhD) and Center for Drug Evaluation and Safety (W-H Lo-Ciganic, J Guo), College of Pharmacy, University of Florida, Gainesville, FL, USA; Department of Health Policy and Management, Graduate School of Public Health (Prof J M Donohue PhD), Center for Pharmaceutical Policy and Prescribing, Health Policy Institute (J Guo, Prof J M Donohue, Q Yang MS, C C Kuza PhD, Prof W F Gellad MD), and Division of General Internal Medicine, School of Medicine (Prof W F Gellad), University of Pittsburgh, Pittsburgh, PA, USA; Heinz College of Information Systems and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA (J C Weiss PhD); Department of Mathematics (Prof H H Zhang PhD) and Division of Rheumatology, Department of Medicine, and the University of Arizona Arthritis Center (Prof C K Kwoh MD), University of Arizona, Tucson, AZ, USA; Program for Addiction Research, Clinical Care, Knowledge, and Advocacy, Division of Epidemiology, Department of Internal Medicine (G Cochran PhD, Prof A J Gordon MD) and Department of Pharmacotherapy, College of Pharmacy (Prof D C Malone PhD), University of Utah, Salt Lake City, UT, USA; Informatics, Decision-Enhancement, and Analytic Sciences Center, Veterans Affairs Salt Lake City Health Care System, Salt Lake City, UT, USA (Prof A J Gordon);

Center for Health Equity
Research Promotion, Veterans
Affairs Pittsburgh Healthcare
System, Pittsburgh, PA, USA
(Prof W F Gellad)

Correspondence to:
Dr Wei-Hsuan Lo-Giganic,
Department of Pharmaceutical
Outcomes and Policy, College of
Pharmacy, University of Florida,
Gainesville, FL 32610-0496, USA
wloigiganic@cop.ufl.edu

Research in context

Evidence before this study

In previous work, we developed machine learning approaches to improve risk prediction and stratification for development of opioid use disorder and subsequent overdose in Medicare beneficiaries. We are not aware of studies applying these methods in a Medicaid population, and then externally validating them. We thus searched PubMed for research articles, with no language restrictions, published from database inception up to June 1, 2021, using the following search terms: "Analgesics, Opioid / therapeutic use* AND Algorithms*" AND "Risk Assessment / methods*". After excluding eight studies that included narrow and specific patient populations (ie, opioid naive, paediatric, oncology, and anaesthesia patients), we identified two previous relevant studies that included Medicaid data for predicting the risk of opioid use disorder development within 1 or 5 years after initiating opioid prescriptions. Neither study examined overdose risk, nor did they include external validation.

Added value of this study

This study developed a machine learning algorithm capable of predicting 3-month risk of opioid overdose using Pennsylvania Medicaid claims data (2013–16) and validated it

in two data sources: more recent years of Medicaid data from Pennsylvania (2017–18) and in claims data from Arizona's Medicaid programme (2015–17). This work found that the algorithm was robust at predicting Medicaid beneficiaries' 3-month risk of opioid overdose, without major changes or adaptations to the model, despite the difference in calendar years and the different characteristics of the US states. Our prediction algorithm has the potential to be applied to Medicaid populations in other states and could be useful to guide clinical decisions and target interventions based on the degree of individual's risk.

Implications of all the available evidence

Machine-learning algorithms that more accurately predict patients' risks for opioid overdose over short time intervals and that provide better risk stratification than currently used tools can be valuable tools in data-informed decisions regarding the allocation of interventions and resources. The model derived from using historical data in one large state was scalable to more recent data and data from another state and provides an opportunity to improve on existing Medicaid programme strategies for addressing opioid overdose risk.

placing them at greater risk of opioid misuse and overdose. However, few studies have developed prediction algorithms to identify Medicaid beneficiaries at high risk of overdose.^{5–7} Furthermore, little is known about whether prediction algorithms developed using earlier years from one US state will perform as well using more recent data or when applied to another state.⁸

To fill these knowledge gaps, we studied two state Medicaid programmes, Pennsylvania and Arizona, with different population characteristics and overdose rates.⁹ In 2018, Pennsylvania ranked fourth and Arizona ranked 21st in US drug overdose mortality.⁹ We first used Pennsylvania Medicaid claims data from 2013–16 to develop a machine-learning algorithm to predict overdose in the subsequent 3 months. Second, we externally validated our prediction algorithm using more recent years (ie, 2017–18) of Pennsylvania Medicaid data and externally validated it using 2015–17 Arizona Medicaid data.

Methods

Study design and data

In this prognostic modelling study we used administrative claims data of Medicaid beneficiaries in Pennsylvania from Jan 1, 2013, to Dec 31, 2016, to develop a machine learning algorithm for overdose prediction. To evaluate whether the algorithm performs well using recent years of data and whether it translates to another state Medicaid programme, we conducted two validations using: (1) Pennsylvania Medicaid data from Jan 1, 2017, to Dec 31, 2018 (ie, the 2017–18 Pennsylvania external

validation dataset), and (2) Arizona Medicaid data from Jan 1, 2015, to December 31, 2017 (ie, the 2015–17 Arizona external validation dataset). Pennsylvania and Arizona have different geographical locations, population characteristics, and overdose rates with different drugs involved.¹⁰ Pennsylvania Medicaid ranks as the fourth largest of the 50 states in Medicaid expenditures¹¹ and fifth in enrolment (approximately 3 million beneficiaries annually),¹⁰ while Arizona Medicaid ranks 14th in total expenditures¹¹ and 12th in enrolment (approximately 1.7 million beneficiaries annually). Pennsylvania and Arizona implemented Medicaid eligibility expansion from the Affordable Care Act in different years (Arizona on Jan 1, 2014; Pennsylvania on Jan 1, 2015).

The Pennsylvania and the Arizona Medicaid datasets captured demographic information, eligibility, and enrolment information, outpatient, inpatient, and professional services, and prescription drugs. Prescription data contained all prescriptions reimbursed by Medicaid and included national drug codes, the dates of prescription fills, quantities dispensed, and days of supply. The prescriber information (eg, specialty) was available in the Pennsylvania Medicaid dataset but unavailable for the Arizona Medicaid dataset. We were able to link Arizona Medicaid data with death certificates data, allowing identification of fatal opioid overdoses not present in Medicaid claims, but not for the Pennsylvania dataset. We constructed study cohorts and created candidate predictors for the 2017–18 Pennsylvania and 2015–17 Arizona datasets as we did for the 2013–16

Pennsylvania dataset. We then applied the prediction algorithm to the external datasets. The study complied with Standards for Reporting of Diagnostic Accuracy and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognostic or Diagnosis reporting guidelines (appendix pp 25–28).^{12,13} It was approved by University of Pittsburgh and University of Florida Institutional Review Boards (human research ethics committees).

We identified Medicaid beneficiaries aged 18–64 years who filled one or more opioid prescription (excluding buprenorphine formulations approved for pain by the US Food and Drug Administration). An index date was defined as the date of an individual's first opioid prescription during the study period. We excluded beneficiaries who: (1) only filled parenteral opioids or cough or cold medications containing opioids, or both; (2) had malignant cancer diagnoses;³ (3) received hospice care; (4) were dually eligible for Medicare or enrolled in special Medicaid programmes that we were unable to completely observe their health services or prescription drug use; (5) had the first opioid prescription during the study period's last 3 months, or did not have at least a 3-month look-back period before their first opioid prescription, to allow measuring candidate predictors; or (6) were not enrolled for 3 months after the first opioid fill (appendix p 9). Beneficiaries remained in the cohort once eligible, regardless of whether they continued to receive opioids or had an occurrence of overdose, until they died or disenrolled from Medicaid. The same criteria were applied to datasets from both states.

We used International Classification of Diseases codes (ICD) versions 9 and 10 (appendix p 3) to identify any occurrence of opioid overdose (including prescription opioids and heroin) from inpatient or emergency department visits in 3-month periods from the first index opioid prescription.³ Overdose was defined as an opioid overdose code as the primary diagnosis, or other drug overdose or substance use disorders as the primary diagnosis (appendix p 4) with opioid overdose as a non-primary diagnosis.³

To be consistent with previous literature and quarterly evaluation periods commonly used by prescription drug monitoring programmes and health plans, we chose 3 months for the predictors' and outcomes' measurement windows.^{14,15} Candidate predictors of overdose ($n=284$) included sociodemographics, patient health status, use patterns of opioid and other non-opioid prescriptions, and provider-level and regional-level factors measured at baseline (during the 3-month period before the first opioid fill) and in 3-month windows after initiating prescription opioids (appendix pp 5–6). We updated the predictors measured in each 3-month period to account for changes over time for predicting overdose risks in each subsequent period (appendix p 10). This time-updating approach mimics active surveillance health systems might adopt.³

Statistical analysis

Our machine learning analysis using 2013–16 Pennsylvania Medicaid data comprised two steps: (1) developing a prediction model and creating overdose risk prediction scores for each individual, and (2) risk stratifying individuals into subgroups with similar overdose risks. We conducted external validation of the developed prediction algorithm using the 2017–18 Pennsylvania and the 2015–17 Arizona Medicaid datasets, respectively.

First, we randomly and equally divided the 2013–16 Pennsylvania Medicaid cohort into training (developing algorithms), testing (refining algorithms), and internal validation (evaluating algorithms' prediction performance) datasets. We compared beneficiaries' characteristics in training, testing, and internal validation datasets using two-tailed Student's *t* test, χ^2 test, and analysis of variance, or corresponding non-parametric tests. We applied several commonly used methods such as multivariate logistic regression, penalised regression, random forests, and gradient-boosting machine (GBM) to develop and test overdose prediction algorithms. Consistent with previous studies,^{3,5} GBM yielded the best prediction results (C-statistic of 0·841 for GBM vs up to 0·820 for other methods; appendix p 11) with an ability to handle complex interactions between predictors and outcomes. The study's objective was to externally validate the best-performing algorithm; thus, we focused on reporting the GBM model (appendix pp 1–2). Using the internal validation dataset of the 2013–16 Pennsylvania Medicaid cohort to assess the prediction algorithm's discrimination performance (ie, the extent to which predicted high-risk patients exhibit higher overdose rates compared with those predicted as low risk), we report C-statistics (0·700–0·800=good; >0·800=very good)¹⁶ and precision-recall curves.¹⁷ However, since C-statistics do not account for outcome prevalence information, which is important given the rarity of overdose events, we also report sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio, number needed to evaluate to identify one overdose, and estimated rate of alerts (appendix pp 12–13).^{18,19} Given that beneficiaries could have multiple 3-month periods until occurrence of a censored event (disenrolment or death), we presented episode-level performance as the main result. We conducted sensitivity analyses that iteratively and randomly selected patient-level random subsets from the internal validation data to ensure consistency of prediction performance.

No single prediction probability threshold to define high risk suits every purpose, as it is determined by the outcome's risk and benefit profile, type of interventions, and resource availability. Therefore, our main analysis classified validation dataset beneficiaries into subgroups using decile thresholds of predicted overdose risk scores from the training algorithm (ie, fixed thresholds) to allow comparison of risk profiles in different validation

See Online for appendix

	2013–16 Pennsylvania Medicaid (n=639 693)			External validation datasets	
	Training (n=213 231)	Testing (n=213 231)	Internal validation (n=213 231)	2017–18 Pennsylvania Medicaid (n=318 585)	2015–17 Arizona Medicaid (n=391 959)
Had ≥1 opioid overdose episode	2894 (1.4%)	2899 (1.4%)	2848 (1.3%)	2705 (0.8%)	2410 (0.6%)
Mean age, years	36.3 (12.2)	36.3 (12.3)	36.2 (12.2)	39.2 (12.5)	37.3 (12.8)
Age group, years					
18–30	87 767 (41.2%)	88 033 (41.3%)	88 294 (41.4%)	101 438 (31.8%)	150 595 (38.4%)
31–40	52 403 (24.6%)	51 767 (24.3%)	52 074 (24.4%)	81 832 (25.7%)	94 507 (24.1%)
41–50	38 348 (18.0%)	38 513 (18.1%)	38 390 (18.0%)	63 958 (20.1%)	71 377 (18.2%)
51–64	34 499 (16.2%)	34 647 (16.2%)	34 246 (16.1%)	71 357 (22.4%)	75 480 (19.3%)
Sex					
Female	141 207 (66.2%)	141 084 (66.2%)	141 166 (66.2%)	209 397 (65.7%)	248 551 (63.4%)
Male	72 024 (33.8%)	72 147 (33.8%)	72 065 (33.8%)	109 188 (34.3%)	143 408 (36.6%)
Race					
White	129 585 (60.8%)	129 668 (60.8%)	129 779 (60.9%)	198 166 (62.2%)	215 020 (54.9%)
Black	56 318 (26.4%)	56 264 (26.4%)	56 476 (26.5%)	79 482 (24.9%)	34 954 (8.9%)
Other or unknown	27 328 (12.8%)	27 299 (12.8%)	26 976 (12.7%)	40 937 (12.8%)	141 985 (36.2%)
Metropolitan residence	186 161 (87.3%)	186 316 (87.4%)	186 479 (87.5%)	276 640 (86.8%)	358 433 (91.4%)
Medicaid eligibility group at index					
Disabled	62 742 (29.4%)	62 609 (29.4%)	62 592 (29.4%)	70 025 (22.0%)	22 018 (5.6%)
Newly eligible	46 819 (22.0%)	47 034 (22.1%)	46 956 (22.0%)	136 536 (42.9%)	190 911 (48.7%)
Non-disabled adults	103 670 (48.6%)	103 588 (48.6%)	103 683 (48.6%)	112 024 (35.2%)	179 030 (45.7%)
Number of opioid fills	2.0 (1.7)	2.0 (1.7)	2.0 (1.7)	2.0 (1.7)	1.8 (1.5)
Average daily MME	37.1 (60.8)	37.0 (48.7)	37.0 (80.4)	37.7 (229.4)	39.3 (42.6)
Cumulative days of concurrent opioid and benzodiazepine use	3.0 (12.6)	2.9 (12.5)	3.0 (12.7)	3.9 (15.0)	2.7 (12.1)
Number of gabapentinoid fills	0.19 (0.7)	0.19 (0.7)	0.19 (0.7)	0.39 (1.1)	0.22 (0.8)
Number of hospitalisations	0.07 (0.3)	0.07 (0.4)	0.07 (0.4)	0.08 (0.3)	0.07 (0.4)
Number of emergency department visits	0.65 (1.2)	0.65 (1.2)	0.65 (1.3)	0.64 (1.3)	0.79 (1.4)
Opioid use disorder	9694 (4.5%)	9729 (4.6%)	9523 (4.5%)	18 730 (5.9%)	10 884 (2.8%)
Alcohol use disorder	5864 (2.8%)	5814 (2.7%)	5900 (2.8%)	9431 (3.0%)	11 554 (2.9%)
Anxiety disorders	26 579 (12.5%)	26 436 (12.4%)	26 368 (12.4%)	60 225 (18.9%)	47 132 (12.0%)
Mood disorders	38 809 (18.2%)	38 539 (18.1%)	38 325 (18.0%)	67 635 (21.2%)	48 132 (12.3%)

Data are n (%) or mean (SD). MME=morphine milligram equivalent.

Table 1: Selected characteristics of Pennsylvania and Arizona Medicaid beneficiaries

datasets. We further split the highest decile into three strata based on top first, second to fifth, and sixth to tenth percentiles to allow closer examination of patients at highest risk of experiencing an overdose. We thus created 12 risk subgroups. As an alternative, we conducted secondary analyses using decile risk score thresholds derived from each validation dataset to stratify beneficiaries into 12 risk subgroups. We created calibration plots (composed of 20 population bins of equal size) to examine the extent to which predicted overdose risks agree with observed risks by risk subgroup. We also present different thresholds along with other metrics at multiple levels of sensitivity and specificity (eg, arbitrarily choosing 90% sensitivity or a threshold with balanced sensitivity and specificity identified by the Youden Index as an anchor).²⁰

In external validation analyses, we applied the developed GBM algorithm to the 2017–18 Pennsylvania

and 2015–17 Arizona datasets. Because the original prediction algorithm included two prescriber-level variables, and we did not have prescriber information in the Arizona Medicaid dataset, the model automatically imputed these two prescriber-level variables for the Arizona dataset using the median values from the 2013–16 Pennsylvania Medicaid dataset. As done in the 2013–16 Pennsylvania Medicaid dataset, we evaluated prediction performance. We also used risk score thresholds derived from the training dataset to classify beneficiaries in these validation cohorts into 12 risk subgroups, the definition of which was based on the training dataset (2013–16 Pennsylvania) risk scores.

In secondary analyses, we first reported the top 25 important predictors from the GBM model. Second, we compared our prediction performance over a 12-month period with any of the opioid measures included in the Core Set of Adult Health Care Quality

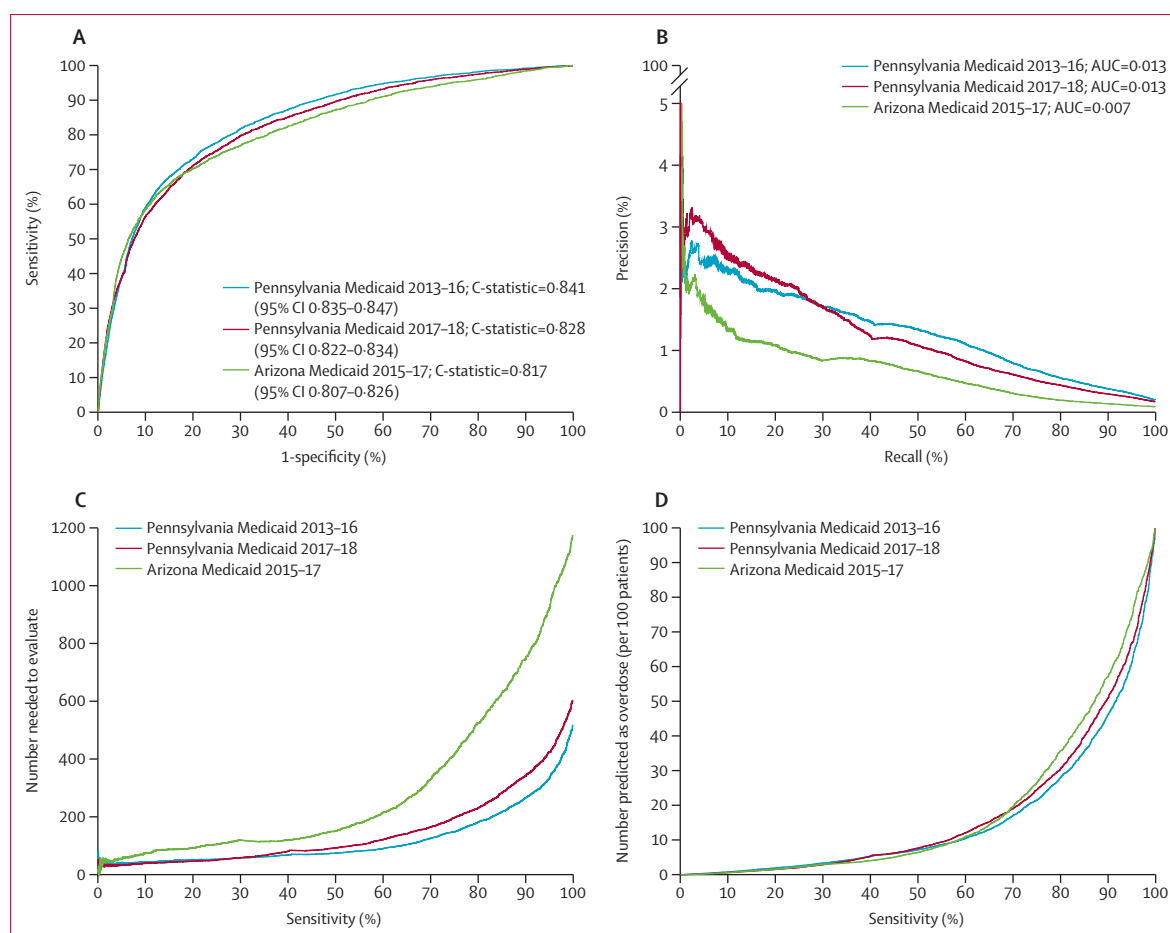


Figure 1: Performance matrix for predicting opioid overdose using GBM in Pennsylvania and Arizona Medicaid beneficiaries

(A) Areas under the receiver operating characteristic curves (or C-statistics). (B) Precision-recall curves (precision=positive predictive value and recall=sensitivity)—precision recall curves that are closer to the upper right corner or have a larger AUC than 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). Arizona Medicaid 2015–17=2015–17 Arizona external validation dataset (391 959 beneficiaries with 2 549 039 non-overdose episodes and 2172 overdose episodes). AUC=area under the curve. GBM=gradient boosting machine Pennsylvania Medicaid 2013–16=2013–16 Pennsylvania internal validation dataset (213 231 beneficiaries with 1745 919 non-overdose episodes and 3377 overdose episodes). Pennsylvania Medicaid 2017–18=2017–18 Pennsylvania external validation dataset (318 585 beneficiaries with 1825 672 non-overdose episodes and 3032 overdose episodes).

Measures for Medicaid²¹ or Medicaid Section 1115 Substance Use Disorder Demonstrations (hereafter Medicaid opioid measures) used by US states to identify high-risk individuals or substance use behaviour in Medicaid. These include three metrics: high-dose use defined as more than 120 morphine milligram equivalent for 90 continuous days or longer; four or more opioid prescribers and four or more pharmacies; and concurrent opioid and benzodiazepine use for 30 days or longer. Third, we used Arizona Medicaid beneficiaries death certificate data to identify fatal opioid overdoses that did not receive medical attention to determine whether the prediction algorithm worked to predict fatal overdoses using International Classification of Diseases, Tenth Revision (ICD-10) underlying cause-of-death codes X42, X44, Y12, and Y14 for accidental and undetermined overdose and multiple cause-of-death codes T40.1

(heroin), T40.2 (natural and semisynthetic opioids), T40.3 (methadone), and T40.4 (synthetic opioids other than methadone).²²

Analyses were performed using SAS 9.4 and Salford Predictive Modeler software suite version 8.2.

Role of the funding source

The funder had no role in the study design and conduct; data collection, management, analysis, and interpretation; manuscript preparation, review, or approval; and decision to submit the manuscript for publication.

Results

Beneficiaries in training, testing, and internal validation datasets of the 2013–16 Pennsylvania dataset (n=213 231 in each dataset used for model development) had similar characteristics and outcome distributions (table 1).

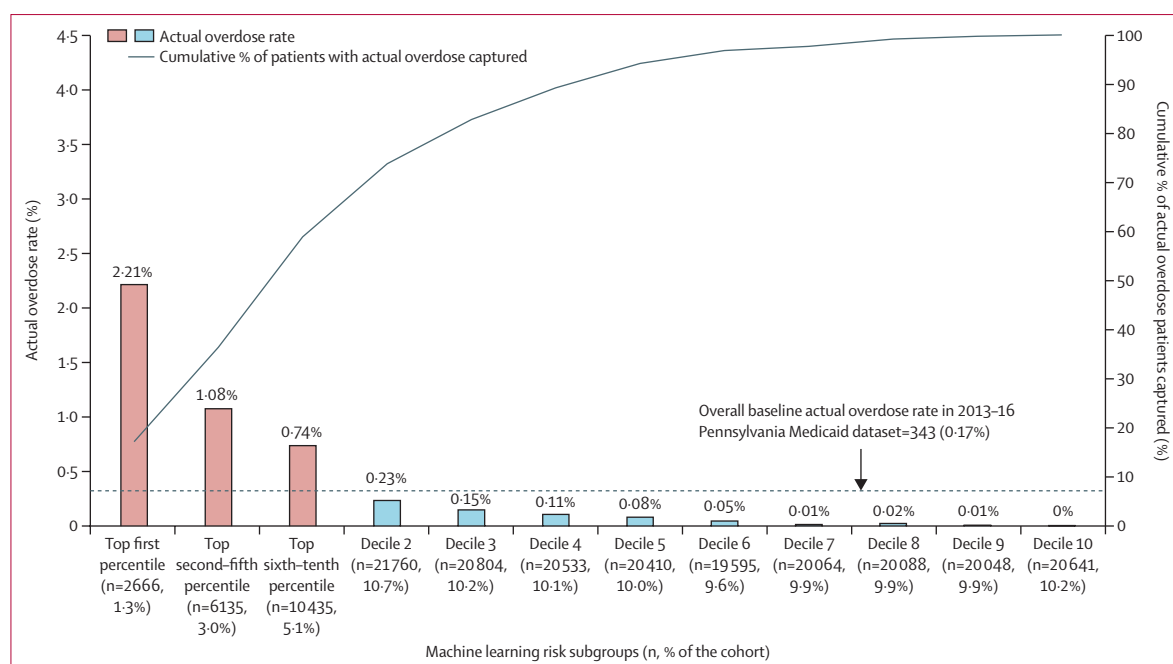


Figure 2: Opioid overdose identified by risk subgroup in the 2013–16 internal validation Pennsylvania Medicaid dataset (n=213 231) using GBM

Based on the individual's predicted probability of an opioid overdose (fatal or non-fatal) event, we classified 203 179 beneficiaries in the validation datasets into modified decile risk subgroups, with the highest decile further split into three additional strata based on the top first, second-fifth, and sixth-tenth percentiles to allow closer examination of beneficiaries at highest risk of experiencing an overdose. We used the thresholds of the risk scores derived from the 2013–16 Pennsylvania training dataset to identify a beneficiary's risk subgroup: top first percentile (≥ 98.3); second-fifth percentile ($96.6 \leq \text{risk score} < 98.3$); sixth-tenth percentile ($64.9 \leq \text{risk score} < 96.6$); decile 2 ($47.6 \leq \text{risk score} < 64.9$); decile 3 ($38.4 \leq \text{risk score} < 47.6$); decile 4 ($32.2 \leq \text{risk score} < 38.4$); decile 5 ($27.5 \leq \text{risk score} < 32.2$); decile 6 ($23.8 \leq \text{risk score} < 27.5$); decile 7 ($20.4 \leq \text{risk score} < 23.8$); decile 8 ($18.8 \leq \text{risk score} < 20.4$); decile 9 ($14.2 \leq \text{risk score} < 18.8$); decile 10 ($14.2 < \text{risk score}$). GBM=gradient boosting machine.

Compared with Pennsylvania Medicaid beneficiaries in the 2013–16 dataset, Pennsylvania Medicaid beneficiaries in the 2017–18 external validation dataset (n=318 585) were more likely to be older (39.2 vs 36.2–36.3 years) and newly eligible for Medicaid (42.9% vs 22.0–22.1%) and to have opioid use disorder diagnoses (5.9% vs 4.5–4.6%), whereas Medicaid beneficiaries in the Arizona dataset (n=391 959) were more likely to be of other or unknown race (36.2% vs 12.7–12.8%) and less likely to have opioid use disorder diagnoses (2.8% vs 4.5–4.6%). Rates of one or more opioid-overdose episodes during the study period were lower in the 2017–18 Pennsylvania (0.8%) and 2015–17 Arizona (0.6%) validation datasets, compared with the 2013–16 Pennsylvania algorithm-development dataset (1.3–1.4%).

Figure 1 summarises four prediction performance measures for the GBM models using the internal validation dataset from the 2013–16 Pennsylvania dataset and the external validation datasets. At the episode level, the prediction algorithm performed well in all three validation datasets (C-statistic of 0.841 [95% CI 0.835–0.847] for the 2013–16 Pennsylvania dataset, 0.828 [0.822–0.834] for the 2017–18 Pennsylvania dataset, and 0.817 [0.807–0.826] for the 2015–17 Arizona dataset; figure 1A). The improved performance in the internal-validation and external-validation Pennsylvania datasets compared with the

Arizona Medicaid dataset in precision-recall curves (figure 1B), number needed to evaluate (figure 1C), and positive alerts per 100 beneficiaries (figure 1D) were mainly driven by higher opioid overdose rates in Pennsylvania (table 1).

Prediction performance measures by varying sensitivity and specificity levels (90–100%) are shown in the appendix (pp 7–8). In the 2013–16 Pennsylvania internal validation dataset, at the balanced threshold using the Youden index, the GBM model had a 75.3% sensitivity, 78.5% specificity, 0.7% positive predictive value, 99.9% negative predictive value, number needed to evaluate of 149 to identify one opioid overdose, and approximately 22 positive alerts per 100 beneficiaries. In the 2017–18 Pennsylvania external validation dataset, at the balanced threshold, the GBM model had a 71.4% sensitivity, 79.9% specificity, 0.6% positive predictive value, 99.9% negative predictive value, number needed to evaluate of 171 to identify one opioid overdose, and approximately 20 positive alerts per 100 beneficiaries. In the 2015–17 Arizona external validation dataset, at the balanced threshold, the GBM model had a 67.2% sensitivity, 84.0% specificity, 0.4% positive predictive value, 100% negative predictive value, number needed to evaluate of 281 to identify one opioid overdose, and approximately 16.1 positive alerts per 100 beneficiaries.

Sensitivity analyses using randomly and iteratively selected patient-level data overall yielded similar results as using episode-level data (appendix pp 14–15).

Figure 2 depicts the actual overdose rate for individuals in the internal validation dataset using risk score thresholds derived from the 2013–16 Pennsylvania

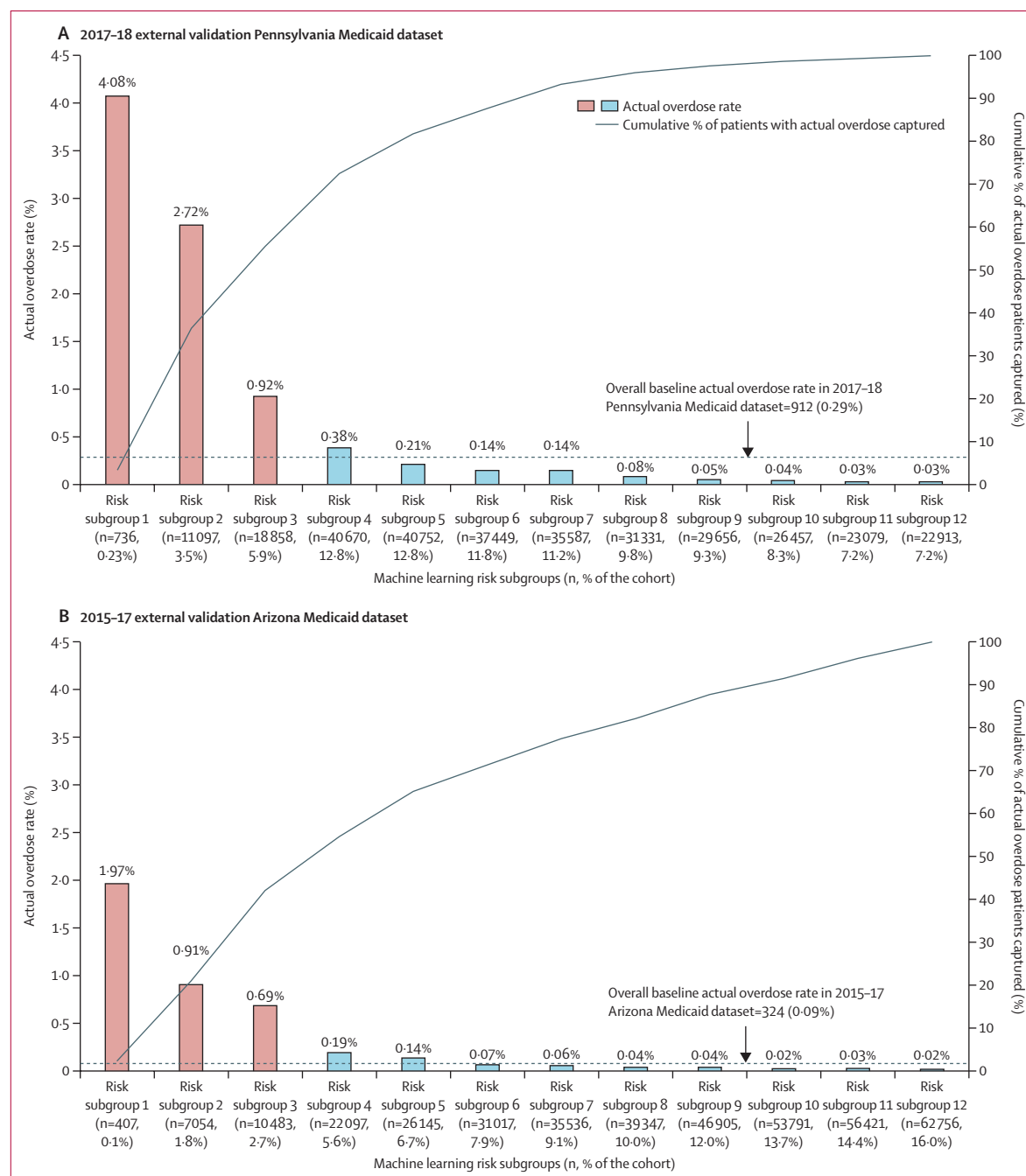


Figure 3: Opioid overdose identified by risk subgroup in the 2017–18 Pennsylvania (n=318 585) and 2015–17 Arizona Medicaid (n=391 959) external validation datasets using GBM

Based on the individual's predicted probability of an opioid overdose (fatal or non-fatal) event, we classified beneficiaries in the two validation datasets into risk subgroups using the modified decile thresholds of the risk scores derived from the 2013–16 Pennsylvania training dataset, with the highest risk decile further split into three additional strata based on the top first, second-fifth, and sixth-tenth percentiles to allow closer examination of beneficiaries at highest risk of experiencing an overdose. The thresholds of the risk scores derived from the 2013–16 Pennsylvania training dataset to identify a beneficiary's risk subgroup are: top first percentile (≥ 98.3); second-fifth percentile ($96.6 \leq \text{risk score} < 98.3$); sixth-tenth percentile ($64.9 \leq \text{risk score} < 96.6$); decile 2 ($47.6 \leq \text{risk score} < 64.9$); decile 3 ($38.4 \leq \text{risk score} < 47.6$); decile 4 ($32.2 \leq \text{risk score} < 38.4$); decile 5 ($27.5 \leq \text{risk score} < 32.2$); decile 6 ($23.8 \leq \text{risk score} < 27.5$); decile 7 ($20.4 \leq \text{risk score} < 23.8$); decile 8 ($18.8 \leq \text{risk score} < 20.4$); decile 9 ($14.2 \leq \text{risk score} < 18.8$); decile 10 ($14.2 < \text{risk score}$). GBM=gradient boosting machine.

	Any Medicaid core set opioid measure†		High risk in GBM using different thresholds‡		
	Low risk (n=122 538, 90.7%)	High risk (n=12 568, 9.3%)	Top first percentile (n=4570, 3.4%)	Top fifth percentile (n=11 053, 8.2%)	Top tenth percentile (n=23 158, 17.1%)
Number of actual overdoses (% of each subgroup)	639 (0.5%)	204 (1.6%)	299 (6.5%)	557 (5.0%)	713 (3.1%)
Number of actual non-overdoses (% of each subgroup)	121 899 (99.5%)	12 364 (98.4%)	4271 (93.5%)	10 496 (95.0%)	22 445 (96.9%)
Number needed to evaluate	NA	62	15	19	32
% of all overdoses over 12 months (n=843) captured	75.8%	24.2%	35.5%	66.1%	84.6%

MME=morphine milligram equivalent. GBM=gradient boosting machine. *To compare with Medicaid opioid measures, beneficiaries were required to have at least a 12-month period of follow-up and the resulting sample size was smaller than the sample size in the main analysis. †If classifying beneficiaries with any of the Medicaid high-risk opioid use measures as opioid overdose, those remaining would be considered as non-overdose. ‡The Medicaid opioid quality measures included in the Core Set of Adult Health Care Quality Measures for Medicaid or Medicaid Section 1115 Substance Use Disorder Demonstrations to identify high-risk individuals or use behaviour in Medicaid. These simple measures include three metrics: (1) high-dose use, defined as >120 MME for ≥90 continuous days, (2) ≥4 opioid prescribers and ≥4 pharmacies, and (3) concurrent opioid and benzodiazepine use ≥30 days. §For GBM, we presented high-risk groups using different cutoff thresholds of prediction probability: individuals with (1) predicted score in the top first percentile (≥98.3); (2) predicted probability in the top fifth percentile (≥96.6); and (3) predicted probability in the top tenth percentile (≥64.9). If classifying beneficiaries in the high-risk group of opioid overdoses, those remaining would be considered as non-overdose.

Table 2: Comparison of prediction performance using any of the Medicaid opioid quality measures versus GBM in the 2013–16 Pennsylvania internal validation sample (n=135 106) over a 12-month period*

training dataset. The highest-risk subgroup (risk scores in the top 1st percentile; 1.3% [n=2666]) had a positive predictive value of 2.2%, a negative predictive value of 97.8%, and number needed to evaluate of 45. Among 343 individuals with overdose in the 2013–16 Pennsylvania internal validation dataset, 253 (73.8%) individuals were in the top two deciles of risk scores (shown in figure 2 as the top four risk subgroups). The 3rd–10th decile subgroups had minimal overdose rates (0–15 per 10 000). In external validation analyses (figure 3), the overall baseline overdose rate in 3-month windows was higher in the 2017–18 than in the 2013–16 Pennsylvania Medicaid dataset (0.29% vs 0.17%). The overall baseline 3-month overdose rate was also lower in the 2015–17 Arizona external validation dataset (0.09%). In the 2017–18 Pennsylvania external validation dataset (figure 3A), the highest-risk subgroup (or risk subgroup 1, 0.23% [n=736]) had a positive predictive value of 4.1%, a negative predictive value of 95.9%, and number needed to evaluate of 25. Of 912 individuals with overdose in the 2017–18 Pennsylvania external validation dataset, 661 (72.5%) individuals were in the top four risk subgroups. The fifth to 12th risk subgroups had minimal overdose rates (ranging from 0.03% to 0.21%). In the 2015–17 Arizona external validation dataset (figure 3B), the highest risk subgroup (risk subgroup 1, 0.10% [n=407]) had a positive predictive value of 1.97%, a negative predictive value of 98.0%, and number needed to evaluate of 51. Of 342 individuals with overdose in the 2015–17 Arizona external validation dataset, 187 (54.7%) individuals were in the top four risk subgroups. Consistent with the Pennsylvania external validation dataset, the fifth to 12th risk subgroups had minimal overdose rates (2–14 per 10 000). Similar magnitudes were found using risk score thresholds derived from each corresponding validation dataset (appendix pp 16–17).

Additional calibration curves are shown in the appendix (pp 18–19).

The top 25 most important predictors identified by the GBM model, such as having a diagnosis of OUD, total number of emergency department visits, race, gender, and age, are shown in the appendix (p 20). Table 2 compared prediction performance with existing Medicaid opioid measures over a 12-month period in the 2013–16 Pennsylvania Medicaid dataset. Using existing Medicaid opioid measures for identifying high-risk individuals (9.3%) captured 24.2% of all actual overdose cases (number needed to evaluate of 62) over a 12-month period. When using the top fifth percentile of our risk scores to identify high risk, our GBM algorithm captured 66.1% of all actual opioid overdose cases (number needed to evaluate of 19), despite there being a similar number of high-risk individuals identified in both models. Our analysis predicting fatal overdose in the Arizona external validation dataset (84% were captured in claims data) yielded similar findings (eg, C-statistic of 0.814; 95% CI 0.796–0.831; appendix pp 21–22).

Discussion

We developed and externally validated a machine-learning algorithm with strong performance for predicting 3-month risk of opioid overdose in Medicaid beneficiaries. Our study shows that an opioid overdose prediction algorithm developed in one state's Medicaid programme can effectively translate to later time periods in the same US state and to a different state, addressing a major concern about the generalisability of opioid overdose prediction models. In addition, the algorithm represents an improvement on less accurate opioid risk measures currently tracked by state Medicaid programmes.^{21,23,24}

To our knowledge, this study is the first study predicting 3-month opioid overdose risk among Medicaid populations with external validation after initiation of prescription opioids. We only identified two previous studies that used Medicaid data to predict risk of opioid use disorder development within 1 or 5 years after initiating opioid prescriptions,^{6,7} but neither examined overdose risk, nor did they include external validation. Studies using advanced methods to more accurately identify individuals at risk of overdose are needed. This study expanded our previous work using machine-learning approaches to improve accuracy of predicting overdose in the subsequent 3 months in a large state Medicaid dataset and broaden applicability of these models across state Medicaid programmes.⁵ Our best-performing GBM has several advantages, including handling missing data automatically, no additional feature selection process required prior to the GBM modelling, greater flexibility in hyper parameter tuning to include complex interactions between predictors and outcomes, and often providing better performance compared with other approaches.^{3,5} We acknowledge, however, that the flexibility during model tuning can be time-consuming and computationally expensive.

Our Arizona Medicaid external validation analysis showed the feasibility of applying our prediction model to other state Medicaid programmes with very different population race and ethnicity, geography, and overdose rates compared with Pennsylvania Medicaid. When using fixed risk score thresholds identified from the Pennsylvania dataset, fewer Arizona Medicaid beneficiaries were classified into the top four high-risk subgroups (eg, 10% for the Arizona dataset vs 22% for the 2017–18 Pennsylvania dataset) with 55% of all overdoses captured. Using various risk scores identified from each validation dataset, over 75% of overdoses were captured in the top three decile groups in all validation datasets (appendix pp 16–17). The model showed good performance across states and the potential clinical and policy use of different risk stratification approaches without any major change or adaptation of the models developed from a US state Medicaid programme. Future work should further validate the applicability of our model to other state Medicaid programmes. To maintain the model's generalisability, the model might need to be recalibrated in different states such as Midwest and South regions.

Although our model had good discrimination (C-statistics >0.80), rare overdose outcomes led to low positive predictive values that could increase false positives and overestimate the benefits or underestimate intervention costs and resources. In cases such as these, reporting additional measures, including sensitivity, number needed to evaluate, and estimated alert rate provides for more thorough evaluation of a clinical prediction model's performance.¹⁸ For example, number needed to evaluate is the number of patients necessary

to evaluate in order to detect one outcome using an early warning tool (ie, the machine-learning algorithm in our study) versus if no tools existed.¹⁸ When predicting a rare outcome like overdose, the number needed to evaluate closely estimates the number needed to treat or number needed to screen because very low baseline overdose rates will make the absolute risk reduction close to the post-screening incidence. Despite low positive predictive values, our number needed to evaluate (149–602 in the Pennsylvania datasets; 281–1170 in the Arizona dataset varying by different risk thresholds) using GBM algorithms is similar to the number needed to screen for commonly used cancer screening tests, such as annual mammography screening to prevent one breast cancer death (number needed to screen of 233–1316 varying by subgroups with different underlying risks).²⁵

Our machine learning model represents an advance on strategies that Medicaid programmes are using to predict overdose risk, allowing them to better target time sensitive or so-called just in time interventions. Our study showed how current rules (eg, Medicaid Core Set opioid measures) for identifying so-called high-risk patients might not accurately predict risk of overdose, compared with using different risk score thresholds (eg, top fifth percentile of risk scores) to identify high-risk patients. Although not perfect, our risk classification with 12 subgroups allows those implementing the algorithm to determine the risk threshold at which to intervene, based on costs and intensity of interventions and resource availability. Resource intensive and burdensome interventions (eg, pharmacy lock-in programmes) could be limited to the small number of individuals in the highest-risk subgroup. Lower cost, less burdensome, or less risky interventions (eg, naloxone distribution)²⁶ could be targeted towards more individuals in moderate to high risk subgroups.^{27,28} Nonetheless, additional screening and assessment are needed to avoid unintended consequences resulting from false positives.

Our study has limitations. First, we could not capture prescriptions paid out of pocket or patients with only illicit opioid use. For example, only 40% of fatal overdoses in Arizona were captured using our cohort definition (ie, having one or more opioid fills in the study period). Second, claims data do not capture overdoses that do not receive medical attention. However, we applied a previously validated algorithm using ICD codes to identify opioid-overdose events reported in medical claims (positive predictive value of 81–84%).²⁹ Our algorithm also did well in a sensitivity analysis using Arizona death certificates data to predict fatal overdose. Third, the Arizona dataset did not have prescriber information nor a separate variable for ethnicity, which could introduce misclassification biases. Fourth, we were unable to capture other potential predictors, including laboratory results in clinical data and sociobehavioural information that might improve the model. Fifth, our

prediction algorithm might not be generalisable to other populations or states. Sixth, as expected, the positive predictive value from our model was low due to overdose being a rare outcome. Nonetheless, given the serious consequences of overdose, our risk-stratified approach appeared to be effective in both Pennsylvania and Arizona. Finally, although our algorithm has many potential uses, there are key implementation barriers to overcome before implementation (eg, infrastructure to automatically generate risk scores or algorithm's usability and effectiveness due to data lags for claims data). Furthermore, the current algorithm included race and ethnicity and requires comprehensive bias evaluations to identify potential approaches for ensuring algorithm fairness to target interventions and provide health services equitably.

In conclusion, a machine-learning algorithm predicting opioid overdose derived from Pennsylvania Medicaid data performed well in external validation with data from more recent years and data from another state with different characteristics. The algorithm could be a valuable and feasible tool to predict and stratify risk of opioid overdose in Medicaid beneficiaries.

Contributors

All authors contributed to the conception and design of the study, analysis and interpretation of data, and critical revision of the Article for important intellectual content. All authors were responsible for the decision to submit the manuscript all approved the final version submitted. W-HL-C and WFG take responsibility for the integrity of the work as a whole, from inception to the finished Article. W-HL-C, QY, C-YC, JLH, JMD, and WFG had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors reviewed the data and results presented in this Article. JCW, HHZ, GC, AJG, DCM, CKK, and DLW could not access the original data directly because of limitations in the data use agreement. W-HL-C, QY, C-YC, JLH, JMD, and WFG verified the data. W-HL-C, JMD, JCW, HHZ, AJG, GC, DCM, CKK, and WFG conceived of and designed the study. W-HL-C, JMD, and WFG acquired data. W-HL-C, JMD, JLH, CKK, DLW, and WFG provided administrative, technical, and material support. W-HL-C and WFG supervised the study.

Declaration of interests

W-HL-C 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 the machine learning algorithm for opioid risk prediction in Medicare described in this Article. W-HL-C, WFG, DLW, and C-YC are recipients of a grant from the National Institute on Aging (NIA; R21 AG060308). W-HL-C, WFG, JMD, AJG, GC, JLH, CKK, DCM, QY, JW, and HHZ are recipients of a grant from the US National Institute on Drug Abuse (NIDA; R01DA044985). W-HL-C declares grants from the Richard King Mellon Foundation–University of Pittsburgh, University of Florida Clinical and Translational Science Institute, the US National Institute of Mental Health (1R03MH114503-01 and 1R01MH121907), Pharmaceutical Research and Manufacturers of America Foundation, NIDA (1R01DA050676-01A1 and R01DA044985), Veterans Affairs (VA) Merit 1 (101HX002191-01A2), and Merck, Sharp & Dohme and Bristol Myers Squibb. CKK declares grants from Lilly, Pfizer, GSK, Cumberland Pharmaceuticals, AbbVie, and EMD Serono; consultant fees from EMD Serono, Express Scripts, and Regeneron; an advisory board role with EMD Serono, Thusane, Regeneron, Taiwan Liposome Company, Amzell, LG Chem, and Novartis; payment or honoraria from Focus Communications and PRIME Education; and being on the Data and Safety Monitoring Committee for Kolon Tissue Gene and on the board of directors for the International Chinese Osteoarthritis Research Society. WFG declares grant or contract to his

institution from Richard King Mellon Foundation. JMD declares salary support from the Pennsylvania Department of Human Services, Richard King Mellon Foundation, and the US National Institute of Health (NIH)–NIDA (R01DA048019). AJG declares grants or contracts from NIH and the VA; royalties from UpToDate; and other financial and non-financial interests from American Society of Addiction Medicine, Association for Multidisciplinary Education and Research in Substance Use and Addiction, and the International Society of Addiction Journal Editors. JCW declares grants or contracts from Allegheny Health Network, Carnegie Mellon University, and University of Pittsburgh Medical Center; payment or honoraria from St Jude's; and receipt of equipment, materials, or other services from Amazon Web Services and Azure. DLW declares grant funding from Merck, Sharp & Dohme, and NIH–NIDA (1R01DA050676-01A1). All other authors declare no competing interests.

Data sharing

National Drug Codes can be provided by request to the corresponding author. Modelling codes can be provided by request to the corresponding author under proper use agreement. Access to the Pennsylvania Medicaid data was made possible through an intergovernmental agreement between the University of Pittsburgh and the Pennsylvania Department of Human Services. Access to the Arizona Medicaid data was made possible through an inter-institutional data use agreement between the University of Florida and the Center for Health Information and Research at Arizona State University. Per the data use agreements, the relevant limited datasets used in this study contained some patient health information variables (eg, dates of services) and thus cannot be made publicly available. This study was approved by the University of Pittsburgh and University of Florida Institutional Review Boards (IRBs; human research ethics committees). Patient consent was waived for this study due to the use of existing secondary data sources per the IRBs' policies.

Acknowledgments

This work was supported by the grant R01DA044985 from the NIH–NIDA, and grant R21 AG060308 from the NIH–NIA. 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 Pennsylvania Department of Human Services.

References

- Centers for Disease Control and Prevention, National Center for Health Statistics. Drug overdose deaths in the U.S. Top 100,000 annually. 2021. https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2021/20211117.htm#:~:text=Provisional%20data%20from%20CDC's%20National,same%20period%20the%20year%20before (accessed Apr 26, 2022).
- The President's Commission on Combating Drug Addiction and the Opioid Crisis. Final report draft. 2017. https://trumpwhitehouse.archives.gov/sites/whitehouse.gov/files/images/Final_Report_Draft_11-15-2017.pdf (accessed Sept 27, 2021).
- 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* 2019; 2: e190968.
- 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: e0235981.
- Lo-Ciganic WH, Donohue JM, Hulsey EG, et al. Integrating human services and criminal justice data with claims data to predict risk of opioid overdose among Medicaid beneficiaries: a machine-learning approach. *PLoS One* 2021; 16: e0248360.
- Reps JM, Cepeda MS, Ryan PB. Wisdom of the CROUD: development and validation of a patient-level prediction model for opioid use disorder using population-level claims data. *PLoS One* 2020; 15: e0228632.
- Hastings JS, Howison M, Inman SE. Predicting high-risk opioid prescriptions before they are given. *Proc Natl Acad Sci USA* 2020; 117: 1917–23.
- Centers for Disease Control and Prevention. Opioids: understanding the epidemic. 2021. <https://www.cdc.gov/opioids/basics/epidemic.html> (accessed Sept 27, 2021).

- 9 Kaiser Family Foundation. Opioid overdose death rates and all drug overdose death rates per 100,000 population (age-adjusted). 2018. [https://www.kff.org/other/state-indicator/opioid-overdose-death-rates/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Opioid%20Overdose%20Death%20Rate%20\(Age-Adjusted\)%22,%22sort%22:%22desc%22%7D](https://www.kff.org/other/state-indicator/opioid-overdose-death-rates/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Opioid%20Overdose%20Death%20Rate%20(Age-Adjusted)%22,%22sort%22:%22desc%22%7D) (accessed June 15, 2020).
- 10 Kaiser Family Foundation. Total monthly Medicaid and CHIP enrollment. 2018. <https://www.kff.org/health-reform/state-indicator/total-monthly-medicaid-and-chip-enrollment/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Total%20Monthly%20Medicaid%20CHIP%20Enrollment%22,%22sort%22:%22desc%22%7D> (accessed Feb 21, 2022).
- 11 Kaiser Family Foundation. Total Medicaid spending. 2018. <https://www.kff.org/medicaid/state-indicator/total-medicaid-spending/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Total%20Medicaid%20Spending%22,%22sort%22:%22desc%22%7D> (accessed Feb 21, 2022).
- 12 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015; **351**: h5527.
- 13 Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162**: W1–73.
- 14 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* 2009; **15**: 897–906.
- 15 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* 2015; **16**: 445–53.
- 16 Hosmer DWH, Lemeshow S. Assessing the fit of the model. In: Shewhart WA, Wilks SS, eds. *Applied logistic regression*, 2nd edn. Hoboken, NJ: John Wiley & Sons, 2000: 143–202.
- 17 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**: e0118432.
- 18 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* 2015; **19**: 285.
- 19 Tufféry S. *Data mining and statistics for decision making*. Hoboken, NJ: John Wiley & Sons, 2011.
- 20 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21**: 128–38.
- 21 Medicaid. 2020 core set of adult health care quality measures for Medicaid (adult core set). <https://www.medicaid.gov/medicaid/quality-of-care/downloads/performance-measurement/2020-adult-core-set.pdf> (accessed 17 June 17, 2020).
- 22 Chalasani R, Lo-Ciganic WH, Huang JL, et al. Occupational patterns of opioid-related overdose deaths among Arizona Medicaid enrollees, 2008–2017. *J Gen Intern Med* 2020; **35**: 2210–13.
- 23 Pharmacy Quality Alliance (PQA) Performance Measures. Use of opioids at high dosage and from multiple providers in persons without cancer. <http://pqaalliance.org/measures/default.asp> (accessed Jan 7, 2016).
- 24 Pharmacy Quality Alliance. Performance measures. Concurrent use of opioids and benzodiazepines. <http://pqaalliance.org/measures/default.asp> (accessed June 13, 2017).
- 25 Hendrick RE, Helvie MA. Mammography screening: a new estimate of number needed to screen to prevent one breast cancer death. *AJR Am J Roentgenol* 2012; **198**: 723–28.
- 26 Centers for Disease Control and Prevention. Evidence-based strategies for preventing opioid overdose: what's working in the United States. 2018. <http://www.cdc.gov/drugoverdose/pdf/pubs/2018-evidence-based-strategies.pdf> (accessed Oct 23, 2018).
- 27 US Congress. S.524—Comprehensive Addiction and Recovery Act of 2016 2016. <https://www.congress.gov/bill/114th-congress/senate-bill/524/text> (accessed July 20, 2016).
- 28 Roberts AW, Gellad WF, Skinner AC. Lock-in programs and the opioid epidemic: a call for evidence. *Am J Public Health* 2016; **106**: 1918–19.
- 29 Green CA, Perrin NA, Janoff SL, Campbell CI, Chilcoat HD, Coplan PM. Assessing the accuracy of opioid overdose and poisoning codes in diagnostic information from electronic health records, claims data, and death records. *Pharmacoepidemiol Drug Saf* 2017; **26**: 509–17.