

BMJ Open Quickly identifying people at risk of opioid use disorder in emergency departments: trade-offs between a machine learning approach and a simple EHR flag strategy

Izabela E Annis ¹, Robyn Jordan,² Kathleen C Thomas¹

To cite: Annis IE, Jordan R, Thomas KC. Quickly identifying people at risk of opioid use disorder in emergency departments: trade-offs between a machine learning approach and a simple EHR flag strategy. *BMJ Open* 2022;**12**:e059414. doi:10.1136/bmjopen-2021-059414

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-059414>).

Received 30 November 2021
Accepted 15 August 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Division of Pharmaceutical Outcomes and Policy, University of North Carolina at Chapel Hill Eshelman School of Pharmacy, Chapel Hill, North Carolina, USA

²Department of Psychiatry, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA

Correspondence to

Izabela E Annis;
izabela@unc.edu

ABSTRACT

Objectives Emergency departments (EDs) are an important point of contact for people with opioid use disorder (OUD). Universal screening for OUD is costly and often infeasible. Evidence on effective, selective screening is needed. We assessed the feasibility of using a risk factor-based machine learning model to identify OUD quickly among patients presenting in EDs.

Design/settings/participants In this cohort study, all ED visits between January 2016 and March 2018 for patients aged 12 years and older were identified from electronic health records (EHRs) data from a large university health system. First, logistic regression modelling was used to describe and elucidate the associations between patient demographic and clinical characteristics and diagnosis of OUD. Second, a Gradient Boosting Classifier was applied to develop a predictive model to identify patients at risk of OUD. The predictive performance of the Gradient Boosting algorithm was assessed using F1 scores and area under the curve (AUC).

Outcome The primary outcome was the diagnosis of OUD. **Results** Among 345 728 patient ED visits (mean (SD) patient age, 49.4 (21.0) years; 210 045 (60.8%) female), 1.16% had a diagnosis of OUD. Bivariate analyses indicated that history of OUD was the strongest predictor of current OUD (OR=13.4, CI: 11.8 to 15.1). When history of OUD was excluded in multivariate models, baseline use of medications for OUD (OR=3.4, CI: 2.9 to 4.0) and white race (OR=2.9, CI: 2.6 to 3.3) were the strongest predictors. The best Gradient Boosting model achieved an AUC of 0.71, accuracy of 0.96 but only 0.45 sensitivity.

Conclusions Patients who present at the ED with OUD are high-need patients who are typically smokers with psychiatric, chronic pain and substance use disorders. A machine learning model did not improve predictive ability. A quick review of a patient's EHR for history of OUD is an efficient strategy to identify those who are currently at greatest risk of OUD.

INTRODUCTION

The rate of opioid-related hospitalisation and emergency department (ED) visits between 2005 and 2014 increased twofold.¹ ED visits for opioid overdose rose across the USA by

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ Large cohort study analysing prevalence of opioid use disorder in emergency departments.
- ⇒ Multivariable analyses taking into consideration demographic information, diagnoses, procedures and prescription orders.
- ⇒ Use of advanced machine learning modelling in conjunction with remedial measures for outcome imbalance.
- ⇒ Detailed presentation of performance statistics, stratified by outcome, interpreted in the light of results' suitability for use in clinical practice.
- ⇒ Data contain records from only one healthcare system with possible variable misclassification due to missing medications and diagnoses from encounters outside of the studied health system.

30% between 2016 and 2017.² Risk from overdose due to opioid consumption has also increased. In one state, emergency medical services trips to the ED increased 17%.³ As a result, EDs are a crucial point of contact for people with opioid use disorder (OUD). In recent years, EDs have received more recognition as places where patients can be identified and connected to care for OUD.⁴⁻⁷ Nevertheless, evidence indicates that there is significant variation in screening and treatment of OUD in EDs.^{8,9} It is unknown how to best selectively identify patients with risk of OUD who visit EDs, in order to accomplish efficient screening and referral to treatment.

Use of previous diagnoses and procedures coded in existing patients' electronic health records (EHRs) may be an efficient way to identify those at risk during their ED visit. Several screening tools have been developed for patients with chronic pain to assess their risk of OUD,^{10,11} and numerous studies have investigated the relationship between individual patient-level factors and OUD in the



population.¹² Risk factors found to be associated with OUD include history of OUD or other substance use disorders, psychiatric disorders, concurrent prescriptions of selected psychiatric medications and certain social settings that foster illicit substance use.^{12 13}

Proliferation and wide availability of administrative EHR databases have propelled advances in predictive analytics, which yield accurate identification of patients at risk of acute outcomes and healthcare utilisation. Some of the efforts include prediction of hospital readmission,¹⁴ repeat ED visits¹⁵ and hospital admission at the time of ED triage.¹⁶ These successes point to the promise of using patients' historical EHR data to predict who is at risk of OUD among patients presenting in EDs. While the Diagnostic and Statistical Manual of Mental Disorders (DSM) and identification using Structured Clinical Interviews is considered as the 'gold standard' for diagnosis of OUD,¹⁷ a readily available alternative, the International Classification of Diseases (ICD) diagnostic codes, could be used in the development of a predictive algorithm. A reliable, predictive model that could identify patients at high risk of OUD could serve as a decision support tool for triage team members and alert them when a patient should be screened for OUD.⁴

The primary objectives of this study were to (1) describe the association between patient demographic and clinical characteristics and diagnosis of OUD among patients presenting at EDs, and (2) investigate the feasibility of constructing a valid machine learning predictive model of OUD in the ED in order to identify the best strategy for EDs to identify patients at high risk of OUD to facilitate targeted screening and diagnosis. Variable selection was informed by the Andersen Behavioural Model.¹⁸

METHODS

Data and study sample

We used administrative EHR data from a large university health system in North Carolina with seven EDs across separate counties, which included patient demographics, medical encounter details, procedures, diagnoses, vitals, laboratory results and prescription medication orders. The EDs are located in two 'mostly rural'¹⁹ counties (Chatham, Johnston) and five urban counties (Caldwell, Guilford, Henderson, Orange, Wake). The main academic hospital is a safety net hospital that serves patients from the entire state. Our cohort included all ED encounters that occurred during January 2016–March 2018, for patients aged 12 years or older, who had at least one clinical encounter during the baseline period. The baseline period was defined as 12 months prior to the ED encounter. Informed presence bias, the idea that inclusion and number of records in an EHR are associated with patients' health status and severity of illness, is a well-known problem affecting studies using EHRs.²⁰ Typically, analyses using administrative databases impose various inclusion/exclusion criteria, for example, continuous enrolment or a requirement to have at least one

prescription fill. While this approach allows for a more rigorous study design, it also limits the relevance of the results to a narrow population of patients. We aimed to make our findings generalisable to the wider population of patients seeking care in ED, and therefore did not impose stricter data sufficiency inclusion criteria.

Outcome

Our outcome, the 'index' OUD, was a dichotomous variable indicating whether a patient was diagnosed with OUD during the ED encounter. Because our goal was to identify patients at high risk of OUD to refer for screening and diagnosis, we used a broad definition of OUD which included a range of opioid-related disorders, adverse effects of opioids and opioid poisoning.²¹ We identified OUD using ICD-10 codes (F11, T40.0, T40.1, T40.2, T40.3, and T40.6) and ICD-9 codes (304.00, 304.01, 304.02, 304.03, 304.70, 304.71, 304.72, 304.73, 305.50, 305.51, 305.52, 305.53, 965.00, 965.01, 965.02, 965.09, E850.0, E850.1, E850.2, E935.0, E935.1, E935.2, E940.1.). While we acknowledge the results from previous studies^{22 23} that examined validity of ICD codes in the identification of OUD and found that ICD diagnoses did not perfectly align with the DSM OUD diagnostic criteria, we proceed with this generous definition with the intention of retaining everyone who might be at risk of OUD and potentially benefit from screening.

Predictors

From EHR patient demographics, we obtained Andersen Model predisposing and enabling characteristics: age, sex, race, ethnicity and marital status.^{24–26} Using the EHR's encounter-level data, we constructed Andersen Model indicators for need factors. We included comorbidity levels measured by the Charlson Comorbidity Index,²⁷ history of alcohol use disorder, tobacco use disorder, mental illness, domestic violence and abuse, chronic pain and an indication of previous hospitalisation.^{24–26} An indicator for previous diagnosis of OUD was also ascertained from the EHRs.²⁸ All clinical conditions were assessed in the 12-month baseline period and defined using previously validated algorithms from the Centers for Medicare and Medicaid Chronic Conditions Warehouse.²⁹ We also used the EHR's prescription order data to assess previous use of analgesics, psychotropics, medications for OUD (MOUD) and opioids.²⁶ Psychiatric disorder was comprised of any anxiety disorder, bipolar disorder, depression, personality disorder, other psychotic disorder, post-traumatic stress disorder and schizophrenia. Pain was captured by dorsopathies, migraine and chronic headache, other pain, fibromyalgia, and chronic pain and fatigue. Physical accidents were comprised of transportation accidents, burns and corrosions and injuries. In sensitivity analyses, we analysed our data using disaggregated list of variables (for instance, instead of psychiatric disorder, we used individual indicators for anxiety, depression, etc). Our EHR database did not capture information on previous

ODU-specific screening and therefore we were unable to include such variables as predictors.

Statistical analysis and machine learning model development

We used absolute standardised differences to evaluate the extent of the difference in means and prevalence of encounter characteristics between OUD and non-OUD groups.³⁰ To assess the magnitude of the association between individual patient characteristics and diagnosis of OUD, we used generalised estimating equations accounting for repeated measures. Previous diagnosis of OUD was the strongest predictor and related to many other risk factors. Therefore, we also explored the relationship between patient characteristics and OUD among patients with no previous record of OUD.

To explore the feasibility of building a predictive model that could be implemented and used in EDs, we applied the Gradient Boosting Classifier,³¹ XGBoost, as implemented in Python V.2.7.17. We set aside a 50% random sample of our dataset as an independent validation dataset and used the remainder for model development. All performance measures are reported for the independent validation dataset. To develop the prediction model, we used a grid search approach, with stratified fivefold cross-validation, to identify the values of the hyperparameters that optimise the model prediction performance. Online supplemental table 1 provides the list of parameters used in optimisation and the range of explored values.

As is the case with many real-world health outcomes, our outcome is infrequent. Unbalanced outcome distributions can result in poor sensitivity in machine learning models and several remedial measures have been proposed to address this issue.^{32–33} We evaluated and compared the performance of three approaches: (1) the naïve model, where no imbalance remedy was applied, (2) a simple oversampling of the minority class and (3) a synthetic minority oversampling technique (SMOTE).³² In this study, the minority class is comprised of the cases where the outcome, index OUD, is equal to 1. In simplified terms, SMOTE is an oversampling method where new cases of the minority class are generated through first randomly selecting a case, then finding its *k* minority class neighbours, and finally generating a synthetic case by randomly drawing from the distributions of the characteristics of its neighbours. SMOTE's key feature differentiating it from a simple oversampling approach is the process of synthesising new cases instead of generating replicates. In our analyses, we set the parameter *k*, the number of nearest neighbours, to five. In our model optimisation process, we used an F1 score as the primary measure of the models' performance.³⁴ The F1 score is a harmonic mean of the precision (positive predictive value) and recall (sensitivity) and therefore is a more robust measure of model performance than other metrics.^{35–37} To comprehensively assess the performance of the final models, we also report area under the receiver operating characteristic curve (AUC), precision and recall. Because the process of selecting threshold values is a multifaceted

and complex process,³⁸ we present all evaluation metrics based on the default threshold of 0.5. We report these performance measures in total for the entire sample and stratified by outcome.

In order to produce findings useful for identifying new or not previously identified OUD, we additionally investigated how well a predictive model would perform among patients with no previous diagnosis of OUD. In primary analyses, we built two sets of machine learning models. We constructed (1) one model for all patients and (2) one for the subset of patients with no history of OUD diagnosis. Because of strong association between baseline use of medications to treat OUD and previous diagnosis of OUD, we also explored models that included both variables as well as one variable at a time and reported results from the model with the best performance metrics. In sensitivity analyses, we built the machine learning models using disaggregated risk factors. The predictive performance of the models based on the full sample was also stratified by presence/absence of previous OUD.

To understand the trade-offs of using the full sample machine learning model versus a simple indicator of history of OUD, we performed simple calculations for 10 000 hypothetical encounters in the ED. For the purpose of this exercise, we assumed that our outcome, index OUD, was a true reflection of patients' OUD status. By applying rates and findings from our data, we evaluated how many patients would be correctly identified for screening, how many would be missed and how many would be screened unnecessarily by each of the two methods.

We used SAS V.9.4 for statistical analyses and Python V.2.7.17 for machine learning model implementation.

Patient and public involvement

In our practice and research, patients have told us that it is very difficult to find and engage with treatment for OUD. An emergency room visit can be a turning point that makes a person open to receiving treatment. For this reason, we designed this study to develop an efficient means of recognising OUD and need for treatment so that people could be connected with treatment at their ED visit. Because we used already existing administrative data, patients were not involved in the study design and implementation.

RESULTS

The final analytical sample included 345 728 ED encounters. The prevalence of index OUD diagnosis in the full cohort was 1.16%. Most patients had no evidence of OUD history prior to index OUD diagnosis (N=327 369 visits, 95%). The prevalence of index OUD diagnosis in this group was 0.57%. OUD was coded as a primary concern in 1024 (25.6%) of the 3995 cases. When OUD was not noted as a chief complaint, the most frequently recorded primary reasons for the remaining 2971 ED visits were: abdominal and pelvic pain (4.4%),

Table 1 Characteristics of the study cohort stratified by presence of OUD diagnosis prior to and at index ED visit

Characteristic	All ED visits			ED visits with no previous diagnosis of OUD in EHR		
	No OUD	OUD	ASD	No OUD	OUD	ASD
N	341 733	3995		325 501	1868	
Demographics						
Age, mean (SD)	49.5 (21.1)	42.6 (16.0)	37.2	49.6 (21.3)	45.1 (17.6)	22.8
Sex			21.3			15.8
Male	133 699 (39.1)	1984 (49.7)		126 523 (38.9)	872 (46.7)	
Female	208 034 (60.9)	2011 (50.3)		198 978 (61.1)	996 (53.3)	
Ethnicity			19.3			13.5
Hispanic	20 467 (6.0)	87 (2.2)		20 115 (6.2)	62 (3.3)	
Non-Hispanic	321 266 (94.0)	3908 (97.8)		305 386 (93.8)	1806 (96.7)	
Race						
Non-white	136 489 (39.9)	771 (19.3)	46.4	131 866 (40.5)	397 (21.3)	42.6
White	205 244 (60.1)	3224 (80.7)	46.4	193 635 (59.4)	1471 (78.7)	42.6
Marital status						
Non-married	221 299 (64.8)	3103 (77.7)	28.8	209 315 (64.3)	1364 (73.0)	18.9
Married	120 434 (35.2)	892 (22.3)	28.8	116 186 (35.7)	504 (27.0)	18.9
Clinical characteristics assessed in baseline period						
Charlson Comorbidity Index, mean (SD)	2.0 (2.9)	1.6 (2.6)	13.5	1.9 (2.8)	1.6 (2.6)	12.8
Chronic pain (any)	156 245 (45.7)	2369 (59.3)	27.4	143 065 (44.0)	1000 (53.5)	19.3
Physical injuries, burns, accidents (any)	135 983 (39.8)	1878 (47.0)	14.6	125 660 (38.6)	783 (41.9)	6.8
Mental health condition (any)	116 719 (34.2)	2364 (59.2)	51.8	104 621 (32.1)	767 (41.1)	18.6
Opioid use disorder	16 232 (4.8)	2127 (53.2)	126.4			
Alcohol use disorder	23 041 (6.7)	703 (17.6)	33.7	19 528 (6.0)	159 (8.5)	9.7
Tobacco use	96 116 (28.1)	2425 (60.7)	69.4	86 570 (26.6)	846 (45.3)	39.7
Domestic abuse/neglect	1735 (0.5)	55 (1.4)	9.0	1419 (0.4)	21 (1.1)	7.8
Inpatient stay	120 494 (35.3)	1906 (47.7)	25.5	108 647 (33.4)	554 (29.7)	8.0
Other analgesics	165 784 (48.5)	2306 (57.7)	18.5	153 659 (47.2)	899 (48.1)	1.8
Opioids	156 546 (45.8)	2349 (58.8)	26.2	143 473 (44.1)	1072 (57.4)	26.9
Psychotropic meds	140 578 (41.1)	2270 (56.8)	31.8	127 939 (39.3)	842 (45.1)	11.7
MOUD	8193 (2.4)	589 (14.7)	45.2	5174 (1.6)	77 (4.1)	15.2

Not reportable, cell size less than 11.

N (%) are reported unless otherwise noted.

.ASD, absolute standardised difference; ED, emergency department; EHR, electronic health record; MOUD, medications for OUD; OUD, opioid use disorder.

alcohol-related disorders (4.1%), symptoms and signs involving emotional state (3.4%), dorsalgias (3.3%) and other sepsis (2.8%). [Table 1](#) presents sample characteristics for the full cohort and for the subcohort of patients with no history of OUD, stratified by index diagnosis of OUD. When compared with patients with no index diagnosis of OUD, patients with a diagnosis of OUD were more likely to be male, white, non-married, and have a history of chronic pain, mental health conditions, alcohol use and tobacco use disorders. Fifty-three per cent of patients with index OUD diagnosis also had

a previous diagnosis of OUD in their EHRs, compared with only 4.8% of patients with no index OUD. Almost 15% of patients with index OUD were previously treated with medications for OUD (methadone, buprenorphine, naltrexone). In contrast, the group with no index OUD received MOUD in only 2.4% of their encounters. While the demographic characteristics in the full cohort and in the subcohort with no index OUD were similar, the prevalence of different clinical conditions (for instance, mental health conditions, alcohol and tobacco use disorders) in the subcohort was, on average, lower than in

Table 2 Generalised estimating equations model of patient-level factors and diagnosis of OUD in the ED

Characteristic	All ED visits (history of OUD included in the model)		All ED visits (history of OUD excluded from the model)		ED visits with no previous diagnosis of OUD	
	OR	95% CI	OR	95% CI	OR	95% CI
Previous diagnosis of OUD	13.36	11.84 to 15.09				
Age*	0.89	0.87 to 0.91	0.85	0.83 to 0.87	0.90	0.88 to 0.93
Sex: male	1.38	1.27 to 1.51	1.53	1.39 to 1.67	1.44	1.29 to 1.61
Ethnicity: Hispanic	0.82	0.62 to 1.07	0.77	0.59 to 1.02	1.06	0.79 to 1.43
Race: white	2.33	2.07 to 2.64	2.95	2.61 to 3.33	2.64	2.29 to 3.03
Marital status: non-married	1.38	1.24 to 1.53	1.63	1.47 to 1.82	1.53	1.35 to 1.73
CCI	0.92	0.90 to 0.94	0.92	0.90 to 0.94	0.95	0.92 to 0.98
Chronic pain (any)	1.12	1.01 to 1.24	1.36	1.24 to 1.49	1.44	1.29 to 1.62
Physical injuries, burns, accidents	0.86	0.78 to 0.94	0.90	0.83 to 0.99	0.90	0.80 to 1.00
Mental health condition	1.14	1.03 to 1.27	1.46	1.32 to 1.62	1.09	0.96 to 1.23
Alcohol use disorder	1.00	0.87 to 1.15	1.23	1.03 to 1.40	1.08	0.88 to 1.33
Tobacco use disorder	1.82	1.65 to 2.00	2.33	2.11 to 2.57	1.72	1.54 to 1.94
Domestic abuse/neglect	0.96	0.62 to 1.47	1.15	0.72 to 1.81	1.43	0.81 to 2.55
Inpatient stay	0.91	0.81 to 1.00	1.16	1.04 to 1.29	0.72	0.62 to 0.82
Other analgesics	0.89	0.81 to 0.98	0.93	0.85 to 1.03	0.79	0.7 to 0.89
Opioids	1.18	1.06 to 1.31	1.23	1.11 to 1.36	1.80	1.59 to 2.04
Psychotropic meds	1.01	0.90 to 1.11	1.08	0.97 to 1.20	1.10	0.97 to 1.25
MOUD	1.57	1.35 to 1.82	3.41	2.88 to 4.03	2.51	1.90 to 3.32

*OR reflects a change of 10 years.

.CCI, Charlson Comorbidity Index; MOUD, medications for OUD; OUD, opioid use disorder.

the full cohort. Online supplemental table 2 provides descriptive statistics for the disaggregated set of factors.

Table 2 shows ORs that quantify the association between OUD diagnosis in the ED and patient demographics and clinical characteristics. Previous diagnosis of OUD was the strongest predictor of index OUD (OR=13.4, 95% CI: 11.8 to 15.1). Given its high correlation with all other substance use disorders and many of the clinical conditions, it was excluded from the second model where baseline use of MOUD (OR=3.4, 95% CI: 2.9 to 4.0), white race (OR=2.9, 95% CI: 2.6 to 3.3), tobacco use disorder (OR=2.3, 95% CI: 2.1 to 2.6) and non-married marital status (OR=1.6, 95% CI: 1.5 to 1.8) were the strongest predictors. Results from the model based on encounters with no history of OUD indicate that white race (OR=2.6, 95% CI: 2.3 to 3.0), previous use of MOUD (OR=2.5, 95% CI: 1.9 to 3.3), prescription of opioids in the baseline (OR=1.8, 95% CI: 1.6 to 2.1) and tobacco use disorder (OR=1.7, 95% CI: 1.5 to 1.9) were the key risk factors for OUD diagnosis in the ED.

Table 3 presents performance metrics from the machine learning models. The best model, the model built using all ED visits and employing the SMOTE algorithm for outcome imbalance, achieved an overall F1 score of 0.97 and AUC of 0.71. This model attained a specificity of 0.97 but sensitivity of only 0.45. Further, after stratifying

the performance metrics based on previous diagnosis of OUD, the results show that, among ED visits with previous OUD, the model achieved a high sensitivity of 0.81 but a low specificity of 0.43. Among encounters with no previous OUD, the model reached a perfect specificity of 1.0 and a poor sensitivity of 0.02. Findings were similar for the model based on visits with no history of OUD; models achieved high specificity while sensitivity remained very low. Results from our sensitivity analyses, models based on the full cohort as well as the subcohort, indicate that the model using the disaggregated set of variables (results in the Online supplemental table 3) did not perform substantially better than the models using the aggregated set of factors.

Table 4 presents trade-off calculations of our two hypothetical scenarios for selecting patients for OUD screening. These calculations operate on the assumption that if the value of the outcome (OUD in the ED) was 0 but the patient had a history of OUD elsewhere in the record, then the patient was either successfully treated or recovered. The first method, based on a simple EHR indicator of history of OUD, would correctly identify 53.2% of patients with OUD but would require 469 patients to be screened unnecessarily. On the other hand, the method based on the machine learning predictive model would correctly identify 45.0% of the patients with OUD but



Table 3 Performance metrics from XGBoost machine learning model using short list of predictors

Cohort	Subgroup	Outcome	N	Naïve model—no remedy for data imbalance			Oversampling of the minority class			Synthetic oversampling using SMOTE					
				Precision	Recall	F1 score	AUC/ACC	Precision	Recall	F1 score	AUC/ACC	Precision	Recall	F1 score	AUC/ACC
Full	All	Y=0	171193	0.99	1.00	0.99	0.54/0.99	0.99	0.94	0.96	0.66/0.93	0.99	0.97	0.98	0.71/0.96
		Y=1	1985	0.21	0.08	0.12		0.06	0.38	0.11		0.15	0.45	0.23	
		All Macro	173178	0.60	0.54	0.55		0.53	0.66	0.54		0.57	0.71	0.60	
		All Wt	173178	0.98	0.99	0.98		0.98	0.93	0.95		0.98	0.96	0.97	
History of OUD		Y=0	8055	0.89	0.94	0.92	0.54/0.85	0.92	0.67	0.78	0.63/0.66	0.94	0.43	0.59	0.62/0.48
		Y=1	1079	0.26	0.14	0.18		0.19	0.58	0.29		0.16	0.81	0.27	
		All Macro	9134	0.58	0.54	0.55		0.56	0.63	0.53		0.55	0.62	0.43	
		All Wt	9134	0.82	0.85	0.83		0.84	0.66	0.72		0.85	0.48	0.56	
No history of OUD		Y=0	163138	0.99	1.00	1.00	0.50/0.99	1.00	0.95	0.97	0.55/0.94	0.99	1.00	1.00	0.51/0.99
		Y=1	906	0.02	0.00	0.01		0.02	0.15	0.03		0.04	0.02	0.03	
		All Macro	164044	0.51	0.50	0.50		0.51	0.55	0.50		0.52	0.51	0.51	
		All Wt	164044	0.99	0.99	0.99		0.99	0.94	0.97		0.99	0.99	0.99	
Sub		Y=0	163050	0.99	1.00	1.00	0.50/0.99	1.00	0.89	0.94	0.58/0.88	0.99	1.00	0.99	0.51/0.99
		Y=1	969	0.02	0.01	0.01		0.01	0.27	0.03		0.03	0.02	0.02	
		All Macro	164019	0.51	0.50	0.50		0.50	0.58	0.48		0.51	0.51	0.51	
		All Wt	164019	0.99	0.99	0.99		0.99	0.88	0.93		0.99	0.99	0.99	

All Macro=performance metrics for the entire sample were summarised using unweighted arithmetic mean (ie, Y=0 and Y=1 are treated equally regardless of their respective support/N).
 All Wt=performance metrics for the entire sample were summarised using weighted average (ie, Y=0 and Y=1 are weighted according to their respective support/N).
 Full=all ED visits were included in the model building.
 All=metrics reported for all ED visits.
 History of OUD=metrics reported for ED visits for patients with previous diagnosis of OUD.
 No history of OUD=metrics reported for ED visits for patients with no previous diagnosis of OUD.
 Sub=ED visits for patients with no history of OUD were included in the model building.
 .ACC, accuracy; AUC, area under the receiver operating characteristic curve; ED, emergency department; OUD, opioid use disorder; SMOTE, synthetic minority oversampling technique.

Table 4 Trade-off calculations, for 10 000 hypothetical ER encounters and a true OUD rate of 1.15%, to show what would happen if we screened (1) everyone indicated by the records of history of OUD, and (2) everyone indicated by the machine model

Method	# of patients indicated for screening	# (%) of patients correctly identified with OUD	# (%) of patients with OUD who were NOT identified	# of patients screened unnecessarily
(1) Indicator of history of OUD	531	61 (53.2)	54 (46.8)	469
(2) Machine learning predictive model	348	52 (45.0)	63 (55.0)	296

ER, emergency room; OUD, opioid use disorder.

would require only 296 patients to have an unwarranted screen.

DISCUSSION

This study examines strategies to identify patients who are at risk of OUD, among individuals presenting at EDs in order to facilitate targeted screening by busy ED staff. Among ED encounters at a large healthcare system, 1.16% resulted in the diagnosis of OUD. Patients who present at the ED with OUD are typically white, smokers, and with psychiatric disorders, chronic pain, substance use disorders and use of prescription opioids. History of OUD is the single strongest predictor of OUD diagnosis at the ED and is simultaneously strongly correlated with many other clinical conditions such as substance use disorders and mental health conditions. Among patients with an OUD diagnosis at the ED, over 50% had a previous diagnosis of OUD, while among encounters with no OUD, only 4.8% had a record of earlier OUD diagnosis.

Even though our predictive model of OUD based on patients' EHRs achieved a high overall accuracy of 0.96, weighted F1 score of 0.97 and AUC of 0.71, its suitability for use in clinical practice is debatable due to low sensitivity. Diagnosis of OUD, like many other health outcomes or adverse effects, is a rare event, posing a significant challenge in development of machine learning predictive models. Our application of remedial approaches, simple oversampling and SMOTE improved the performance of the models slightly; however, poor sensitivity remained. While we would be inclined to accept a low specificity of the models, as it would cause no harm to the patients to have them selected for an OUD screening questionnaire, low sensitivity makes us pause and assess these results in the light of other viable options.

Even though both of our modelling approaches, logistic regression and machine learning models, suggest that there are several clinical and demographic factors that are strongly associated with OUD, the machine learning model implies that other non-clinical variables are needed to make accurate prediction of risk of OUD. For example, inclusion of social context and history in the EHR has the potential to improve model outcomes.³⁹ Currently, history of OUD stands out as the most valuable predictor to select patients for OUD screening in EDs. A flag in the medical record indicating that a patient was

diagnosed with OUD in the past year could serve as a simple first step in a decision-making protocol to select patients for OUD screening. Our trade-off calculation of 10 000 hypothetical ED encounters suggests that having such a flag in the EHR would correctly identify 53% of patients with OUD for screening, while 469 out of the 10 000 would be screened unnecessarily. However, this calculation was based on the generous assumption that our outcome, index OUD in the ED, was recorded properly if the patient truly had a current OUD. Additionally, if OUD was not noted in the ED, and we assume that the patient had ongoing chronic OUD as indicated by historical diagnosis of OUD, then some, if not all, of the 469 screening tests might be justified. If the machine learning model were to be incorporated into the EHR system, the model would correctly identify 45% of the patients with true OUD and only 296 additional patients without OUD would be screened. Given the complexity of incorporating a machine learning model into an EHR system, a simple indicator of previous diagnosis of OUD that alerts triage nurses seems like a natural choice to help identify patients for OUD screening. On the other hand, since staff time is scarce, incorporation of a machine learning model would cut nearly in half the number of ED patients unnecessarily screened. If, instead, a patient had no previous record of OUD, then ED staff would have to rely heavily on patient report of OUD or a chief complaint related to substance use as their reason for coming to the ED.

Our investigation makes a critical point with regard to construction of machine learning models for infrequent outcome prediction. It is necessary to not only assess the overall predictive performance measures such as accuracy and AUC, but also conduct careful review of sensitivity, specificity and stratification of the results by key predictors to appropriately assess the model's usefulness in practice and consequences of its use. For instance, in our study, the false-positive result has minimal consequences, but a false negative might have serious ramifications. With opioid overdoses increasing at an alarming rate in the setting of COVID-19, it becomes paramount that people with a risk of OUD are identified and offered treatment.

Our data show that among ED visits where no OUD was noted, 4.8% had a record of previous OUD. These



cases may represent patients who recovered from OUD or missed diagnosis of OUD. The median time since the last OUD diagnosis among ED encounters with OUD was 59 days, while among encounters where no OUD was noted, the median time was 98 days, a significantly longer time. Nevertheless, OUD is a chronic disease, and it is unlikely that the patients with OUD recovered in this length of time. Because ED clinicians list only diagnoses for problems that were actively treated or acknowledged while in the ED, this suggests that OUD may not have been addressed during these subsequent ED encounters. This is a missed opportunity and further strengthens the argument that an automated screening approach would increase ED clinicians' ability to help patients with OUD.

While this study portrays a solid picture of patients diagnosed with OUD in EDs, it has some important limitations. Like many studies using real-world data, the data available to us were limited and imperfect. First, the validity of using ICD codes to diagnose OUD has been found to be adequate at best. Second, the documentation of ICD diagnoses in EDs might be infrequent and inconsistent. To the extent that OUD was consistently missed in patients, our models may perform better than we indicate. Because our EHR data contain records from only one healthcare system, there is a possibility of variable misclassification due to missing medications and diagnoses from encounters outside of the studied health system. These limitations emphasise and draw attention to earlier calls for development of validated algorithms to accurately identify patients with OUD in observational data.⁴⁰ Additionally, the protocols for OUD screening in an ED and resources to support them might be health system specific and not generalisable to other settings. Our study used EHRs coded as structured data. Since many medical data elements are coded in the notes as free text, future research should explore the potential of inclusion of unstructured data in model building to increase the accuracy of the predictions. These limitations emphasise why implementation of machine learning models in real-world clinical settings continues to be rare.^{41 42}

Our findings have important implications for research and practice. First, while it is difficult to accurately predict who has OUD, it is clear that history of OUD should be considered when selecting patients for screening. Given that it is not feasible to screen every single individual who presents at the ED, ED conversations about chronic pain, source of injury, substance use and psychiatric disorders may yield important information about risk of OUD in order to connect patients to needed services.

Contributors IA helped design the study, performed all analyses and wrote the draft of the manuscript. IA is the guarantor of the manuscript and accepts full responsibility for the work and conduct of the study. KT contributed to the study design and critically revised the article. RJ provided subject matter expertise in the area of addiction medicine and critically reviewed the manuscript. All authors discussed the results and contributed to the final manuscript.

Funding This work was supported by North Carolina Department of Health and Human Services.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval This study was reviewed by the Office of Human Research Ethics of the University of North Carolina at Chapel Hill, which determined that this submission did not constitute human subjects research as defined under federal regulations (45 CFR 46.102 (d or f) and 21 CFR 56.102(c)(e)(l)) and did not require IRB approval.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The data used in this study cannot be made publicly available. These data were obtained from the University of North Carolina Data Warehouse for Health, and are subject to a data use agreement that does not allow release or sharing of these data.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Izabela E Annis <http://orcid.org/0000-0002-5011-3471>

REFERENCES

- Weiss AJ, Elixhauser A, Barrett ML. Opioid-related inpatient stays and emergency department visits by state. *Statistical brief*:219 www.hcup-us.ahrq.gov/reports/statbriefs/sb219-OpioidHospital-Stays-ED-Visits-by-State.pdf
- CDC. Identifying increases in opioid overdoses. centers for disease control and prevention. *Published* 2018 https://www.cdc.gov/vitalsigns/opioid-overdoses/index.html#anchor_1519932571
- Slavova S, Rock P, Bush HM, et al. Signal of increased opioid overdose during COVID-19 from emergency medical services data. *Drug Alcohol Depend* 2020;214:108176.
- Bogan C, Jennings L, Haynes L, et al. Implementation of emergency department-initiated buprenorphine for opioid use disorder in a rural southern state. *J Subst Abuse Treat* 2020;112:73–8.
- Samuels EA, D'Onofrio G, Huntley K, et al. A quality framework for emergency department treatment of opioid use disorder. *Ann Emerg Med* 2019;73:237–47.
- Hawk K, D'Onofrio G. Emergency department screening and interventions for substance use disorders [published correction appears in *Addict Sci Clin Pract* 2018;13:18.
- D'Onofrio G, O'Connor PG, Pantalon MV, et al. Emergency department-initiated buprenorphine/naloxone treatment for opioid dependence: a randomized clinical trial. *JAMA* 2015;313:1636–44.
- Glober NK, Hern G, McBride O, et al. Variations in the California emergency medical services response to opioid use disorder. *West J Emerg Med* 2020;21:671–6.
- Hoyeck P, Wiercigroch D, Clarke C, et al. Understanding current practice of opioid use disorder management in emergency departments across Canada: a cross-sectional study. *CJEM* 2020;22:494–8.
- Webster LR, Webster RM. Predicting aberrant behaviors in opioid-treated patients: preliminary validation of the opioid risk tool. *Pain Med* 2005;6:432–42.
- Atluri SL, Sudarshan G. Development of a screening tool to detect the risk of inappropriate prescription opioid use in patients with chronic pain. *Pain Physician* 2004;7:333–8.
- Klimas J, Gorfinkel L, Fairbairn N, et al. Strategies to identify patient risks of prescription opioid addiction when initiating opioids for pain: a systematic review. *JAMA Netw Open* 2019;2:e193365.

- 13 Webster LR. Risk factors for opioid-use disorder and overdose. *Anesth Analg* 2017;125:1741–8.
- 14 Kansagara D, Englander H, Salanitro A, *et al.* Risk prediction models for hospital readmission: a systematic review. *JAMA* 2011;306:1688.
- 15 Hao S, Jin B, Shin AY, *et al.* Risk prediction of emergency department revisit 30 days post discharge: a prospective study. *PLoS One* 2014;9:e112944.
- 16 Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS One* 2018;13:e0201016.
- 17 First MB, Williams JB, Karg RS. User's guide for the SCID-5-CV Structured Clinical Interview for DSM-5® disorders: Clinical version. *American Psychiatric Publishing* 2016.
- 18 Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav* 1995;36:1–10.
- 19 Census Bureau US. Available: <https://www.census.gov/content/dam/Census/library/publications/2016/acs/acsgeo-1.pdf>
- 20 Bower JK, Patel S, Rudy JE, *et al.* Addressing bias in electronic health Record-based surveillance of cardiovascular disease risk: finding the signal through the noise. *Curr Epidemiol Rep* 2017;4:346–52.
- 21 Weiss AJ, Heslin KC, Stocks C. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Rockville (MD): Agency for Healthcare Research and Quality (US). In: *Hospital Inpatient Stays Related to Opioid Use Disorder and Endocarditis, 2016: Statistical Brief #256*, 2016.
- 22 Lagisetty P, Garpestad C, Larkin A, *et al.* Identifying individuals with opioid use disorder: validity of international classification of diseases diagnostic codes for opioid use, dependence and abuse. *Drug Alcohol Depend* 2021;221:108583.
- 23 Howell BA, Abel EA, Park D, *et al.* Validity of incident opioid use disorder (OUD) diagnoses in administrative data: a chart verification study. *J Gen Intern Med* 2021;36:1264–70.
- 24 Clement KC, Canner JK, Lawton JS, *et al.* Predictors of new persistent opioid use after coronary artery bypass grafting. *J Thorac Cardiovasc Surg* 2020;160:954–63.
- 25 Bahorik AL, Satre DD, Kline-Simon AH, *et al.* Alcohol, cannabis, and opioid use disorders, and disease burden in an integrated health care system. *J Addict Med* 2017;11:3–9.
- 26 Shei A, Rice JB, Kirson NY, *et al.* Characteristics of high-cost patients diagnosed with opioid abuse. *J Manag Care Spec Pharm* 2015;21:902–12.
- 27 Quan H, Sundararajan V, Halfon P, *et al.* Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43:1130–9.
- 28 Rhon DI, Snodgrass SJ, Cleland JA, *et al.* The risk of prior opioid exposure on future opioid use and comorbidities in individuals with non-acute musculoskeletal knee pain. *J Prim Care Community Health* 2020;11:215013272095743.
- 29 Condition Categories - Chronic Conditions Data Warehouse. www2.ccwdata.org. Available: <https://www2.ccwdata.org/web/guest/condition-categories> [Accessed Dec 2019].
- 30 Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun Stat Simul Comput* 2009;38:1228–34.
- 31 Friedman JH. Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* 2001;29:1189–232.
- 32 Chawla NV, Bowyer KW, Hall LO, *et al.* SMOTE: synthetic minority Over-sampling technique. *J Artif Intell Res* 2002;16:321–57.
- 33 He H, Bai Y, Garcia E. Adaptive synthetic sampling approach for imbalanced learning. *IEEE International Joint Conference on Neural Networks* 2008:1322–8.
- 34 Perner P. *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13–20, 2012. Proceedings.* Springer;2012.
- 35 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.
- 36 Wang F. Machine learning for predicting rare clinical Outcomes-Finding needles in a haystack. *JAMA Netw Open* 2021;4:e2110738.
- 37 Liu Z, Cao, W G, *et al.* Self-paced ensemble for highly imbalanced massive data classification. In: ICDE, ed. In: *IEEE 36th international conference on data engineering*, 2020: 841–52.
- 38 Wynants L, van Smeden M, McLernon DJ. Three myths about risk thresholds for prediction models. *BMC Med.*;2019:192.
- 39 Strain E. UpToDate. Uptodate.com. published, 2019. Available: <https://www.uptodate.com/contents/opioid-use-disorder-epidemiology-pharmacology-clinical-manifestations-course-screening-assessment-and-diagnosis> [Accessed Jan 2020].
- 40 Morgan JR, LaRochelle MR, Commentary on Karnik. Commentary on Karnik *et al.*: harmonization now-the need for consistent, validated measures to identify opioid use disorder in observational data. *Addiction* 2022;117:2448–9.
- 41 Kelly CJ, Karthikesalingam A, Suleyman M. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.*;2019:195.
- 42 Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N Engl J Med* 2017;376:2507–9.