JAMIA Open, 2023, **6(3)**, ooad081
https://doi.org/10.1093/jamiaopen/ooad081
Research and Applications

Research and Applications

Enhanced phenotypes for identifying opioid overdose in emergency department visit electronic health record data

Ralph Ward, PhD[1,*], Jihad S. Obeid, MD, FAMIA[2], Lindsey Jennings, MD, MPH[3], Elizabeth Szwast, MPH[2], William Garrett Hayes, MD[4], Royal Pipaliya, MD[4], Cameron Bailey, MD[4], Skylar Faul, MD[5], Brianna Polyak[6], George Hamilton Baker, MD, MS[7], Jenna L. McCauley, PhD[8], Leslie A. Lenert, MD, MS, FACP, FACMI[2]

[1]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC 29425, United States, [2]Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC 29425, United States, [3]Department of Emergency Medicine, Medical University of South Carolina, Charleston, SC 29425, United States, [4]College of Medicine, Medical University of South Carolina, Charleston, SC 29425, United States, [5]School of Medicine, Mercer University, Macon, GA 31207, United States, [6]School of Medicine, University of Texas Rio Grande Valley, Edinburg, TX 78539, United States, [7]Department of Pediatric Cardiology, Medical University of South Carolina, Charleston, SC 29425, United States, [8]Department of Psychiatry, Medical University of South Carolina, Charleston, SC 29425, United States

*Corresponding author: Ralph Ward, PhD, Department of Public Health Sciences, Medical University of South Carolina, 135 Cannon St., Charleston, SC 29401 (wardrc@musc.edu)

ABSTRACT

Background: Accurate identification of opioid overdose (OOD) cases in electronic healthcare record (EHR) data is an important element in surveillance, empirical research, and clinical intervention. We sought to improve existing OOD electronic phenotypes by incorporating new data types beyond diagnostic codes and by applying several statistical and machine learning methods.

Materials and Methods: We developed an EHR dataset of emergency department visits involving OOD cases or patients considered at risk for an OOD and ascertained true OOD status through manual chart reviews. We developed and validated prediction models using Random Forest, Extreme Gradient Boost, and Elastic Net models that incorporated 717 features involving primary and second diagnoses, chief complaints, medications prescribed, vital signs, laboratory results, and procedural codes. We also developed models limited to single data types.

Results: A total of 1718 records involving 1485 patients were manually reviewed; 541 (36.4%) patients had one or more OOD. Prediction performance was similar for all models; sensitivity varied from 94% to 97%; and area under the receiver operating characteristic curve (AUC) was 98% for all methods. The primary diagnosis and chief complaint were the most important contributors to AUC performance; primary diagnoses and medication class contributed most to sensitivity; chief complaint, primary diagnosis, and vital signs were most important for specificity. Models limited to decision support data types available in real time demonstrated robust prediction performance.

Conclusions: Substantial prediction performance improvements were demonstrated for identifying OODs in EHR data. Our e-phenotypes could be applied in surveillance, retrospective empirical applications, or clinical decision support systems.

LAY SUMMARY

Methods to accurately identify opioid overdose (OOD) cases in electronic healthcare record (EHR) data are important tools for surveillance, empirical research, and clinical interventions needed to help mitigate the opioid crisis. We sought to improve existing OOD EHR phenotypes through 2 innovations: (1) incorporating new data types beyond diagnostic codes and (2) applying several advanced statistical and machine learning methods. We first developed an EHR dataset of 1718 emergency department (ED) visits that was a mixture of 621 OOD cases and 1097 non-OOD visits involving patients considered at high risk for an overdose. These non-overdose patients frequently had symptoms that a model might confuse with an overdose, such as withdrawal symptoms. The goal was to train our models to differentiate between a true OOD and other closely related non-OOD conditions. We determined true OOD visits through manual chart reviews. Next, we developed 3 competing prediction model designs that each used 6 types of EHR data (primary and second diagnosis codes, chief complaints, medications prescribed, vital signs, laboratory results, and procedural codes). We showed that these models had improved prediction performance over previously published methods. We believe these models could be used to drive real-time clinical decision support.

Key words: opiate overdose; electronic phenotype; prediction models.

Introduction

Opioid overdose remains a serious and growing public health concern, with the number of overdose-related deaths continuing to rise. According to the Centers for Disease Control and Prevention (CDC), opioid overdose (OOD) in the United States was responsible for 68 630 deaths in 2020, a 37.8% increase from the previous year with synthetic opioids such as fentanyl comprising 82.4% of OOD deaths.[1] The emergency department (ED) and emergency medical services are often the first point of care for OOD

Received: May 31, 2023; **Revised:** August 22, 2023; **Editorial Decision:** August 28, 2023; **Accepted:** August 29, 2023

patients, making the ED a critical setting for the timely identification and appropriate management of survivors. Accurate retrospective identification of cases is also essential for epidemiological surveillance studies, assessing the burden of OOD, and evaluating management strategies such as dispensing naloxone kits, initiating medications for opioid use disorder and/or treatment referral. Electronic health records (EHR) contain a wealth of data that can be used to develop electronic phenotyping (e-phenotype) strategies, a notion at the heart of the learning health system.[2] However, identifying OOD within EHRs can be challenging since administrative data, including billing codes and other structured data, are inconsistent and frequently missing, and manual review of patient records is time-consuming.

Several studies have examined EHR-based OOD phenotyping approaches with promising results; however, each had its limitations. Chartash et al[3] examined opioid use disorder instead of OOD and focused on the absence or presence of International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes for opioid-related disorders as well as the absence or presence of related key words in chief complaints, and achieved a positive predictive value (PPV) of 95%. Green et al[4] tested several algorithms with ICD codes and natural language processing derived variables from clinical notes and achieved an overall PPV of 87.4% on the validation dataset. Badger et al[5] used random forest and penalized logistic regression to develop OOD phenotypes using ICD codes, opioid medications, procedures, lab results, observations, and clinical notes from $N = 278$ OOD patients who had been identified initially through an ICD code phenotype. The maximum AUC achieved was 0.89, with positive predictive value 0.67 and sensitivity 0.65.

An important attribute for OOD predictive models is data temporality, or the point during an ED visit at which information needed to drive the model actually becomes available. Models based on data assigned post-visit such as ICD-10-CM codes are useful for generation of retrospective registries. Models that use clinical data available in nearly real-time can also be applied to drive decision support tools within the visit, such as prompting for additional data on the overdose cause, planning for follow up care, initiating medications for patients with opioid use disorder, or prescribing take home naloxone.[6] Examples of clinical data that are available within a visit include emergency department chief complaints, triage nursing notes, administration and response to antidotes (eg, naloxone), vital signs, and laboratory results such as drug screening testing. In addition, baseline data from prior to a visit, such as ICD-10-CM diagnostic codes from prior emergency department visits or from outpatient care may also be useful.

In this study, we sought to evaluate previous approaches from both an accuracy and temporality of data perspective and improve on them by adding more opioid overdose-related ICD codes including recent definitions established by the CDC for surveillance of drug overdoses in ED settings.[7] We also included a multitude of EHR-based variables, including chief complaint, vital signs, procedure codes, administered medications, and laboratory values. We evaluated multiple machine learning and statistical models while developing our e-phenotype, and examined the impact of various predictors on the accuracy of the models to streamline future implementation of the e-phenotype at other sites.

Methods

Population and data sources

Electronic medical record data were retrieved for emergency department (ED) visits between January 1, 2012 and April 30, 2021 at the Medical University of South Carolina (MUSC) for patients over age 18 years who: (1) were treated for a suspected opioid overdose (OOD) or (2) met defined conditions for being considered at risk for an OOD at some point (further elaboration below). The MUSC Health Network includes the main campus in Charleston that is an academic, quaternary referral center, along with several community hospitals across the state. Definitions for "suspected OOD" and "at-risk for future OOD" were based on international classification of diseases (ICD) code listings established by Green et al.[4] Because these listings were limited to the ICD-9-CM system, we used a general equivalence mapping originally developed by the Centers for Medicare and Medicaid Services to map the listings to the ICD-10-CM system.[8] For the at-risk group, we included all ED visits that were within 1 year before or after the date on which the patient met inclusion criteria. The suspected OOD group was expected to be reasonably small (less than 1000 patients), but the at-risk group was expected to be substantially larger. We used several rules to shrink this group to focus on cases that might confuse an algorithm by appearing to be similar to an OOD but which in fact involved other diagnoses. These rules and a listing of the ICD codes used to identify suspected OOD and at-risk patients are included in the Supplementary Material. The study was approved by the MUSC Institutional Review Board (Pro00088536).

Development of validated cases

Manual chart reviews of the ED visits were used to establish a "gold standard" OOD status against which the phenotype models could be trained and validated. Clinical notes and relevant information from ED visits were imported into a REDCap database,[9] which allowed reviewers to annotate each visit as OOD or not OOD. A group of 13 reviewers consisting of 4 clinicians, several medical students, and study team members assisted with the reviews. Detailed review guidelines were defined by the lead clinical team (which included an ED physician and an MD informatics expert) and made available to the reviewers via a link from REDCap. All reviewers were added to the IRB protocol and trained using the review guidelines and a video recorded training session. The charts were assigned evenly to all reviewers with some overlap allowing for a portion of the charts to be reviewed independently by 2 reviewers to allow for inter-rater reliability (IRR) assessment. Reviewers were allowed to check a box in REDCap to request group reviews for difficult cases. Nonconcordant cases and group review requests were reviewed by the lead clinical team. Cases were excluded if the ED clinical notes were incomplete or ambiguous.

Predictor variables

The prediction models used predictor variables (or features) derived only from coded data recorded during an ED visit; clinical text notes were used only during the gold standard chart reviews. We used the following data types as predictors: primary and secondary diagnosis ICD codes, chief complaint (selected from a fixed menu of conditions), medication therapeutic class, vital signs recorded during ED visit triage (respiration rate, pulse, systolic, and diastolic blood pressure), Current Procedural

Terminology (CPT) codes, and laboratory results. A total of 717 features were used to develop the prediction models.

Statistical and machine learning methods

We used 3 methods to develop competing prediction models: Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Elastic Net penalized regression (ENET). For each method, we generated 1000 bootstrapped datasets (sampled with replacement) that were the same size as the full dataset. These were used to determine 95% CIs from the prediction performance statistics collected from 1000 model runs. For XGBoost and ENET, we held out a 10% random sample during each run to use as validation data from which we reported prediction performance. For RF, we used 750 independent decision trees during each run to train a "forest" to determine which variables were most important in classifying a visit as an OOD or non-OOD; the hyperparameter mtry (number of variables randomly sampled as candidates at each split) was set at 27 (square root of feature size).[10] In each RF tree, part of the data was held out (out of bag) and used to test that tree's predictive performance on validation data. We used the out of bag results for the full forest to assess the model's predictive performance. XGBoost is also a decision tree method, but unlike RF, the trees are not independent.[11] Each tree was generated sequentially, using the prediction performance of the previous tree to "boost" performance of the next. We set hyperparameters based on previous experience; these are provided in the sample code in the Supplementary Material. ENET is a penalized regression method designed to "shrink" some predictor estimates to 0 based on their lower importance in predicting an OOD. Related methods are LASSO and Ridge Regression; each is optimal in certain situations and ENET is a compromise between them.[12] The ENET model was first tuned using a nested 10-fold cross-validation step using training data to find the best performing hyperparameters (alpha and lambda), after which the model was retrained on the full training set; validation data was not part of the tuning or training steps. For each RF, XGBoost and ENET model, we report the area under the Receiver Operating Characteristic (ROC) curve (AUC), and statistics derived from the 2×2 table of actual versus predicted OOD (sensitivity, specificity, positive predictive value, F1 score, and accuracy). Cut points were determined based on the Youden score, which is based on the point at which an ROC curve reaches the maximum vertical distance above the diagonal marking 50% prediction probability.[13] For each method, we also report variable importance results; for RF, this was based on the mean decrease of the Gini Index; for XGBoost this was the fractional contribution of each predictor based on the total gain for the times that predictor was used in a tree split; for ENET this was based on the relative magnitude of the absolute value of parameter estimates. The primary analyses involved models that included all available predictors; we also ran secondary models that included a single type (ie, primary diagnosis, or chief complaint, etc.) to assess the relative importance of each type. All analyses were performed in R using packages randomForest, xgboost, and glmnet.[12,14–16] Sample code is included in the Supplementary Material.

Results

A total of 1719 records belonging to 1485 patients were manually reviewed, with 105 records (6%) assigned to 2 reviewers. The IRR was high with Kappa = 0.96. Table 1 provides a summary of the demographic characteristics of the patients. Of the 1485 patients, 541 (36.4%) had one or more ED visits involving an OOD. Approximately 11% of those with an OOD had multiple visits involving an OOD. Non-Hispanic White (NHW) and non-Hispanic Black (NHB) patients comprised most of the cohort (70.9% and 26.1%, respectively), with NHW patients being significantly more likely to have an overdose visit (40.6% for NHW vs 25.1% for NHB, $P<.0001$). Females comprised 42.1% of the cohort and were less likely to have an OOD visit compared to males (28.8% vs 42.0%, $P<.0001$). The median

Table 1. Demographic characteristics of the validated cohort.

Variable	Category	One or more OOD	No OOD	Total	*P*-value
Group size	*N* (%)	541 (36.4%)	944 (63.6%)	1485	
Race and ethnicity	Non-Hispanic White	427 (40.6%)	626 (59.4%)	1053 (70.9%)	<.0001
	Non-Hispanic Black	97 (25.1%)	290 (74.9%)	387 (26.1%)	
	Hispanic	9 (33.3%)	18 (66.6%)	28 (1.8%)	
	Other or missing	8 (44.4%)	10 (55.6%)	18 (1.2%)	
Sex	Female	180 (28.8%)	445 (71.2%)	625 (42.1%)	<.0001
	Male	361 (42.0%)	499 (58.0%)	860 (57.9%)	
Age	Median (IQR)	29 (22, 42)	33 (22, 48)	31 (22, 46)	.0006
Total OOD visits per patient	0	0 (0%)	944 (100%)	944	NA
	1	480 (88.7%)	0	480	
	2	47 (8.7%)	0	47	
	3	11 (2.0%)	0	11	
	4	2 (0.4%)	0	2	
	6	1 (0.2%)	0	1	

Abbreviation: IQR, interquartile range.

Table 2. Performance statistics for full models.

Model	Sensitivity	Specificity	Accuracy	Positive predictive value	F-1 score	ROC-AUC
Random Forest (RF)	0.94 (0.93, 0.96)	0.93 (0.91, 0.95)	0.94 (0.93, 0.94)	0.89 (0.86, 0.91)	0.91 (0.91, 0.92)	0.98 (0.98, 0.99)
Extreme Gradient Boost (XGBoost)	0.97 (0.90, 1.0)	0.94 (0.88, 0.97)	0.94 (0.92, 0.97)	0.90 (0.82, 0.95)	0.92 (0.89, 0.95)	0.98 (0.96, 0.99)
Elastic Net (ENET)	0.97 (0.89, 1.0)	0.96 (0.90, 0.99)	0.96 (0.92, 0.99)	0.93 (0.84, 0.99)	0.94 (0.90, 0.98)	0.98 (0.97, 0.99)

(interquartile) age was 31 years (22, 46); OOD patients were likely to be slightly younger: 29 (22, 42) versus non-OOD patients 33 (22, 48), respectively, $P = .0006$.

Table 2 provides prediction performance statistics and 95% CIs from validation data for the 3 full models (RF, XGBoost, and ENET) based on 1000 model runs on bootstrapped datasets. Performance was similar for all methods; sensitivity varied from 94% to 97%; specificity varied from 93% to 96%, and AUC was 98% for all models. The top 30 out of 717 predictor variables included in the models are shown in Table 3, ranked by variable importance for the ENET model, with similar rankings given in Tables S1 and S2 for the RF and XGBoost models. Each data type has a different color shading (yellow=chief complaint, green=primary or secondary diagnosis ICD code, blue=vital signs, tan=medication class, pink=positive lab result, purple=CPT

Table 3. Relative importance of predictors by model type.

Rank			Importance statistic					
RF	XGB	ENET	RF	XGB	ENET	Code	Data type	Description
2	2	1	87.970	0.145	1.824	T401X1A	Primary dx	Poisoning by heroin, accidental (unintentional), initial encounter
6	3	2	32.156	0.093	0.992	T402X1A	Primary dx	Poisoning by other opioids, accidental (unintentional), initial encounter
1	1	3	114.583	0.483	0.939		Chief Complaint	Drug Overdose
5	4	4	36.024	0.047	0.642		MEDICATIONS	ANTIDOTES
15	10	5	6.344	0.010	0.410	T401X1A	Secondary dx	Poisoning by heroin, accidental (unintentional), initial encounter
10	9	6	11.197	0.015	0.398	96501	Primary dx	Accidental overdose of heroin
83	31	7	1.276	0.002	0.211	T401X4A	Primary dx	Poisoning by heroin, undetermined, initial encounter
4	6	8	43.923	0.034	−0.185		Vitals	Respirations
3	12	9	51.962	0.007	0.184	Y929	Primary dx	Unspecified place or not applicable
34	11	10	2.281	0.009	0.147	96509	Secondary dx	Poisoning by opiates and related narcotics, other
111	74	11	0.832	0.001	0.129	T402X1A	Secondary dx	Poisoning by other opioids, accidental (unintentional), initial encounter
64	58	12	1.739	0.001	0.086	E9800	Secondary dx	Poisoning by analgesics, antipyretics, and antirheumatics
9	13	13	15.091	0.005	−0.085		MEDICATIONS	PSYCHOTHERAPEUTIC DRUGS
35	47	14	2.231	0.001	0.082	E9500	Primary dx	Suicide and self-inflicted poisoning by analgesics, antipyretics, and antirheumatics
75	37	15	1.365	0.002	0.081	96501	Secondary dx	Heroin overdose
76	38	16	1.365	0.002	0.081	96501	Secondary dx	Intentional heroin overdose
77	39	17	1.365	0.002	0.081	96501	Secondary dx	Poisoning by heroin(965.01)
7	5	18	15.934	0.034	0.074		Vitals	Pulse
8	82	19	15.576	0.001	0.072	X58XXXA	Primary dx	Exposure to other specified factors, initial encounter
202		20	0.413		0.068	T402X4A	Primary dx	Poisoning by other opioids
294		21	0.254		−0.068	E9803	Secondary dx	Poisoning by tranquilizers and other psychotropic agents
20	24	22	5.004	0.003	−0.064	F329	Secondary dx	Major depressive disorder, single episode
291		23	0.258		0.060	T40601A	Primary dx	Poisoning by unspecified narcotics, accidental (unintentional), initial encounter
101	33	24	0.997	0.002	0.060	E8502	Secondary dx	Accidental poisoning by other opiates and related narcotics(E850.2)
129	29	25	0.704	0.002	0.051	T401X2A	Primary dx	Poisoning by heroin, intentional self-harm, initial encounter
84	34	26	1.273	0.002	0.050	96500	Primary dx	Opiate overdose
26	83	27	3.125	0.001	−0.050	30000	Secondary dx	Anxiety state, unspecified
25	158	28	3.654	0.000	0.050	Y9289	Primary dx	Other specified places as the place of occurrence of the external cause
245		29	0.325		−0.046	E11319	Secondary dx	Type 2 diabetes
104	42	30	0.936	0.002	0.044	T50901A	Secondary dx	Poisoning by unspecified drugs

The model ranks are shaded for the top 30 of 717 predictors by model type and ordered by elastic-net rank. The models independently assigned high ranks to many of the same predictors. For example, a medication therapeutic class of "ANTIDOTES" was in the top 5 for all models and respiration rate was in the top 10. The data types for Chief Complaint, Medications, and Vitals frequently appeared among the top-ranked predictors, particularly for RF and XGBoost models; previous phenotypes relied solely on ICD codes. The Supplementary Material include tables ordered by XGBoost and RF ranks. Note that no ranks are given for some ENET variables because the model "shrank" those parameter estimates to 0. Importance statistics for each model type are: RF=mean decrease of the Gini Index; XGBoost=fractional contribution of each predictor; ENET=absolute value of parameter estimates. Data type colors: (yellow=chief complaint, green=primary or secondary diagnosis ICD code, blue=vital signs, tan=medication class, pink=positive lab result, purple=CPT code).

code). Some predictors were independently ranked as highly important by the 3 models. For example, a chief complaint for drug overdose was among the top 3 in all models; a medication in the antidotes class was among the top 5; respiration rate (vital signs) was among the top 10. For XGBoost, positive lab results for methadone, opiates, and amphetamines were among the top 30 predictors. Several CPT codes were also highly ranked. For ENET, a larger number of ICD codes were prioritized compared to the other models, but vitals (respirations and pulse), 2 medication classes, and 1 chief complaint were among the top 30. Note that for ENET results, a negative importance statistic indicates that predictor is negatively associated with an OOD. For example, a higher respiration rate or CPT codes for a CT scan or electrocardiograms were estimated to be less likely associated with an OOD.

Although prediction performance was similar among the 3 designs, we noted that RF importance rankings included 18 non-ICD code predictors among the top 30, while XGBoost and ENET had 14 and 5, respectively. To assess contributions by predictor type, we ran several additional RF models, each based on single predictor types; Figure 1 shows the ROC curves for these models. The primary diagnosis code and chief complaint were the most important contributors to model performance. Figure 2 provides a comparison of diagnostic performance by predictor type. Models based on primary diagnosis or medication class had the highest sensitivity, whereas models based on chief complaint, primary diagnosis, or vital signs had the highest specificity.

Figure 3 shows prediction performance for the full model compared to smaller models based on the top 2 predictor categories (primary diagnosis and chief complaint) when they were used either separately or together in models. The smaller model based on primary diagnosis and chief complaint performed nearly as well (AUC=0.97) as the full model (AUC=0.98). Figure 4 provides a comparison of ROC curves for the full model (AUC = 0.98) and the model limited to decision support data types (AUC = 0.94) that would be available in near real-time during the ED visit. Decision support data types included chief complaints, medications prescribed during the visit, vital signs, laboratory results; excluded data types were primary and secondary ICD codes (billing codes) and CPT codes since they were entered following the visit. The model based on decision support data types achieved 88% sensitivity, 89% specificity, and 89% positive predictive value.

Sensitivity analyses

Since several of the top predictors were "obvious" labels for a drug overdose (ie, chief complaint=drug overdose), we examined how strongly model predictions for opioid overdose

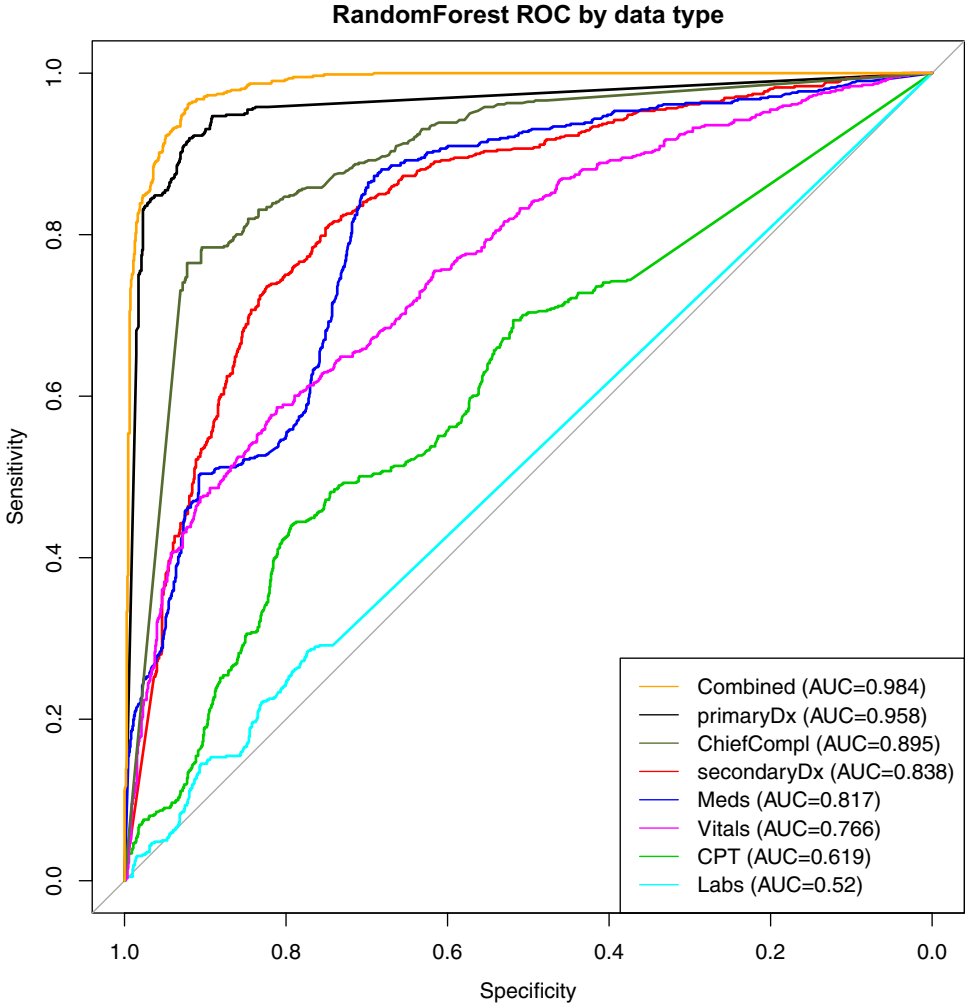

Figure 1. ROC curves by predictor category for RF models (other model types were similar). The primary diagnosis and chief complaint were the most important contributors to model performance.

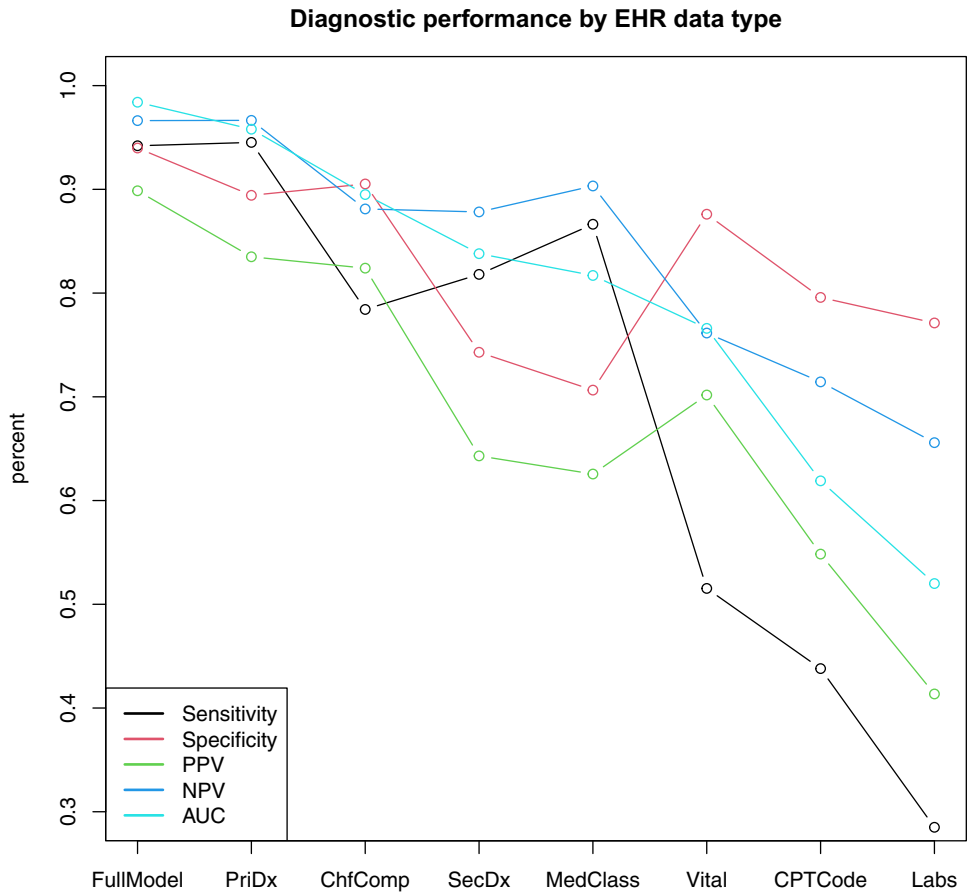

Figure 2. Diagnostic performance by EHR data type in RF models. Primary diagnosis and medication class were the EHR data types with greatest sensitivity. For specificity, chief complaint, primary diagnosis, and vital signs were the top performers.

depended on these predictors, though we also noted that previous published phenotypes appropriately included them for the stated purpose of identifying an OOD in EHR data.[3,4] We compared model results when 3 key predictors were all excluded: chief complaint of drug overdose, primary diagnosis code T401X1A (poisoning by heroin) and antidotes medication class. Each of these was among the top 5 predictors for each model design (Table 3 and Tables S1 and S2). For the chief complaint of drug overdose, 471 (75.8%) of true OOD cases and 83 (7.6%) of non-OOD cases had this complaint; out of 554 cases with this complaint, 83 (15%) were not true OODs. For the T401XA ICD-10 code, 313 (50.4%) of true OOD cases and 10 (0.9%) of non-OOD cases had this code; out of 323 cases with this code, 10 (3.1%) were not true OODs. For the "antidotes" medication class, 173 (27.9%) of true OOD cases and 19 (1.7%) of non-OD cases included this class; out of 192 cases with this class, 19 (9.9%) were not true OODs. Table S3 gives RF model results when these 3 features were excluded; model performance was not substantially affected; AUC dropped from 0.98 to 0.96, sensitivity from 0.94 to 0.92, specificity dropped from 0.93 to 0.88, and positive predictive value from 0.89 to 0.81.

The Supplementary Material also include a description of several subgroup analyses in which patients were excluded if they had some (or any) of these "obvious" overdose labels in order to better identify what other predictors were important in this group. In the subgroup that excluded patients with the top-5 most important "obvious" labels, the top-10 most important predictors from the RF model included 4 vital signs, 3 ICD codes related to suicide, poisoning, or place of occurrence, and 3 diagnosis codes directly related to OOD. We then examined the most limiting subgroup without any ICD-9 or -10 primary or secondary diagnosis directly related to an OOD, and found that only 0.5% of this group had an OOD. Thus, nearly all OOD cases had at least one ICD code related to an OOD, and it was not feasible to train and validate a predictive model using this subgroup.

Next, we checked whether our results were biased by multiple ED visits per patient, since the machine learning and penalized regression methods applied here did not account for multiple ED visits per patient (13% of our patients had more than one ED visit and 11% of those with an OOD had multiple visits). We created a subset of the full dataset in which each patient had only one visit by randomly selecting one row from each multiple-visit patient. Results from the RF model are shown in Table S2; overall AUC was unchanged, though specificity dropped slightly from 0.93 to 0.92 and positive predictive value fell from 0.89 to 0.86.

Finally, we checked whether knowledge of a previous ED visit involving an OOD would be important if added to the original RF model, which relied only on current ED visit EHR data. In Table S3, we see that AUC was unchanged, and sensitivity increased only slightly from 0.94 to 0.95. Table S4 provides a comparison between the original RF model and the new model that included the "previous OOD" variable; the new variable ranked only 15th in importance.

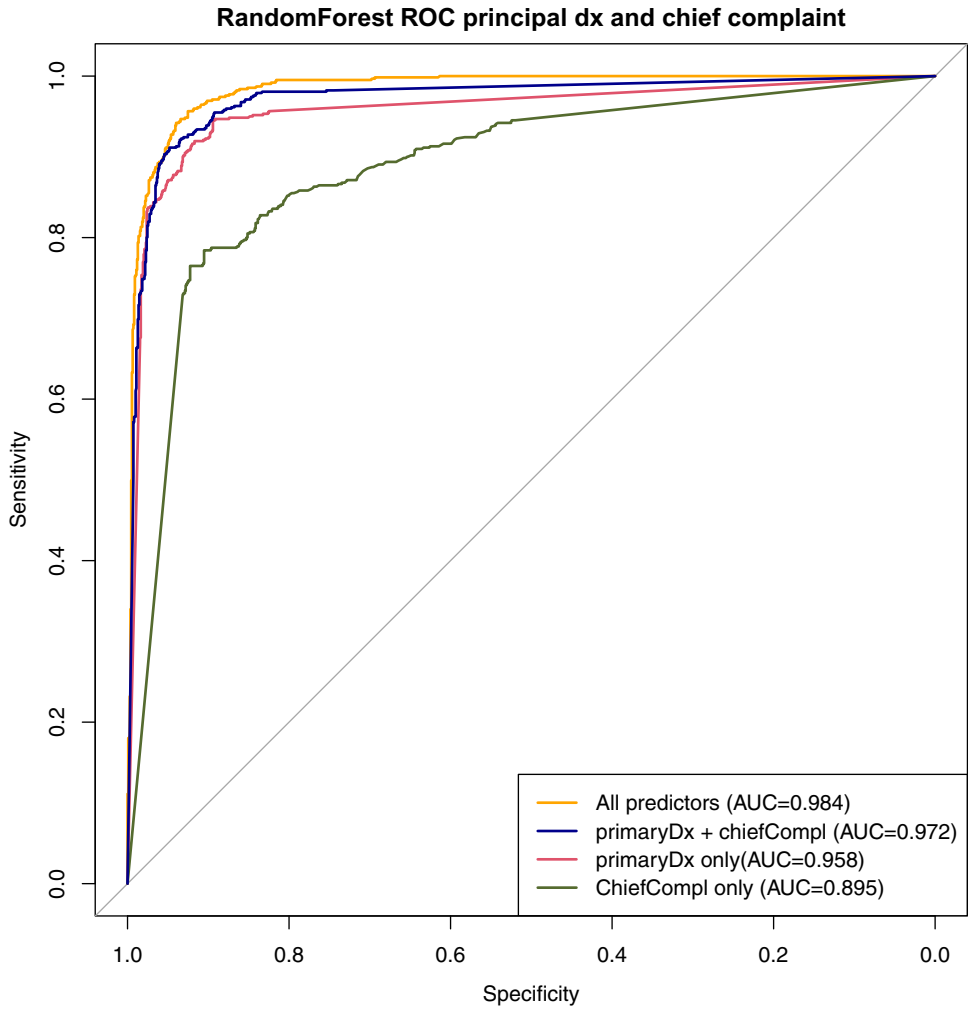

Figure 3. ROC curves based on the top 2 most important predictor types (RF). ROC curve for "top 2" (dark blue) shows only slightly degraded prediction performance compared to the full model with all predictor types (gold). Exporting this smaller model to other healthcare data systems would be less complex than the full model (which also includes medications, secondary diagnoses, CPT codes, labs, and vital signs).

Discussion

We sought to develop an improved phenotype for identifying an OOD in EHR data to boost the effectiveness of epidemiological surveillance efforts and to assist in healthcare management strategies, such as dispensation of naloxone kits after emergency department care. Green et al[4] had earlier developed an OOD phenotype based only on ICD-9 codes and reported sensitivity of 97.2%, specificity 84.6%, and PPV 87.4% on validation data. By expanding the data types to include chief complaints, primary, and secondary diagnoses using both ICD-CM-9 and -10 codes, medication classes, vital signs, CPT codes, and laboratory results, we were able to boost overall prediction performance: our machine learning and penalized regression models achieved sensitivity between 94% and 97%, specificity between 93% and 96%, positive predictive value between 89% and 93%, with 98% AUC for each model based on validation data. None of the 3 models (RF, XGBoost, and ENET) were distinctly superior to the others in prediction performance. We also assessed which data types were largest contributors to diagnostic performance, and found that the primary diagnosis ICD codes, coupled with the chief complaint (selected from a pull-down menu in our EHR system) were most important based on overall AUC, and the other data types may not be needed in many instances to achieve satisfactory performance. Related to this, we also found that nearly every OOD case had at least one ICD code directly related to OOD. This more "portable" model based on primary diagnosis and chief complaint would be far easier to implement in a new system compared with the full model. However, as shown in Figure 2, if specificity is important, then other data types such as vital signs may be useful (respiration rate, pulse rate, blood pressure). Incorporation of medication therapeutic class may help to boost sensitivity. We also demonstrated that our phenotype could be a useful input for decision support systems similar to one we previously described[5] since prediction performance remained fairly robust when limited to data types available in near real-time (chief complaints, medications prescribed during the visit, vital signs, laboratory results). Compared to the full model (RF), sensitivity declined from 92% to 88%, specificity declined from 95% to 89%, and positive predictive value declined from 91% to 89%. Finally, we ran multiple sensitivity checks that demonstrated our phenotype was robust to several concerns, particularly potential bias from patients with multiple ED visits.

Limitations

This was a single site study based on a population within a limited geographic region and was based on data extracted

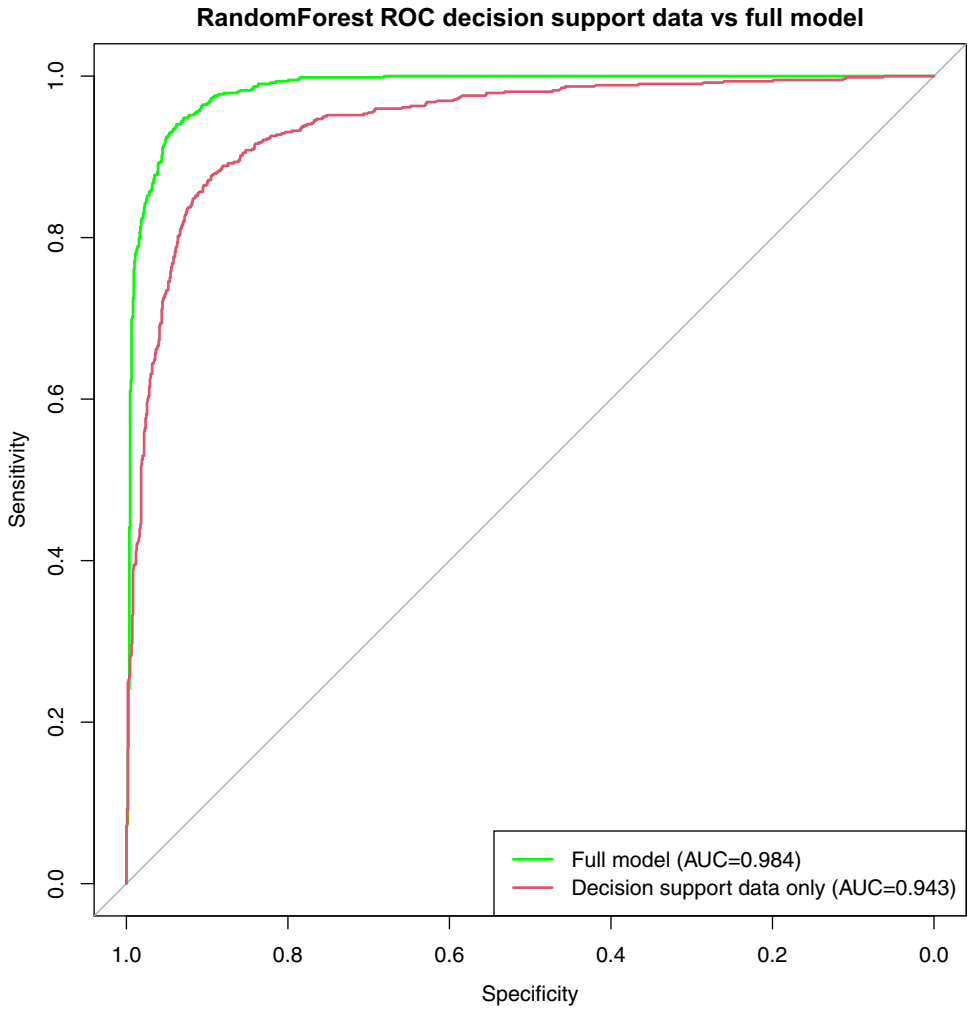

Figure 4. Comparison of RF ROC curves for full model (AUC = 0.984) and model limited to decision support data types (AUC = 0.943) that would be available in near real-time during the ED visit. Decision support data types included chief complaints, medications prescribed during the visit, vital signs, laboratory results; asynchronous data types excluded from this model were primary and secondary ICD codes (billing codes) and CPT codes. The model based on decision support data types achieved 88% sensitivity, 89% specificity, and 89% positive predictive value.

from a single EHR data system. Specific data types such as "Medication Therapeutic Class" may not be available in other systems, so tailored phenotype development may be required in other settings. Additional work is needed to determine if this phenotype is generalizable to other populations, and to determine if the models we developed here are portable to other EHR systems. Further, our models did not incorporate clinical text notes. Natural language processing methods could potentially boost predictive performance even further.

Conclusion

To our knowledge, our work provides an important update in methods for identifying OODs in EHR data with e-phenotypes that could be used both in retrospective empirical applications and in clinical decision support systems. We showed that incorporating new EHR data types can lead to substantial improvement in overall prediction performance.

Acknowledgments

The authors acknowledge the outstanding project support provided by the data management and project management teams at the Biomedical Informatics Center at MUSC.

Author contributions

Study design: R.W., J.S.O., L.J., J.L.M., and L.A.L. Acquisition and interpretation of data: R.W., J.S.O., L.J., W.G.H., R.P., C.B., S.F., AS, B.P., G.H.B., and J.L.M. Analysis: R.W. and J.S.O. Drafting of the manuscript and revision of content: all authors. Final approval of the manuscript and agreement to be accountable for all aspects of the work and resolution of all questions related to accuracy or integrity of the manuscript: all authors.

Supplementary material

Supplementary material is available at *JAMIA Open* online.

Funding

This publication was supported, in part, by the National Center for Advancing Translational Sciences of the National Institutes of Health under Grant Numbers U01TR002628, UL1TR001450, UL1TR001998, UL1TR001442, and UL1TR001086. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of interest

None declared.

Data availability

The data underlying this article cannot be shared publicly due to privacy requirements associated with protected healthcare information. A synthetic and fully deidentified version of the dataset will be shared on reasonable request to the corresponding author.

References

1. Centers for Disease Control and Prevention (CDC). *Provisional Drug Overdose Death Counts*. National Center for Health Statistics; 2023. https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm. Accessed March 6, 2023.
2. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med*. 2010; 2(57):57cm29.
3. Chartash D, Paek H, Dziura JD, *et al*. Identifying opioid use disorder in the emergency department: multi-system electronic health record-based computable phenotype derivation and validation study. *JMIR Med Inform*. 2019;7(4):e15794.
4. Green CA, Perrin NA, Hazlehurst B, *et al*. Identifying and classifying opioid-related overdoses: a validation study. *Pharmacoepidemiol Drug Saf*. 2019;28(8):1127-1137. https://doi.org/10.1002/pds.4772.
5. Badger J, LaRose E, Mayer J, Bashiri F, Page D, Peissig P. Machine learning for phenotyping opioid overdose events. *J Biomed Inform*. 2019;94:103185.
6. Jennings LK, Ward R, Pekar E, *et al*. The effectiveness of a noninterruptive alert to increase prescription of take-home naloxone in emergency departments. *J Am Med Inform Assoc*. 2023;30(4):683-691. https://doi.org/10.1093/jamia/ocac257.
7. Vivolo-Kantor A, Pasalic E, Liu S, Martinez PD, Gladden RM; Overdose Morbidity Team. Defining indicators for drug overdose emergency department visits and hospitalisations in ICD-10-CM coded discharge data. *Inj Prev*. 2021;27(S1):i56-i61. 10.1136/injuryprev-2019-043521.
8. Centers for Medicare and Medicaid Services. *ICD-9-CM to and from ICD-10-CM and ICD-10-PCS Crosswalk or General Equivalence Mappings*. 2020. Downloaded from https://www.nber.org/research/data/icd-9-cm-and-icd-10-cm-and-icd-10-pcs-crosswalk-or-general-equivalence-mappings. Accessed November 22, 2022.
9. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009; 42(2):377-381. https://doi.org/10.1016/j.jbi.2008.08.010.
10. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.
11. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189-1232.
12. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Soft*. 2010;33(1):1-22.
13. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3(1):32-35.
14. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021. https://www.R-project.org/. Accessed July 13, 2022.
15. Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2(3):18-22.
16. Chen T, He T, Benesty M, *et al*. *xgboost: Extreme Gradient Boosting*. R package version 1.6.0.1. 2022. https://CRAN.R-project.org/package=xgboost. Accessed June 12, 2022.