Received: 27 February 2022 Revised: 5 December 2022 Accepted: 22 December 2022
DOI: 10.1002/pds.5591

Development and validation of an electronic health records-based opioid use disorder algorithm by expert clinical adjudication among patients with prescribed opioids

Shabbar I. Ranapurwala[1,2] | Ishrat Z. Alam[1,2] | Brian W. Pence[1,2] | Timothy S. Carey[1,3,4] | Sean Christensen[5] | Marshall Clark[3] | Paul R. Chelminski[6] | Li-Tzy Wu[7,8] | Lawrence H. Greenblatt[8] | Jeffrey E. Korte[9] | Mark Wolfson[10] | Heather E. Douglas[11] | Lynn A. Bowlby[8] | Michael Capata[5] | Stephen W. Marshall[1,2]

[1]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, North Carolina, USA

[2]Injury Prevention Research Center, University of North Carolina, Chapel Hill, North Carolina, USA

[3]North Carolina Translational and Clinical Sciences Institute, School of Medicine, University of North Carolina at Chapel Hill, North Carolina, USA

[4]Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA

[5]Department of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, South Carolina, USA

[6]Division of General Internal Medicine and Clinical Epidemiology, Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA

[7]Department of Psychiatry and Behavioral Sciences, School of Medicine, Duke University, Durham, North Carolina, USA

[8]Department of Medicine, School of Medicine, Duke University, Durham, North Carolina, USA

[9]Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina, USA

[10]Department of Social Medicine, Population, and Public Health, School of Medicine, University of California, Riverside, California, USA

[11]Department of Psychiatry and Behavioral Medicine, School of Medicine, Wake Forest University, Winston-Salem, North Carolina, USA

Correspondence
Shabbar I. Ranapurwala, Epidemiology, 135 Dauer Drive, 2102A McGavran-Greenberg Hall, CB# 7435, Chapel Hill, NC, USA.
Email: sirana@email.unc.edu

Funding information
Centers for Disease Control and Prevention; National Center for Advancing Translational Sciences; National Institute on Drug Abuse

Abstract

Background: In the US, over 200 lives are lost from opioid overdoses each day. Accurate and prompt diagnosis of opioid use disorders (OUD) may help prevent overdose deaths. However, international classification of disease (ICD) codes for OUD are known to underestimate prevalence, and their specificity and sensitivity are unknown. We developed and validated algorithms to identify OUD in electronic health records (EHR) and examined the validity of OUD ICD codes.

Methods: Through four iterations, we developed EHR-based OUD identification algorithms among patients who were prescribed opioids from 2014 to 2017. The algorithms and OUD ICD codes were validated against 169 independent "gold standard" EHR chart reviews conducted by an expert adjudication panel across four healthcare systems. After using 2014–2020 EHR for validating iteration 1, the experts were advised to use 2014–2017 EHR thereafter.

Results: Of the 169 EHR charts, 81 (48%) were reviewed by more than one expert and exhibited 85% expert agreement. The experts identified 54 OUD cases. The experts endorsed all 11 OUD criteria from the Diagnostic and Statistical Manual of

Mental Disorders-5, including craving (72%), tolerance (65%), withdrawal (56%), and recurrent use in physically hazardous conditions (50%). The OUD ICD codes had 10% sensitivity and 99% specificity, underscoring large underestimation. In comparison our algorithm identified OUD with 23% sensitivity and 98% specificity.

Conclusions and relevance: This is the first study to estimate the validity of OUD ICD codes and develop validated EHR-based OUD identification algorithms. This work will inform future research on early intervention and prevention of OUD.

KEYWORDS

diagnostic and statistical manual of mental disorders-V, international classification of disease, opioid use disorder, validation

Key Points

- ICD diagnosis codes for opioid use disorders (OUD) are known to underestimate OUD prevalence, but reliability measures (specificity and sensitivity) are not available.
- This is the first study to estimate the sensitivity and specificity of OUD ICD codes.
- We found that the OUD ICD codes had 10% sensitivity and 99% specificity.
- We developed an algorithm to identify OUD in electronic health records data, which had up to 23% sensitivity and 98% specificity.
- We also developed an algorithm that excluded cancer patients, which had 16% sensitivity and 98% specificity.

Plain Language Summary:

It is understood that many patients with opioid use disorders (OUD) remain undiagnosed and therefore do not receive appropriate treatment. We developed and examined the accuracy of an algorithm to identify OUD among patients receiving opioid prescriptions in four large integrated healthcare systems in southern US. In addition, we examined the accuracy of OUD diagnoses codes that are commonly used in healthcare systems to understand the proportion of people that may remain undiagnosed. We found that the diagnoses codes only capture 10% of OUD cases and miss the remaining 90%. In comparison, our algorithms identify twice (23%) as many OUD cases, many of whom remain undiagnosed. Still, our algorithms miss about 77% of OUD cases. More work needs to be done to identify patients with OUD so that they can be provided adequate treatment and prevented from harms like overdose and death.

1 | INTRODUCTION

Each day more than 200 individuals die from opioid overdoses in the United States (US).[1,2] The overdose death estimates represent only the tip of the opioid epidemic iceberg, with >2 million Americans suffering from opioid use disorders (OUD), a term which encompasses addiction, abuse, and dependence. Another 10 million or more Americans are misusing opioids and are at risk of developing OUD.[3] Even as most opioid research to date has focused on preventing overdose deaths,[4–7] opioid overdose deaths from prescription and illicit opioids have increased.[1,2] Continued progress in combatting the opioid epidemic requires a shift towards earlier intervention to prevent and treat opioid use disorder.[8]

The limited research focus on OUD prevention derives from difficulties in reliably identifying OUD in large healthcare databases.[9,10] International classification of disease (ICD) codes for OUD are likely under-utilized because OUD can be challenging to identify clinically[11] and the stigma surrounding the documentation of an OUD diagnosis can negatively affect patients' insurance, employment, and medical care. As a result, ICD codes may have low sensitivity in identifying OUD and substantially underestimate the true prevalence of OUD.[12,13]

Clinical review of patients' medical records is an alternative to relying on ICD codes. This approach allows for more accurate identification of OUD than ICD codes alone, but it is a time-intensive strategy, beyond the resources for most research projects and infeasible in the large healthcare databases.[12]

The early and accurate identification of patients with OUD is critical in linking patients to treatment to prevent overdose deaths, develop OUD prevention strategies by examining its predictors, and reduce suffering and loss of productivity due to medical

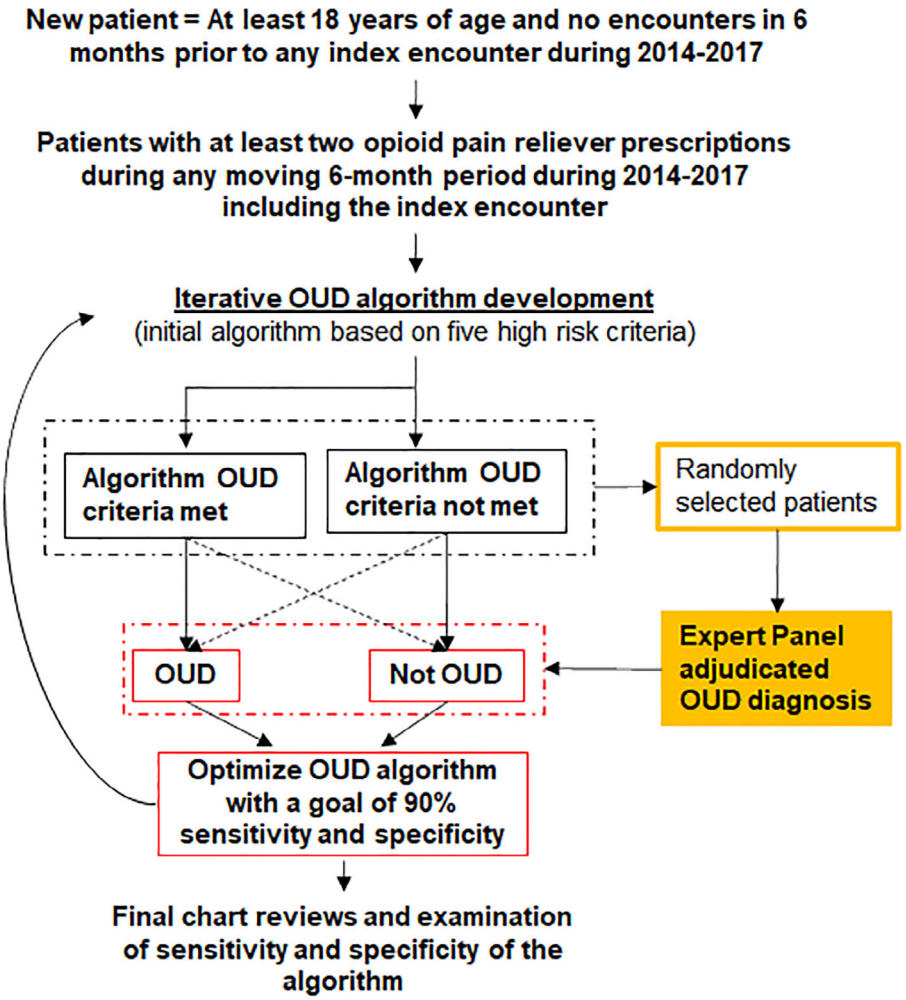

FIGURE 1 Patient selection and algorithm development

comorbidities.[9–13] Using electronic health record (EHR) data from four large healthcare systems in two US states, we iteratively developed and validated algorithms to identify OUD from EHR data. We used expert-adjudicated OUD diagnosis as a gold standard. We compared sensitivity and specificity for identifying OUD using our algorithms and using ICD codes alone.

2 | METHODS

We conducted a validation study in four large academic integrated healthcare systems including the University of North Carolina at Chapel Hill, Duke University, Wake Forest Baptist Health, and the Medical University of South Carolina. The study was approved by institutional review boards of all four sites. We used structured EHR data from 2014 to 2017 and variables available in the PCORnet® common data model[14] to increase the generalizability of the algorithm. The variables included age, gender, ICD 9/10 diagnosis codes, encounter information, medications, and healthcare common procedure coding system (HCPSC) codes.

2.1 | Gold standard

To develop and validate the algorithms to identify OUD and examine the validity of the ICD codes alone, we established a gold standard expert adjudication panel comprising two experts at each institution (8 total). The experts included psychiatrists with specialized medical training in substance use disorders, pain medicine specialists, general internal medicine physicians with experience in treating OUD, and substance use disorder treatment specialists. All experts reviewed and applied clinical judgement to the 11 criteria for OUD from the Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5); meeting any two of the 11 criteria is considered sufficient for a diagnosis of OUD (APA, 2013)–experts used clinical judgement to ascertain whether the patients met these criteria or not.[15]

2.2 | Algorithm building and validation

Figure 1 shows a schematic of the methods involved in developing the algorithm to identify OUD. We first identified a cohort of all new patients at least 18 years of age, defined as a patient with no medical

encounters in the 6 months prior to the first observed encounter (index encounter) between 2014 and 2017. Thus, we also used EHR data from July 1, 2013, to December 31, 2013, for the look-back, in addition to 2014–2017 data. We then identified new patients who received at least two opioid prescriptions during any moving 6-month period including the index encounter from 2014 to 2017. This represented 10.5% of all new patients seen at all four institutions.

We used this cohort of new patients with at least two opioid prescriptions in a 6-month period for developing algorithms to determine OUD status. Our initial (stage-1) algorithm included five criteria (Table 1), intentionally designed to be highly sensitive rather than specific. Patients who met at least one of these criteria during 2014–2017 were classified as probable OUD patients, while patients who met none of the criteria were classified as probable non-OUD patients.

We randomly selected 10 stage-1 patients, 5 probable OUD and 5 probable non-OUD, from each institution (40 total). The expert panel, who was blinded to the algorithm specifications and probable OUD classification, reviewed full clinical details and triage notes from these patients without the stage-1 classification results and adjudicated the cases. Each reviewer was allotted eight of the 10 site-specific cases such that six cases were reviewed by two reviewers at each site.

After each case was reviewed, the experts provided a decision of OUD diagnosis or no OUD diagnosis and reported the factors that met the DSM-5 OUD criteria or other factors they used to decide that the patient did or did not have an OUD. For cases that were reviewed by both reviewers, if any one of the reviewers classified it as OUD, we considered the gold standard to be OUD.

We then compared expert-adjudicated OUD diagnoses with our stage-1 algorithm results and calculated stage-1 sensitivity and specificity. To improve the algorithm's performance, we refined the stage-1 algorithm by incorporating the factors experts used for their decision making, particularly in discordant cases (algorithm positive/expert negative or vice versa). The resulting stage-2 algorithm was then utilized to repeat the whole process, including identifying probable OUD cases in the original cohort of patients, randomly selecting a new set of 10 patients from each site for expert adjudication, and examining validity of the algorithm. We repeated this process for four iterations. Details about the iterations are presented in Table 1 and coding can be accessed here: https://github.com/ShabbarIR/OUD-algorithms-development-and-validation.

Note that in iteration 1, instead of using only data from 2014 to 2017 for validation, the experts used the data from 2014 to 2020. This was a mistake, which was rectified by additional written instructions in iteration 2. Therefore, iterations 1 and 2 of the algorithm had the same specifications but used different time periods for validation, 2014–2020 and 2014–2017, respectively. Iterations 2, 3, and 4 were validated based on data from 2014 to 2017.

2.3 | Statistical analysis

We report positive and negative predictive values (PPV and NPV) of each stage of algorithm iteration, as well as for using ICD codes alone for OUD based just on the cases reviewed at that stage. Following this, we used the total sample of all patients with adjudicated OUD status from any of the four algorithm iterations as the full sample to evaluate the final performance of stage-3 and stage-4 algorithms as well as for ICD 9/10 codes alone. Since the expert adjudication was conducted in a sample of the charts, we used inverse probability of sampling proportion (IPSP) weights, using site-level sampling proportions from individual sites, to adjust the sensitivities and specificities of the algorithms, using methods described by Katki et al.[16] We present weighted sensitivity, specificity, PPV, and NPV with 95% confidence intervals (CI) and likelihood ratios (LR+/LR−).[17,18]

We estimated adjusted OUD prevalence using the sensitivity and specificity estimates to quantify the OUD prevalence underestimation from our algorithms and the ICD codes (Table 3). Prevalence estimates were adjusted using the following formula[19]:

$$Adjusted Prevalence = \frac{Crude Prevalence + Specificity - 1}{Sensitivity + Specificity - 1},$$

We used Cohen's kappa[20] to examine agreement between the reviewers. One reviewer dropped out during stage-2 leaving only one reviewer at one of the sites. At that site Cohen's kappa was calculated using the stage-1 results only.

To ensure consistency throughout the study procedures, EHR extraction, programming, and analysis steps were developed at one institution, and then the SQL/SAS codes were shared with the other three institutions.

We also present the frequency with which the expert reviewers endorsed each of the 11 DSM-5 criteria, and the frequency of select themes and words used to describe the reasons for adjudicating a chart as OUD.

Lastly, we conducted four sensitivity analyses: (1) we excluded iteration 1 altogether since the expert reviewers had use of charts beyond 2017 in evaluation of OUD; (2) we changed the definition of OUD gold standard, such that cases reviewed by both the reviewers were classified as OUD only if both reviewers agreed, else they were classified as non-OUD; (3) we examined the validity of our algorithm and ICD codes to identify moderate to severe OUD cases defined as an OUD case where the expert reviewers endorsed at least four DSM-5 criteria; and (4) we calculated site specific sensitivities and specificities for our algorithms and ICD codes.

3 | RESULTS

The expert adjudication panel reviewed 169 charts to validate the algorithms and the ICD codes. During stage-2, one reviewer at a site dropped out, so only eight charts were reviewed there. During stage-3 and stage-4 iterations, there was a programming error at another site, which necessitated the review of an additional 11 charts

TABLE 1 Opioid use disorder algorithm iterations, definitions, and performance compared to expert panel adjudication

Iteration	# charts reviewed	Criteria	Positive predictive value (True+/algorithm+)	Negative predictive value (True−/algorithm−)
1	40	Any of: A, B, C, D, E	50% (10/20)	90% (18/20)
	40	A only (ICD codes)	71% (2/7)	79% (26/33)
2[a]	38	Any of: A, B, C, D, E	32% (6/19)	89% (17/19)
	38	A only (ICD codes)	50% (4/8)	87% (26/30)
3	46	Any of: A, D, F, G, H	71% (15/21)	88% (22/25)
	46	A only (ICD codes)	100% (12/12)	82% (28/34)
	46	Any of: A, D, G, H	83% (15/18)	89% (25/28)
4[b]	45	Any of: A, D, F, G, I	65% (15/23)	95% (21/22)
	45	A only (ICD codes)	85% (11/13)	84% (27/32)
	45	Any of: A, D, G, I	70% (14/20)	92% (23/25)

Note: A. ICD9/10 codes for abuse/dependence. B. transition to buprenorphine, methadone, and naltrexone that were prescribed to treat OUD OR co-prescription of Naloxone to prevent overdose. C. ≥50 MME of opioids for ≥180 days. D. procedure codes for Medication Assisted Treatment for OUD. E. ≥1 overdose event during the study period. F. ≥90 MME of opioids for ≥180 days. G. 'B'+'E'. H. 3+ ED visits with opioid RX in 30-day window. I. 'H,' excluding ED visits where surgery was performed up to 2 weeks before visit.

[a]Iteration 2 had the same criteria as iteration 1, except that in iteration 2, we provided additional instructions to the expert adjudication panel that all abstraction should be restricted to the years 2014–2017 only.

[b]Excludes patients receiving medications for metastatic cancers or have cancer diagnoses.

during those two iterations for that particular site only. The Cohen's kappa for agreement between the site experts was 0.65 (95% confidence interval: 0.46–0.83). The average observed agreement between reviewers was 85.2% with a minimum of 83.3%.

The expert reviewers identified 54 cases of OUD out of the 169 chart reviews. All DSM-5 criteria were endorsed by the reviewers with high frequency (Supplementary Table 1). The most common DSM-5 criteria endorsed among the 54 OUD cases were craving (72%), tolerance (65%), withdrawal (56%), and recurrent use in physically hazardous conditions (50%). The expert reviewers also noted additional factors that helped them deduce OUD cases including intravenous drug use (e.g., heroin, cocaine, etc.), overdoses, emergency visits, medication assisted treatment, and urine drug screening results (Supplementary Table 2). Experts noted ICD codes for OUD in only a handful of cases (Supplementary Table 2). Conversely, the experts noted that 84% of not-OUD included those receiving clinically indicated opioids for chronic pain, surgery, trauma, and cancer, even "while patient showed tolerance and withdrawal symptoms" as experts noted.

The stage-3 unweighted PPV (71%) and NPV (88%) along with increased PPV and NPV suggest that it is the most balanced algorithm (Table 1) and stage-4 algorithm, which excludes cancer patients, has the highest unweighted NPV (95%). The ICD-based OUD definition produced a 100% unweighted PPV (at stage-3) and 82% unweighted NPV suggesting underestimation of OUD prevalence.

However, when we used IPSP weights to adjust for the stratified sampling, we found that the sensitivities of both stage-3 (23.5%) and stage-4 (16.7%) algorithms and the ICD-based OUD definition (10.3%) were very low (Table 2). The specificity of both the algorithms and the ICD-based OUD definition were 98.3%, 98.5%, and 99.5%, respectively. Thus, compared to ICD-codes, a small decrease in specificity increased the sensitivity of the OUD algorithms by a factor of 1.6–2.3 times. The sensitivity of the ICD codes declined further when OUD among cancer patients were excluded during stage-4 algorithm, however the specificity improved marginally (Table 2). For both stage-3 and stage-4 algorithms, removing the '≥90 morphine milligram equivalents (MME) of opioids for ≥180 days' criterion reduced sensitivity, but increased specificity; the reduced sensitivity was still higher than that of the ICD codes in isolation (Table 2).

Sensitivity analyses by excluding the iteration 1, excluding OUD case disagreement where both reviewers reviewed cases, and examining moderate to severe OUD (at least four DSM-5 criteria) showed that the results are robust for both iterations 3 and 4, and for ICD codes (Table 3). Low prevalence of OUD at one of the sites and small samples per site resulted in the variability in site-specific algorithm sensitivity for one site (Supplementary Table 3). However, the algorithm specificity and ICD code sensitivity and specificity were robust between all sites (Supplementary Table 3).

The weighted PPV and NPV estimates suggest (Table 2), that even at seemingly high values of PPV and NPV, the ICD codes and our algorithms can have low sensitivity and greatly underestimate the true prevalence of disease (Table 4). Based on the low weighted sensitivity of the algorithms, the overall prevalence of OUD was estimated to be between 14–15% (Table 3) among new patients who received at least two opioid prescriptions during a 6-month period subsequent to their index encounter between 2014–2017. Table 4 shows that the OUD ICD codes underestimate the OUD prevalence by a factor of 10, or almost 11 times underestimation if we exclude the cancer population. The crude prevalence estimate from the stage-3 algorithm was found to be the least underestimated, and yet it underestimated OUD prevalence by over 5 times (Table 4).

TABLE 2 Validity of stage 3 and 4 iterations and ICD9/10 codes for OUD using all charts and inverse probability of sampling proportion weights and 95% confidence intervals

Iteration	Criteria	Sensitivity[a] (95% CI)	Specificity[a] (95% CI)	PPV[a] (95% CI)	NPV[a] (95% CI)	Likelihood ratio+	Likelihood ratio−
3	Any of: A, D, F, G, H	23.04% (22.64–23.44)	98.36% (98.32–98.40)	61.66% (60.90–62.42)	91.79% (91.70–91.87)	14.05	0.78
	A only (ICD codes)	10.34% (10.10–10.59)	99.54% (99.52–99.57)	81.68% (80.80–82.56)	84.91% (84.79–85.03)	22.48	0.90
	Any of: A, D, G, H	17.99% (17.62–18.35)	98.79% (98.75–98.82)	62.96% (62.10–63.82)	91.33% (91.24–91.41)	14.86	0.83
4[b]	Any of: A, D, F, G, I	16.37% (16.03–16.71)	98.55% (98.51–98.59)	58.16% (57.31–59.02)	90.53% (90.44–90.62)	11.29	0.85
	A only (ICD codes)	10.11% (9.86–10.37)	99.66% (99.64–99.68)	84.38% (83.48–85.28)	85.83% (85.71–85.95)	29.74	0.90
	Any of: A, D, G, I	13.89% (13.57–14.20)	98.94% (98.91–98.98)	61.80% (60.86–62.74)	90.31% (90.22–90.40)	13.10	0.87

Note: A. ICD9/10 codes for abuse/dependence. B. transition to buprenorphine, methadone, and naltrexone that were prescribed to treat OUD OR co-prescription of Naloxone to prevent overdose. C. ≥50 MME of opioids for ≥180 days. D. procedure codes for Medication Assisted Treatment for OUD. E. ≥1 overdose event during the study period. F. ≥90 MME of opioids for ≥180 days. G. 'B'+'E'. H. 3+ ED visits with opioid RX in 30-day window. I. 'H,' excluding ED visits where surgery was performed up to 2 weeks before visit.

Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

[a]Weighted using site-specific inverse probability of sampling weights.

[b]Excludes patients receiving medications for metastatic cancers or have cancer diagnoses.

4 | DISCUSSION

OUD has significant impact on patients, families, and healthcare delivery systems.[21] Improved OUD case identification may help intervene early in the opioid use cycle, thereby facilitating targeted OUD and overdose prevention. This is the first study to define and validate EHR-based OUD algorithms using expert clinical adjudication as a gold standard, and the first study to estimate the validity of ICD-based OUD definitions. Prior studies have found ~58%–62% PPV for the ICD-based OUD definition, but none so far have examined sensitivity of these codes.[9,10] Our study underscores how seemingly reasonable estimates of PPV may be associated with very low sensitivity and underestimate prevalence. With little loss of specificity, our algorithms improve upon the low sensitivity of ICD-based OUD definition and provide the foundation to build more sensitive algorithms in future. The robust findings of our algorithm and its future refinements could help healthcare delivery systems identify patients with probable OUD who would benefit from further evaluation and linkage to appropriate care.

The use of the highly specific ICD-based OUD definition yields fewer false positive diagnoses; however, its low-sensitivity misses most patients with OUD (high false negative rates). This is problematic in devising robust clinical and population-based responses to stem the opioid epidemic. Implementing an algorithm with greater sensitivity optimizes opportunities for treating a condition with low prevalence such as OUD.[12,13] The slightly lower specificity of these algorithms underscores the importance of further clinical assessment in determining appropriate care. Experienced clinicians are accustomed to either confirming or excluding diagnoses based on first pass diagnostic blood testing which is sensitive but not specific. Multiple algorithm versions permit preferred balance of sensitivity and specificity for a specific research or clinical purpose.

This study is the first to allow for stratification of the algorithm to include or exclude cancer patients who are being treated with opioids. Management of pain for cancer patients is complex, often involving collaborative care models to prescribe opioids for pain management in hospice settings. The ability to exclude cancer patients and evaluate OUD among non-cancer patients is therefore invaluable.[4,5]

Our adjusted OUD prevalence estimates may seem much higher than estimates from the National Survey on Drug Use and Health (NSDUH) of self-reported OUD and opioid misuse prevalence in the general population.[3] However, our estimates are among people who received at least two opioid prescriptions in a six-month period (10.5% of all new patients). Adjusting for this population selection, our estimates of OUD prevalence in the general population would be slightly lower than NSDUH's estimates,[3] which are perhaps more clinically relevant than the NSDUH estimates.

Limitations: First, our algorithms may not be sensitive to people who primarily use illicit opioids, because these individuals may not consistently encounter the healthcare system. The expert reviewer notes indicated that OUD patients with illicit opioid use such as injection drug use, heroin, and cocaine were captured in this data, albeit their prevalence may be lower in EHR data. Yet, prescription opioid misuse is often a gateway to using illicit opioids,[6–8] and early identification of prescription-related OUD may help prevent new illicit opioid use. Further, there are no datasets yet

TABLE 3 Comparison of sensitivity and specificity of OUD algorithm and ICD codes from the main analysis to that from the sensitivity analyses

		Main analyses	Exclude iteration 1	OUD only where both reviewers agree[a]	OUD only if at least four DSM criteria were met
Stage 3	Se (95% CI)	23.04% (22.64–23.44)	24.60% (24.08–25.12)	25.12% (24.58–25.66)	26.90% (26.39–27.51)
	Sp (95% CI)	98.36% (98.32–98.40)	98.64% (98.60–98.69)	97.51% (97.47–97.56)	97.35% (97.30–97.40)
ICD at stage 3	Se (95% CI)	10.34% (10.10–10.59)	12.24% (11.92–12.55)	11.75% (11.44–12.06)	17.09% (16.65–17.53)
	Sp (95% CI)	99.54% (99.52–99.57)	99.64% (99.62–99.67)	99.17% (99.14–99.21)	99.21% (99.18–99.24)
Stage 4	Se (95% CI)	16.37% (16.03–16.71)	17.13% (16.70–17.55)	14.41% (14.00–14.82)	15.53% (15.07–15.99)
	Sp (95% CI)	98.55% (98.51–98.59)	98.74% (98.70–98.78)	97.73% (97.68–97.78)	97.67% (97.63–97.72)
ICD at stage 4	Se (95% CI)	10.11% (9.86–10.37)	11.35% (11.04–11.66)	12.07% (11.73–12.41)	15.67% (15.24–16.11)
	Sp (95% CI)	99.66% (99.64–99.68)	99.63% (99.60–99.65)	99.33% (99.30–99.36)	99.35% (99.32–99.37)

Abbreviations: DSM, diagnostic and statistical manual of mental Di; ICD, international classification of diseases; OUD, opioid use disorder; Se, sensitivity; Sp, specificity.

[a]This applies only to cases where two reviewers reviewed a case–if both reviewers considered the case to be OUD then the case was considered to be OUD; if one reviewer considered the case to be OUD and another did not, OR both reviewers considered the case to be non-OUD, then the case was considered to be non-OUD; if only one reviewer reviewed the case, then that reviewer's adjudication was considered as it is.

TABLE 4 OUD prevalence estimates from four healthcare systems, 2014–2107

Algorithms	Cohort	Estimated OUD cases	Unadjusted Prevalence (% and 95% CI)	Adjusted[b] prevalence (% and 95% CI)	Adjusted/crude
Stage 3	371 019	9716	2.62 (2.11, 3.26)	14.3 (6.39, 31.8)	5.5
ICD9/10 at Stage 3	371 019	7889	2.13 (1.60, 2.83)	21.3 (14.8, 30.8)	10.0
Stage 4[a]	371 019	7488	2.02 (1.61, 2.53)	15.3 (6.41, 36.4)	7.6
ICD9/10 at Stage 4[a]	371 019	6339	1.71 (1.28, 2.27)	18.6 (10.4, 33.2)	10.9

Abbreviations: CI, confidence intervals; OUD, opioid use disorders.

[a]Excluding OUDs among patients with cancer.

[b]Adjusted for site-specific weighted sensitivity and specificity derived using inverse probability of sampling proportion weights.

that can help identify OUD patients who do not interact with the healthcare system. Hence, an EHR-based algorithm may be one of the best methods for early identification and linkage to care for people with OUD.

Second, some of the criteria used in our algorithms may change over time due to the dynamic nature of the opioid epidemic. However, incorporating such algorithms in healthcare delivery systems can allow future improvements through machine learning and artificial intelligence methods.

Third, EHR data may have limited longitudinal follow up which may cause low sensitivity of our algorithms. At one site, among patients with acute, post-surgical, chronic, and cancer pain between 2014 and 2017, 50%–70% had at least 12 months longitudinal follow-up. Insurance claims data provide better longitudinal follow-up and potentially improve sensitivity. However, claims data may lack detailed ICD diagnoses codes, out-of-pocket medications, and do not have clinical notes to validate claims-based algorithms. Further, the implementation of OUD algorithms in EHR data, rather than claims, have more clinical value in aiding providers in making treatment decisions.

Fourth, it was not possible to examine OUD onset since this is a gradual process, and secondary retrospective data are not equipped for examining disease onset. Future longitudinal cohort studies with prospective data collection will be needed to examine onset.

Finally, our algorithms are validated in four healthcare systems. Though they represent diverse patient populations, further testing in healthcare settings across the US is necessary to improve generalizability. The algorithms are primed for such testing since we used the PCORnet common data model to develop them.

This work will fuel new OUD-focused research to help identify clinical prescribing strategies to prevent OUD.[4–7] The sensitivity and specificity of the ICD codes for opioid abuse and dependence OUD will allow adjustment of findings from previous studies making them more internally valid.[17,22] Future algorithm refinements include machine learning approaches for text mining, inclusion of urine testing results,[23] and comparison with ICD-10 OUD codes alone.

5 | CONCLUSIONS

Using expert clinical adjudication as a gold standard, we underscore the substantial underestimation of OUD prevalence when using ICD-based OUD definitions in EHR data and present the validity of more refined EHR-based OUD algorithms. These estimates and algorithms will allow adjustment of findings from previous studies using quantitative bias correction methods[17,22] and facilitate new research focused on early intervention on OUD prevention and treatment to respond to the ongoing opioid epidemic.[4–7]

AUTHOR CONTRIBUTIONS

Conceptualization: Shabbar I. Ranapurwala, Stephen W. Marshall.

Methodology: Shabbar I. Ranapurwala, Brian W. Pence, Stephen W. Marshall.

Implementation: Shabbar I. Ranapurwala, Ishrat Z. Alam, Marshall Clark, Timothy S. Carey, Sean Christensen, Paul R. Chelminski, Li-Tzy Wu, Lawrence H. Greenblatt, Jeffrey E. Korte, Mark Wolfson, Heather E. Douglas, Lynn A. Bowlby, Michael Capata.

Data abstraction: Shabbar I. Ranapurwala, Marshall Clark, Li-Tzy Wu, Jeffrey E. Korte, Mark Wolfson.

Data analyses: Shabbar I. Ranapurwala, Ishrat Z. Alam.

Interpretation: Shabbar I. Ranapurwala, Brian W. Pence, Timothy S. Carey, Li-Tzy Wu, Jeffrey E. Korte, Stephen W. Marshall.

First draft writing: Shabbar I. Ranapurwala, Brian W. Pence, Sean Christensen, Timothy S. Carey. Writing and revising subsequent drafts: All authors.

ACKNOWLEDGEMENTS

This project was supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (UL1TR002489, UL1TR002553, UL1TR001450, UL1TR001420) and the Duke Endowment through a pilot grant. Shabbar I. Ranapurwala, Ishrat Z. Alam, Stephen W. Marshall, Timothy S. Carey, Paul R. Chelminski, and Li-Tzy Wu were supported by a Centers for Disease Control and Prevention grant (R01CE003009). Brian W. Pence and Shabbar I. Ranapurwala were also supported by a National Institute for Drug Abuse grant (R21DA046048). Stephen W. Marshall was partly supported by a grant (R49CE003092) from CDC's National Center for Injury Prevention and Control for an Injury Control Research Center.

CONFLICT OF INTEREST

Tha author declares there is no potential conflict of interest.

DATA AVAILABILITY STATEMENT

Electronic health records data used for this study can be obtained by making data applications at each of the four healthcare systems included in this study. All other data and methods for this study are freely available here: https://github.com/ShabbarIR/OUD-algorithms-development-and-validation.

ORCID

Shabbar I. Ranapurwala https://orcid.org/0000-0002-3944-3912

REFERENCES

1. Stephenson J. Drug overdose deaths head toward record number in 2020, CDC warns. *JAMA Health Forum*. 2020;1(10):e201318. doi:10.1001/jamahealthforum.2020.1318
2. CDC Health Advisory. Increase in Fatal Drug Overdoses Across the United States Driven by Synthetic Opioids Before and During the COVID-19 Pandemic. Distributed via the CDC Health Alert Network. December 17, 2020, 8:00 AM ET. CDCHAN-00438 https://emergency.cdc.gov/han/2020/han00438.asp.
3. Substance Abuse and Mental Health Services Administration. (2019). Results from the 2018. National Survey on Drug Use and Health: Detailed tables. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. https://www.samhsa.gov/data/.
4. Ranapurwala SI, Naumann RB, Austin AE, Dasgupta N, Marshall SW. Methodologic limitations of prescription opioid safety research and recommendations for improving the evidence base. *Pharmacoepidemiol Drug Saf*. 2019;28(1):4-12. doi:10.1002/pds.4564
5. Staffa J, Meyer T, Secora A, McAninch J. Commentary on "methodologic limitations of prescription opioid safety research and recommendations for improving the evidence base". *Pharmacoepidemiol Drug Saf*. 2019 Jan;28(1):13-15. doi:10.1002/pds.4650
6. Dowell D, Haegerich TM, Chou RCDC. Guideline for prescribing opioids for chronic pain—United States. *MMWR*. 2016;65(1):1-49.
7. Chou R, Deyo R, Devine B, Hansen R, Sullivan S, Jarvik J. G, Blazina I, Dana T, Bougatsos C, Turner J. The Effectiveness and Risks of Long-Term Opioid Treatment of Chronic Pain. Evidence Report/Technology Assessment No. 218. (Prepared by the Pacific Northwest Evidence-Based Practice Center under Contract No. 290–2012-00014-I.) AHRQ Publication No. 14-E005-EF. Rockville, MD: Agency for Healthcare Research and Quality; 014. www.effectivehealthcare.ahrq.gov/reports/final.cfm.
8. Collins FS, Koroshetz WJ, Volkow ND. Helping to end addiction over the long-term: the research plan for the NIH HEAL initiative. *JAMA*. 2018;320(2):129-130. doi:10.1001/jama.2018.8826
9. Howell BA, Abel EA, Park D, Edmond SN, Leisch LJ, Becker WC. Validity of incident opioid use disorder (OUD) diagnoses in administrative data: a chart verification study. *J Gen Intern Med*. 2021;36(5):1264-1270. doi:10.1007/s11606-020-06339-3
10. 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;1(221):108583. doi:10.1016/j.drugalcdep.2021.108583
11. Stein BD, Sorbero M, Dick AW, Pacula RL, Burns RM, Gordon AJ. Physician capacity to treat opioid use disorder with buprenorphine-assisted treatment. *JAMA*. 2016;316:1211-1212.
12. Wu LT, McNeely J, Subramaniam GA, et al. DSM-5 substance use disorders among adult primary care patients: results from a multisite study. *Drug Alcohol Depend*. 2017;179:42-46. doi:10.1016/j.drugalcdep.2017.05.048
13. Hallgren KA, Witwer E, West I, et al. Prevalence of documented alcohol and opioid use disorder diagnoses and treatments in a regional primary care practice-based research network. *J Subst Abuse Treat*. 2020;110:18-27. doi:10.1016/j.jsat.2019.11.008
14. Garza M, del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. *J Biomed Inform*. 2016;64:333-341. doi:10.1016/j.jbi.2016.10.016
15. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub; 2013.
16. Katki HA, Li Y, Edelstein DW, Castle PE. Estimating the agreement and diagnostic accuracy of two diagnostic tests when one test is conducted on only a subsample of specimens. *Stat Med*. 2012;31(5):436-448.
17. Marshall RJ. Validation study methods for estimating exposure proportions and odds ratios with misclassified data. *J Clin Epidemiol*. 1990;43:941-947.
18. Morgan DJ, Pineles L, Owczarzak J, et al. Accuracy of practitioner estimates of probability of diagnosis before and after testing. *JAMA Intern Med*. 2021;181(6):747-755. doi:10.1001/jamainternmed.2021.0269
19. Buderer NM. Statistical methodology: I. incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med*. 1996;3(9):895-900.
20. Kraemer HC. Kappa coefficient. *Wiley StatsRef Stat Ref*. 2014;14:1-4.

21. Center for Substance Abuse Treatment. *Medications for Opioid Use Disorder*. Substance Abuse and Mental Health Services Administration (US); 2020 https://store.samhsa.gov/sites/default/files/SAMHSA_Digital_Download/PEP20-02-01-006_508.pdf
22. Hubbard RA, Tong J, Duan R, Chen Y. Reducing bias due to outcome misclassification for epidemiologic studies using EHR-derived probabilistic phenotypes. *Epidemiology*. 2020;31(4):542-550. doi:10.1097/EDE.0000000000001193
23. Larochelle MR, Cruz R, Kosakowski S, et al. Do urine drug tests reveal substance misuse among patients prescribed opioids for chronic pain? *J Gen Intern Med*. 2022;37(10):2365-2372.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ranapurwala SI, Alam IZ, Pence BW, et al. Development and validation of an electronic health records-based opioid use disorder algorithm by expert clinical adjudication among patients with prescribed opioids. *Pharmacoepidemiol Drug Saf*. 2023;32(5):577-585. doi:10.1002/pds.5591