





## ORIGINAL REPORT

# Identifying and classifying opioid-related overdoses: A validation study

Carla A. Green<sup>1</sup>  | Nancy A. Perrin<sup>1,9</sup> | Brian Hazlehurst<sup>1</sup>  | Shannon L. Janoff<sup>1</sup> | Angela DeVeaugh-Geiss<sup>2</sup> | David S. Carrell<sup>4</sup>  | Carlos G. Grijalva<sup>5</sup> | Caihua Liang<sup>6,7</sup>  | Cheryl L. Enger<sup>6,7</sup> | Paul M. Coplan<sup>3,8</sup>

<sup>1</sup>Center for Health Research, Kaiser Permanente Northwest, Portland, Oregon

<sup>2</sup>Indivior, Inc. North Chesterfield, Virginia

<sup>3</sup>Epidemiology, Johnson & Johnson, New Brunswick, New Jersey

<sup>4</sup>Health Research Institute, Kaiser Permanente Washington, Seattle, Washington

<sup>5</sup>Department of Health Policy, Vanderbilt University Medical Center, Nashville, Tennessee

<sup>6</sup>Epidemiology, Optum, Boston, Massachusetts

<sup>7</sup>Epidemiology, Optum, Ann Arbor, Michigan

<sup>8</sup>Adjunct, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

<sup>9</sup>Johns Hopkins School of Nursing, Johns Hopkins University, Baltimore, Maryland

## Correspondence

B. Hazlehurst, Kaiser Permanente Center for Health Research, 3800 N. Interstate Avenue, Portland, OR 97227-1110.  
Email: brian.hazlehurst@kpchr.org

## Funding information

Research Sponsors: This project was conducted as part of a Food and Drug Administration (FDA)-required postmarketing study for extended-release and long-acting opioid analgesics and was funded by the Opioid Postmarketing Consortium (OPC) consisting of the following companies at the time of study conduct: Allergan; Assertio Therapeutics, Inc.; BioDelivery Sciences, Inc.; Collegium Pharmaceutical, Inc.; Daiichi Sankyo, Inc.; Egalet Corporation; Endo Pharmaceuticals, Inc.; Hikma Pharmaceuticals USA Inc.; Janssen

## Abstract

**Purpose:** The study aims to develop and validate algorithms to identify and classify opioid overdoses using claims and other coded data, and clinical text extracted from electronic health records using natural language processing (NLP).

**Methods:** Primary data were derived from Kaiser Permanente Northwest (2008–2014), an integrated health care system (~n > 475 000 unique individuals per year). Data included *International Classification of Diseases, Ninth Revision (ICD-9)* codes for nonfatal diagnoses, *International Classification of Diseases, Tenth Revision (ICD-10)* codes for fatal events, clinical notes, and prescription medication records. We assessed sensitivity, specificity, positive predictive value, and negative predictive value for algorithms relative to medical chart review and conducted assessments of algorithm portability in Kaiser Permanente Washington, Tennessee State Medicaid, and Optum.

**Results:** Code-based algorithm performance was excellent for opioid-related overdoses (sensitivity = 97.2%, specificity = 84.6%) and classification of heroin-involved overdoses (sensitivity = 91.8%, specificity = 99.0%). Performance was acceptable for code-based suicide/suicide attempt classifications (sensitivity = 70.7%, specificity = 90.5%); sensitivity improved with NLP (sensitivity = 78.7%, specificity = 91.0%). Performance was acceptable for the code-based substance abuse-involved classification (sensitivity = 75.3%, specificity = 79.5%); sensitivity improved with the NLP-enhanced algorithm (sensitivity = 80.5%, specificity = 76.3%). The opioid-related overdose algorithm performed well across portability assessment sites, with sensitivity greater than 96% and specificity greater than 84%. Cross-site sensitivity for heroin-involved overdose was greater than 87%, specificity greater than or equal to 99%.

**Conclusions:** Code-based algorithms developed to detect opioid-related overdoses and classify them according to heroin involvement perform well. Algorithms for classifying suicides/attempts and abuse-related opioid overdoses perform adequately for

Research sponsors: Member Companies of the Opioid PMR Consortium.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. *Pharmacoepidemiology and Drug Safety* Published by John Wiley & Sons Ltd.

Pharmaceutic; Mallinckrodt Inc.; Pernix Therapeutics Holdings, Inc.; Pfizer, Inc.; Purdue Pharma, LP

use for research, particularly given the complexity of classifying such overdoses. The NLP-enhanced algorithms for suicides/suicide attempts and abuse-related overdoses perform significantly better than code-based algorithms and are appropriate for use in settings that have data and capacity to use NLP.

#### KEYWORDS

abuse, algorithms, heroin, methods, opioid overdose, pharmacoepidemiology, suicide

## 1 | INTRODUCTION

Opioid use disorders and fatal and nonfatal opioid-related overdoses (OODs) are significant public health problems.<sup>1-8</sup> Initiatives to reduce prescription opioid-related risks include clinical guidelines,<sup>9</sup> restricted access to extended-release/long-acting (ER/LA) opioids,<sup>10-12</sup> added abuse-deterrent properties,<sup>13-15</sup> FDA's Risk Evaluation and Mitigation Strategies (REMS),<sup>16-19</sup> opioid management plans,<sup>10,20,21</sup> and state prescription drug monitoring programs.<sup>22-24</sup> Accurate identification of OODs is essential to quantify the burden of the problem, evaluate risk-reduction strategies, monitor population-level outcomes, and improve prevention and quality of care. Furthermore, differentiation of type, such as suicides or heroin-related overdoses, is needed to target and evaluate specific interventions.

To date, few studies have validated methods used to identify OODs.<sup>25-28</sup> The purpose of this study was to improve upon and conduct a full validation of a previously developed algorithm to identify overdose events<sup>26</sup> and to develop and validate algorithms that classify types of overdoses (eg, heroin-related and suicides). The original study was limited in scope, excluding events that occurred within 3 days of surgery, and only assessed positive predictive value (PPV). The present study attempted to improve algorithm performance using (a) additional coded data, including a sample of individuals without identified OODs; (b) data extracted with natural language processing (NLP) of EHR clinical notes; (c) tests in three additional health care systems; and (d) additional performance measures.

## 2 | METHODS

### 2.1 | Study populations, data sources, and sampling procedures

Kaiser Permanente Northwest (KPNW) was the primary site for algorithm development and validation; additional sites allowed performance assessment in other data systems. KPNW is an integrated health plan providing comprehensive inpatient and outpatient medical care, including addiction and mental health treatment to members in Oregon and southwest Washington State. KPNW served about 500 000 members at study end and is demographically representative of its service area. The study population included all members with any eligibility from January 1, 2008 to December 31, 2014. Data sources included

administrative, clinical, inpatient and outpatient records, claims, and clinical care information received from non-KPNW settings.

### 2.2 | Samples

Mutually exclusive samples were constructed for development and validation. Suspected OOD cases composed approximately 56% of each sample, defined by a patient and a point in time (based on ICD9-CM/ICD10-CM codes), using the previously developed OOD algorithm (the "OOD algorithm"; Table 1) and opioid-related adverse effects codes. The remainder of each sample (approximately 44%) consisted of "at-risk" individuals likely to have an OOD but for whom no suspected OODs were identified. Individuals at risk were identified using diagnoses commonly comorbid with opioid/other substance use disorders. Diagnoses were chosen on the basis of prior research and existing literature and included the following: (a) substance abuse diagnoses, (b) mental health diagnoses, and (c) diagnoses associated with substance abuse (see Data S1). Individuals with at least two diagnoses from two categories, without suspected OODs, were considered at risk. We drew stratified random samples of at-risk individuals, half with 0-29 days' supply of ER/LA opioids in the prior year, and half with greater than or equal to 30 days' supply. At-risk periods were converted to events/nonevents for analyses.

#### 2.2.1 | Development sample

We created the development sample first, using OOD cases previously identified, to maximize information available for development. Cases included events identified using opioid-related *International Classification of Diseases, Ninth/Tenth Revisions (ICD-9/10)* poisoning codes and opioid-related adverse effects codes. We included the latter to provide information that would help differentiate OODs from adverse effects. Together, these codes together formed the "suspected overdose" stratum in the sample.

#### 2.2.2 | Validation sample

We randomly selected algorithm-identified OOD events and at-risk individuals not used in development, proportionate to the number of suspected overdoses (approximately 56%) and at-risk cases (approximately 44%) in that sample. This "balancing" allowed performance comparisons between development and validation samples. Once

development was complete, performance was assessed in the validation sample.

### 2.2.3 | Datasets

The development sample included 1006 events (suspected OODs and at-risk individuals) from 977 unique people, of which 872 (845 unique people) had EHR information available for chart audit. These 872 events comprised the development dataset. The validation sample included 1696 suspected OOD events and 1136 of which had EHR data available for audit (1100 unique people). These 1136 events formed the validation dataset. Table 2 shows counts of events by type in each sample, the base populations, and sample sizes.

## 2.3 | Portability sites

- Kaiser Permanente Washington (KPWA) provides coverage and care to approximately 760 000 individuals in Puget Sound and Spokane. KPWA's researchers have access to EHR and claims data from integrated health plan members. Limiting factors for some analyses were lack of chart auditor access to chemical dependency records and hospital data that were restricted to insurance claims.
- Optum. Data were derived from Optum's Integrated Database, which included 12.3 million individuals and links the following data: ambulatory, inpatient, medical claims, prescription, and practice management. Chronologic profiles were created from claims data, EHR structured data, and EHR notes extracted using NLP. Because profile data had limitations, we relaxed criteria for ambiguous cases, allowing "possible" responses for outcomes that were coded yes/no at other sites.
- Tennessee Medicaid (TennCare) is a managed Medicaid program, covering about 1.2 million Medicaid-eligible, state residents. TennCare maintains an enrollee registry, records of patient-provider encounters, pharmacy benefits and usage. Chart audits were restricted to inpatient hospital encounter records to facilitate record access; lack of access to outpatient records, particularly mental health follow-up after suicides/suicide attempts and outpatient chemical dependency treatment, made these classifications difficult for auditors.

## 2.4 | Chart audit process

Chart audit data were considered the "gold standard" against which algorithm results were compared. Index dates were dates of suspected OODs or, for those at risk, dates of the second qualifying diagnosis. We examined clinical records of at-risk patients for overdoses for a 2-year period centered upon the index date (greater than one OOD allowed per person). If no overdoses were found, event dates for at-risk individuals became dates of qualifying diagnoses; individuals were coded as having no overdose on that date.

At KPNW, professional chart auditors determined whether suspected OODs were actual OODs, and classified confirmed OODs

### KEY POINTS

- OODs can be identified using coded insurance claims data or electronic health records.
- Heroin-involved OODs can be accurately identified.
- OODs that are suicides/suicide attempts can be identified with adequate accuracy using coded data.
- OODs involving substance abuse can be identified with adequate accuracy using coded data.
- Algorithms for classifying suicides/suicide attempts and substance abuse-involved overdoses can be significantly improved using data derived from natural language processing of clinical text in electronic health records.

according to whether or not they involved: (a) substance abuse; (b) misuse of prescribed medications (ie, therapeutic use not consistent with directions); (c) heroin; (d) inpatient pain management or anesthesia; (e) patient errors; (f) clinician prescribing errors; (g) alcohol; (h) other substances; and (i) suicide/attempted suicide. We completed 100% duplicate review and resolved discrepancies with review by clinicians and team discussion. Portability sites completed audits using similar procedures, though some adaptations were needed to account for site and data differences. Detailed chart audit procedures, including definitions, are available in Data S2.

## 2.5 | Algorithm performance standards

Algorithm performance was evaluated using sensitivity, specificity, PPV, NPV, and F-scores. No universal standards exist for sensitivity or specificity because acceptability depends on context and use. We expected OODs and heroin involvement to be readily identifiable, thus set sensitivity and specificity of 85% as acceptable and 90% as excellent for these algorithms. We expected other classifications (abuse, suicides/attempts, misuse, polysubstance involvement, and medication error) to be more difficult, thus set 75% as acceptable for sensitivity/specificity. F-scores measure the accuracy of the algorithm and vary from 0 to 1, with 1 representing perfect fit. We considered F-score values of 0.90 or greater to indicate excellent performance. Chi-square tests assessed differences between performance in development and validation datasets.

## 2.6 | Code-based algorithm development procedures

We used the least absolute shrinkage and selection operator (LASSO) regression<sup>29,30</sup> to select variables based on predicted probabilities of case status, and classification and regression trees (CART) with random forest (500 trees)<sup>31</sup> to evaluate cutoff values for continuous

**TABLE 1** ICD-9 and ICD-10 codes used in the initial algorithm to detect opioid overdoses<sup>a</sup>

	ICD-9 Code	ICD-10 Code
Poisoning by opium (alkaloids) unspecified	965.00	
Poisoning by heroin	965.01	
Poisoning by methadone	965.02	
Poisoning by other opiates and related narcotics	965.09	
Accidental poisoning by heroin	E850.0	
Accidental poisoning by methadone	E850.1	
Accidental poisoning by other opiates and related narcotics	E850.2	
COD: poisoning by opiates and related narcotics <sup>b</sup>	9650	
COD: poisoning by opium		T40.0
COD: poisoning by heroin		T40.1
COD: poisoning by other opioids		T40.2
COD: poisoning by methadone		T40.3
COD: poisoning by other synthetic narcotic		T40.4
COD: accidental poisoning by and exposure to narcotics and psychodysleptics, not elsewhere classified		X42
COD: intentional self-poisoning by and exposure to narcotics and psychodysleptics, not elsewhere classified		X62
COD: undetermined poisoning by and exposure to narcotics and psychodysleptics, not elsewhere classified		Y12

Note. COD: cause of death; ICD-9/10: International Classification of Diseases, Ninth/Tenth Revision.

<sup>a</sup>Presence of any of these codes was considered indicative of the presence of an OOD.

<sup>b</sup>Though the study period was prior to the nationwide switch to ICD-10 diagnostic codes, ICD-10 cause-of-death codes were in use nationwide throughout the study period.

variables. Variables selected using LASSO were entered into logistic regression analyses to estimate final parameters. Predicted probabilities from logistic regression analyses were used to classify events. Parameters were then applied to the validation dataset and compared with chart audit findings.

**TABLE 2** Development and validation sample descriptions

Sampling Goal	Events Identified by:	Base Population	Development Sample	Validation Sample
Suspected overdose cases	Opioid poisoning diagnostic codes	2,271	483	848
	Opioid adverse effects diagnostic codes	254	78	103
At risk for overdose	Pain, mental health, and substance abuse diagnostic codes	87,550	222 with ≥30 days supply of ER/LA opioids 223 with ≤30 days supply of ER/LA opioids	373 with ≥30 days supply of ER/LA opioids 372 with ≤30 days supply of ER/LA opioids
Total			1006	1696

Note. ER/LA: extended-release/long-acting.

## 2.6.1 | OOD algorithm

We began by forcing the ICD-9/10 codes from our prior algorithm<sup>26</sup> into the model (Table 1), then tested the following additional variables: prescribed opioids in the 9 months prior to index date (immediate release [IR], ER/LA, or both); substance abuse diagnoses; opioid withdrawal diagnoses; weight change; and hospitalizations. None improved performance. The best fitting model used a binary variable coded “1” if any ICD-9/10 codes listed in Table 1 were found.

## 2.6.2 | OODs with heroin involvement

Among chart-audit-confirmed OODs in the development sample (n = 423), codes shown in Table 3 were tested to predict heroin involvement. The best fitting model performed well, and used a binary variable coded “1” if E965.01, E850.0, or T40.1 were present.

## 2.6.3 | Suicides/suicide attempt-related OODs

We modeled overdoses in the development sample (n = 423) according to whether or not they were suicides/attempted suicides (see Tables 4 and 5 for codes evaluated). The initial model assessed predictors identified in the Mental Health Research Network's work predicting suicides/suicide attempts.<sup>32</sup> We created a binary variable if any suicide-related codes in Table 4 were present, binary variables for single or multiple episode depression diagnoses, and interaction terms for suicide and recurrent depression. Next, we tested diagnoses

**TABLE 3** ICD-9 and ICD-10 codes tested in the model to predict if an opioid overdose was related to heroin

E950.0	Suicide and self-inflicted poisoning by solid or liquid substances
E965.01	Poisoning by heroin
E850.0	Accidental poisoning by heroin
T40.1	COD: poisoning by heroin

Note. COD: cause of death; ICD-9/10: International Classification of Diseases, Ninth/Tenth Revision.

**TABLE 4** ICD-9 diagnostic codes evaluated to predict whether or not an opioid overdose was intentional

Suicide/Suicide Attempt	
E950.0	Suicide and self-inflicted poisoning by solid or liquid substances
E950.1	Barbiturates
E950.2	Sedatives and hypnotics
E950.3	Tranquilizers and other psychotropic agents
E950.4	Other specified drugs and medicinal substances
E950.5	Unspecified drug or medicinal substance
E950.9	Other and unspecified solid and liquid substances
E956	Suicide and self-inflicted injury by cutting and piercing instrument
E958.8	Suicide and self-inflicted injury by other specified means
E958.9	Suicide and self-inflicted injury by unspecified means
V62.84	Suicidal ideation
Single Episode Depression	
296.20	Major depressive affective disorder, single episode, unspecified
296.22	Major depressive affective disorder, single episode, moderate
296.23	Major depressive affective disorder, single episode, severe, without mention of psychotic behavior
Multiple Episode Depression	
296.30	Major depressive affective disorder, recurrent episode, unspecified
296.31	Major depressive affective disorder, recurrent episode, mild
296.32	Major depressive affective disorder, recurrent episode, moderate
296.33	Major depressive affective disorder, recurrent episode, severe, without mention of psychotic behavior
296.36	Major depressive affective disorder, recurrent episode, in full remission

Note. ICD-9: International Classification of Diseases, Ninth Revision.

**TABLE 5** Diagnoses tested for predicting suicides/suicide attempts

Diagnoses	ICD-9 Codes
Alcohol use disorders	291.x, 303.x, 305.0
Anxiety disorders	300.0, 300.2, 300.3, 309.20, 309.21, 309.24, 309.81
Bipolar disorders	296.0, 296.1, 296.4–296.7, 296.80, 296.81, 296.89
Drug use disorders	292.x, 304.x, 305.2–305.9
Other psychoses	297.1, 297.3, 298.8, 298.9, 301.22
Schizophrenia spectrum disorders	295.x
Tobacco use disorder	305.1, 649.0, 989.84, V15.82

Note. ICD-9: International Classification of Diseases, Ninth Revision.

for alcohol use disorders, anxiety disorders, bipolar disorders, tobacco and drug use disorders, other psychoses, and schizophrenia spectrum disorders (Table 5). The best fitting model included main effects for suicide codes, single episode depression, recurrent depression, alcohol use disorder (AUD), nonalcohol substance use disorder (SUD), and three interaction terms for suicide with multiple episode depression, AUD and SUD.

## 2.6.4 | Substance use involved unintentional OODs

For unintentional overdoses ( $n = 268$ ), we developed an algorithm to detect substance abuse involvement, using logistic regression to evaluate diagnostic codes listed in Table 6. The final model included two terms: (a) a binary indicator of substance abuse coded "1" if any ICD-9 code related to heroin was present, or if there were no dispenses for an ER/LA or IR opioid in the year prior to the overdose (including the event date), and (b) indication of opioid abuse from ICD-9 codes in the 2 years prior to the event.

**TABLE 6** Diagnoses and other variables tested for identifying substance abuse involvement in opioid-related overdoses

Diagnoses and Other Variables	ICD-9 and ICD-10 Codes
Poisoning by heroin	965.01
Accidental poisoning by heroin	E850.0
Poisoning by heroin	T40.1(COD)
Poisoning by opium (alkaloids), unspecified	965.00
Poisoning by methadone	965.02
Poisoning by other opiates/narcotics	965.09
Opioid dependence	304.00–304.03, 304.70–304.72
Opioid abuse	305.50–305.53
Drug use disorders	292.x, 304.x, 305.2–305.9
Alcohol use disorders	291.x, 303.x, 305.0
Tobacco use disorders	305.1, 649.0, 989.84, V15.82
Count of unique opioid prescribers in the 2 years prior to event	n/a
Count of early ER/LA opioid dispenses in the 2 years prior to event (early refill defined as two consecutive fills of an ER/LA where the number of days between prescriptions was less than or equal to 85% of the days' supply in the first prescription)	n/a
Count of early IR opioid dispenses in 2 years prior to the event	n/a
Medicaid/Medicare dual eligibility	n/a

Note. COD: cause of death; ER/LA: extended-release/long-acting; ICD-9/10: International Classification of Diseases, Ninth/Tenth Revision.

## 2.7 | NLP-enhanced algorithm development

We attempted to enhance code-based algorithm performance with indicators extracted using NLP of EHR clinical notes. The companion paper (Hazlehurst et al<sup>33</sup>) provides details about development and validation of NLP-derived variables. Each NLP-derived variable was tested using logistic regression analyses to determine whether or not its addition improved performance beyond that of each respective code-based algorithm. We used DeLong's test for two correlated receiver operator curves to compare areas under the curve for code-based and NLP-enhanced models.

Table 7 shows definitions of NLP-derived binary variables to code-based algorithms for testing. The NLP classification "polysubstance including opioid" was broken into six binary subcomponents: named opioids; general "narcotics"; named opioid-interacting medications; named recreational/illicit drugs; alcohol; named over-the-counter

medications. The NLP classification "substance abuse" was broken into two subcomponents: alcohol or substance abuse noted; alcohol presence mentioned.

## 2.8 | Portability assessment

Portability sites contributed data from commercial insurance, Medicaid and integrated health care settings. The OOD algorithm was applied to each site's population from 2008 to 2014 (2008-2013 for TennCare). Classification algorithms were tested at sites with adequate data. At each site, a random sample of approximately 250 suspected OODs was selected for inclusion. The remaining approximately 250 individuals per sample were selected using the same criteria used at KPNW to identify at-risk individuals. At KPWA, 435/500 cases had EHR information available (159 confirmed OODs);

**TABLE 7** Natural language processing (NLP) classifications and their alignment with gold standard chart review

	NLP-only Classification	Chart Review Gold Standard Comparator
Event type, irrespective of substance involved	Intentional overdose (suicide/suicide attempt) Unintentional overdose (excludes intentional overdose) Overdose of any type (combines intentional and unintentional overdose) Adverse drug event (excludes any overdose)	Intentional overdose = clearly or possible Unintentional overdose Unintentional overdose or intentional overdose = clearly or possible Adverse drug reaction
Substance involved in overdose or adverse drug event	Heroin NLP identifies an overdose or adverse drug event in combination with heroin, regardless of other opioid or nonopioid prescription or over-the-counter medication Opioid only (excludes heroin) NLP identifies an overdose or adverse drug event in combination with a named opioid (or generic "narcotic") in the absence of heroin and additional nonopioid prescription or over-the-counter medications. Polysubstance including opioid (excludes heroin) NLP identifies an overdose or adverse drug event in combination with a named opioid (or generic "narcotic") AND additional nonopioid prescription or over-the-counter medications, but in the absence of heroin. Any opioid (excludes heroin, include polysubstance)	Heroin involved = yes or possible  A single opioid event (excludes heroin)  A polydrug, opioid event (excludes heroin)  A single or polydrug opioid event (excludes heroin)
Substance abuse involved in opioid-related overdose	Prescription medication abuse (whether prescribed or not) NLP identifies an opioid overdose in combination with abuse of medications (ie, nontherapeutic goals/actions noted about prescription medications), or documented conclusion of abuse by clinician. Substance abuse (including alcohol abuse or presence) NLP identifies an opioid overdose in combination with notations of alcohol abuse or just alcohol present Illicit drug abuse NLP identifies an opioid overdose in combination with heroin or other named recreational drugs (marijuana, cocaine, and methamphetamine). Any substance abuse Any of the above types of abuse.	Opioid or nonopioid prescription med abuse = yes, AND NOT heroin = yes or possible in an opioid overdose event  Alcohol present = yes in an opioid overdose event  Abuse of nonprescribed substances = yes in an opioid overdose event  Opioid or nonopioid prescription med or nonprescribed substance abuse or alcohol present = yes in an opioid overdose event
Patient error in opioid-related overdose or adverse drug event	Patient error (EXCLUDES ALL ABUSE above) NLP identifies an opioid overdose or adverse drug event in combination with mention of mistake/accident in taking medications	Opioid or nonopioid medication-taking error = yes AND NOT ABUSE AS DEFINED ABOVE

500 cases were audited at Optum (258 confirmed OODs), and 516 at TennCare (240 OODs).

Classification algorithms demonstrating acceptable performance at KPNW were implemented at portability sites, comparing results with chart audit. Lack of adequate numbers of cases prevented testing some classification algorithms: heroin involvement at TennCare; substance abuse involvement at KPWA (code-based,

NLP-enhanced). The NLP-enhanced algorithm for suicides/suicide attempts was only tested at KPWA as other sites did not have NLP capability.

Mean age and sex in development and validation datasets were similar (48.8 and 46.9 years; 39.5% and 39.6% male, respectively). Confirmed OOD prevalence was 48.5% in the development dataset (n = 423) and 53.3% (n = 605) in the validation dataset. The validation

**TABLE 8** Cross validation of models with 95% confidence intervals and significance tests comparing the performance in the two samples

	Development Dataset	Validation Dataset	P Value for the Difference
<b>Opioid-related overdose</b>			
Sensitivity (95% CI)	97.9 (96.0-99.0)	97.2 (95.5-98.4)	0.493
Specificity (95% CI)	88.9 (85.6-91.6)	84.6 (81.3-87.5)	0.028
PPV (95% CI)	89.2 (86.4-91.5)	87.4 (85.0-89.3)	0.3421
NPV (95% CI)	97.8 (95.9-98.8)	96.5 (94.5-97.8)	0.205
F-score	0.93	0.92	
<b>OOD<sup>a</sup> classified as heroin involved</b>			
Sensitivity (95% CI)	94.0 (83.5-98.8)	91.8 (84.6-96.4)	0.635
Specificity (95% CI)	96.8 (94.5-98.3)	99.0 (97.7-99.7)	0.018
PPV (95% CI)	79.7 (69.1-87.3)	94.7 (88.3-97.7)	0.004
NPV (95% CI)	99.2 (97.6-99.7)	98.4 (97.0-99.2)	0.330
F-score	0.86	0.93	
<b>OOD classified as suicide/suicide attempt</b>			
Sensitivity (95% CI)	77.4 (70.0-83.7)	70.5 (63.8-76.7)	0.142
Specificity (95% CI)	88.1 (83.6-91.7)	90.2 (86.9-92.9)	0.380
PPV (95% CI)	79.0 (82.8-84.0)	78.9 (73.3-83.6)	0.995
NPV (95% CI)	87.1 (83.4-90.1)	85.5 (82.6-87.9)	0.551
F-score	0.78	0.74	
<b>OOD classified as abuse involved</b>			
Sensitivity (95% CI)	82.0 (73.6-88.6)	75.3 (68.2-81.5)	0.184
Specificity (95% CI)	83.4 (76.7-88.9)	79.5 (73.6-84.6)	0.329
PPV (95% CI)	77.8 (70.9-83.4)	74.0 (68.5-78.9)	0.462
NPV (95% CI)	86.8 (81.4-90.7)	80.5 (76.0-84.4)	0.117
F-score	0.80	0.75	
<b>Overdose classified as suicide/suicide attempt using NLP-enhanced algorithm</b>			
Sensitivity (95% CI)	88.4 (82.3-93.0)	78.7 (72.5-84.1)	0.016
Specificity (95% CI)	91.8 (87.8-94.8)	91.0 (87.7-93.6)	0.707
PPV (95% CI)	86.2 (80.6-90.3)	81.9 (76.7-86.2)	0.278
NPV (95% CI)	93.2 (89.8-95.5)	89.2 (86.3-91.5)	0.079
F-score	0.87	0.80	
<b>Overdose classified as involving abuse using NLP-enhanced algorithm</b>			
Sensitivity (95% CI)	90.1 (83.0-95.0)	80.5 (73.8-86.1)	0.030
Specificity (95% CI)	79.6 (72.5-85.6)	76.3 (70.2-81.8)	0.449
PPV (95% CI)	75.8 (69.5-81.1)	72.5 (67.4-77.2)	0.517
NPV (95% CI)	91.9 (86.6-95.2)	83.4 (78.7-87.3)	0.023
F-score	0.82	0.76	

Note. NLP: natural language processing; OOD: opioid-related overdose.

<sup>a</sup>Opioid-related overdose.

dataset had a slightly higher prevalence of heroin-involved OODs than the development dataset (16.2% vs. 11.8%) and similar prevalence of intentional overdoses (36.6% development and 34.2% validation), and substance use involved unintentional overdoses (41.4% and 43.7%, respectively). The study protocol was reviewed and approved by all sites' institutional review boards.

### 3 | RESULTS

Table 8 summarizes algorithm performance in development and validation datasets for algorithms reaching acceptable performance during development. The OOD algorithm performed well: sensitivity (97.2%), specificity (84.6%), PPV (87.4%), and NPV (96.5%). There were no differences between development and validation datasets on sensitivity, PPV, or NPV, though specificity declined. The heroin-involved classification also performed well in validation: sensitivity was 91.8%, specificity 99.0%, PPV 94.7%, and NPV 98.4%. Some measures of performance (specificity and PPV) were significantly ( $P \leq 0.018$ ) better in the validation compared with the development dataset (Table 8).

For the classification identifying suicides/suicide attempts, performance was acceptable for the code-based model in both datasets and did not differ significantly ( $P \geq 0.142$ ) between the two. Sensitivity was 70.5%, specificity 90.2%, PPV 78.9%, and NPV 85.5% for the coded-based algorithm.

Performance of the code-based classification for substance abuse-involved OODs was also acceptable: Sensitivity was 75.3%, specificity 79.5%, PPV 74.0%, and NPV 80.5%. Performance did not differ between development and validation datasets ( $P \geq 0.117$ ).

Models classifying opioid misuse, patient medication errors, and polysubstance involvement did not reach acceptable (greater than or equal to 0.75%) levels of sensitivity and specificity so were not validated. We identified few events involving clinician error and few inpatient events making modeling unfeasible for these classifications. For inpatient overdose/oversedation, we adopted a different identification strategy (see Green et al, companion paper<sup>34</sup>).

#### 3.1 | NLP-enhanced models

The addition of NLP variables did not improve performance of the code-based OOD algorithm or the heroin algorithm, though NLP-enhanced models outperformed code-based models for classifying suicides/suicide attempts (sensitivity = 78.7%, specificity = 91.0%, PPV = 81.9%, and NPV = 89.2%) and those involving substance abuse (sensitivity = 80.5%, specificity = 76.3%, PPV = 72.5%, and NPV = 83.4%). Performance declined for the NLP-enhanced models in the validation dataset compared with the development dataset (Table 8), but validation results remained above acceptable limits for suicides/suicide attempts and substance abuse-involved overdoses for sensitivity, specificity, and NPV.

#### 3.2 | Algorithm portability performance

Table 9 presents performance across portability sites. The OOD algorithm performed well, with sensitivity greater than 96%, specificity greater than 84%, and F-scores greater than 0.92. Cross-site sensitivity for sites with adequate data to test the heroin-involved classification was greater than 87%, specificity greater than 99%, and F-scores greater than or equal to 0.90. The code-based algorithm for suicides/suicide attempts performed equally at KPNW and KPW (F-scores of 0.74), though there was a decline in sensitivity at Optum and TennCare, likely a result of data limitations. The NLP-enhanced algorithm for suicides/attempts performed better than the code-based algorithm and also performed better at KPWA than at KPNW (F-scores of 0.85 and 0.80, respectively). The substance abuse-involved algorithm performed poorly at Optum and TennCare. Upon

**TABLE 9** Final results of portability assessment of code-based and NLP-enhanced algorithms

	KPNW	KPW	Optum	TennCare
<b>OOD</b>				
Sensitivity	97.2	100.0	96.9	99.2
Specificity	84.6	89.2	100.0	92.4
PPV	87.4	84.1	100.0	91.9
NPV	96.5	100.0	96.9	99.2
F-score	0.92	0.92	0.98	0.95
<b>Heroin</b>				
Sensitivity	91.8	87.5	98.5	N/A
Specificity	99.0	99.3	100.0	N/A
PPV	94.7	93.3	100.0	N/A
NPV	98.4	98.5	99.5	N/A
F-score	0.93	0.90	0.99	N/A
<b>Suicide/suicide attempt</b>				
Sensitivity	70.5	74.1	63.2	44.9
Specificity	90.2	86.7	91.0	87.1
PPV	78.9	74.1	81.1	64.5
NPV	85.5	86.7	80.1	75.1
F-score	0.74	0.74	0.67	0.53
<b>Suicide/suicide attempt—NLP enhanced</b>				
Sensitivity	78.7	81.5	N/A	N/A
Specificity	91.0	95.2	N/A	N/A
PPV	81.9	89.8	N/A	N/A
NPV	89.2	90.9	N/A	N/A
F-score	0.80	0.85	N/A	N/A
<b>Abuse involved</b>				
Sensitivity	75.3	N/A	67.1	41.5
Specificity	79.5	N/A	31.9	62.5
PPV	74.0	N/A	46.2	29.0
NPV	80.5	N/A	52.6	74.4
F-score	0.76	N/A	0.55	0.34



further investigation, it appeared that chart audit determinations for substance abuse-related OODs at these two sites were made difficult by data limitations.

## 4 | DISCUSSION

The code-based OOD algorithm, using *ICD-9* diagnostic codes and death data using *ICD-10* codes, has excellent performance across health systems, whether applied to EHR-based databases or commercial or Medicaid claims databases. Few inpatient events were identified using the algorithm, suggesting that it does not confound inpatient overdoses that are typically medically related with accidental or intentional overdoses occurring elsewhere. Given the excellent performance of the code-based OOD algorithm, the NLP-enhanced algorithm was unable to improve performance.

In short, our results show that a simple code-based algorithm can be used to accurately identify overdoses in widely differing settings. Similarly, the code-based algorithm classifying overdoses as heroin-involved showed excellent performance across settings when adequate numbers of heroin-related events were available for testing. As with the OOD algorithm, however, there was little room from improved performance in an NLP-enhanced model.

Identifying suicides/attempted suicides presented a greater challenge, though the code-based algorithm performed adequately. Performance was not as good at Optum or TennCare, but PPV and NPV were acceptable in all but TennCare, where reviews were focused on inpatient health care encounters rather than both inpatient and outpatient data. Other sites had access to OOD-related follow-up visits that often provided information necessary to determine when an OOD was a suicide/suicide attempt. The NLP-enhanced algorithm significantly improved performance for detecting suicides/suicide attempts suggesting that including NLP-derived data from clinical notes is beneficial when available. Performance of the NLP-enhanced model was good in both sites with the necessary data.

The code-based algorithm for substance abuse involvement showed reasonable, moderate performance in KPNW, particularly given the complexity of identifying substance abuse with coded data alone. With room for improvement and the likelihood that clinicians document suspicions about substance abuse in clinical notes rather than using diagnostic codes, we expected and found that the NLP-enhanced algorithm performance was better. Unfortunately, chart auditors did not have access to chemical dependency records at KPWA, preventing detection of adequate numbers of abuse-related overdoses to test the model there. Nevertheless, results in KPNW suggest that using NLP to identify substance abuse-involved overdoses will be more fruitful than code-based algorithms alone.

As a result of limited cases for some classifications, we were unsuccessful with code-based or NLP-enhanced models classifying prescription medication misuse or patient errors. Other limitations included that development was based primarily on data extracted from a single integrated health care system. Portability assessments designed to overcome this limitation provide important confirmation

for some outcomes but limitations in sample sizes and data sources at portability sites made some comparisons unfeasible. Also, NLP-enhanced algorithms rely on a specific NLP system (MediClass) for extracting information from clinical notes. Although the “knowledge” used by MediClass is easily extracted for use by other NLP systems, results may not be identical.

## 5 | CONCLUSIONS

To our knowledge, this work is the most comprehensive validation of algorithms developed to identify and classify OODs. The code-based OOD algorithm shows excellent performance across different health care systems using *ICD-9* encounter codes. The same holds true for classifying overdoses involving heroin. Algorithms for identifying opioid-related suicides/suicide attempts and substance abuse-involved overdoses perform adequately, particularly given the complexity of identifying these types of OODs. The NLP-enhanced algorithms for suicides/suicide attempt-related overdoses and abuse-related overdoses substantially enhance classification, which should improve ascertainment in settings with NLP capacity. Finally, we used a conservative strategy for evaluating algorithm performance by using an at-risk sample of noncases. A random sample of the population would likely have resulted in no identified opioid overdoses and in better algorithm performance but would not have allowed us to learn from missed cases as part of algorithm development. Additional research is now needed to translate and assess the algorithms for use with *ICD-10* encounter data.

## ETHICS STATEMENT

The study was reviewed and approved by the Kaiser Permanente Northwest Institutional Review Board for the Protection of Human Subjects. Participating sites reviewed the study and ceded oversight to the KPNW IRB.

## ACKNOWLEDGEMENTS

This project was conducted as part of a Food and Drug Administration (FDA)-required postmarketing study for extended-release and long-acting opioid analgesics and was funded by the Opioid Postmarketing Consortium consisting of the following companies at the time of study conduct: Allergan; Assertio Therapeutics, Inc.; BioDelivery Sciences, Inc.; Collegium Pharmaceutical, Inc.; Daiichi Sankyo, Inc.; Egalet Corporation; Endo Pharmaceuticals, Inc.; Hikma Pharmaceuticals USA Inc.; Janssen Pharmaceuticals, Inc.; Mallinckrodt Inc.; Pernix Therapeutics Holdings, Inc.; Pfizer, Inc.; and Purdue Pharma, LP. The investigators would like to thank Elizabeth Shuster, MS and Ana G. Rosales, MS for help with data extraction and analysis, respectively.

## CONFLICT OF INTEREST

This project was conducted as part of a Food and Drug Administration (FDA)-required postmarketing study of extended-release and long-

acting opioid analgesics (<https://www.fda.gov/downloads/Drugs/DrugSafety/InformationbyDrugClass/UCM484415.pdf>) funded by the Opioid Postmarketing Consortium (OPC). The OPC is composed of companies that hold NDAs of extended-release and long-acting opioid analgesics and, at the time of publication, they included the following companies: Allergan; Assertio Therapeutics, Inc; BioDelivery Sciences, Inc; Collegium Pharmaceutical, Inc; Daiichi Sankyo, Inc; Egalet Corporation; Endo Pharmaceuticals, Inc; Hikma Pharmaceuticals USA Inc; Janssen Pharmaceuticals, Inc; SpecGX, LLC; Pernix Therapeutics Holdings, Inc; Pfizer, Inc; and Purdue Pharma, LP. The study was designed in collaboration between OPC members and independent investigators with input from FDA. Investigators maintained intellectual freedom in terms of publishing final results. This study was registered with ClinicalTrials.gov as study NCT02667197 on January 28, 2016. All authors received research funding from the OPC. Drs Green and Perrin and Ms Janoff received prior funding from Purdue Pharma, LP to carry out related research. Dr Green provided research consulting to the OPC. Kaiser Permanente Center for Health Research (KPCHR) staff had primary responsibility for study design, though OPC members provided comments on the protocol. The protocol and statistical analysis plan were reviewed by FDA, revised following review, and then approved. All algorithm development and validation analyses were conducted by KPCHR; analyses of algorithm portability were completed by each participating site. KPCHR staff made all final decisions regarding publication and content, though OPC members reviewed and provided comments on the manuscript. Drs DeVeaugh-Geiss and Coplan were employees of Purdue Pharma, LP at the time of the study. Dr Carrell has received funding from Pfizer Inc and Purdue Pharma, LP to carry out related research. Dr Grijalva has served as a consultant for Pfizer, and Merck for unrelated work; he has also received funding from NIH, AHRQ, CDC, FDA, and Sanofi-Pasteur. Drs Liang and Enger are employees of Optum.

## ORCID

Carla A. Green  <https://orcid.org/0000-0002-0000-4381>

Brian Hazlehurst  <https://orcid.org/0000-0001-9365-3964>

David S. Carrell  <https://orcid.org/0000-0002-8471-0928>

Caihua Liang  <https://orcid.org/0000-0001-6934-3587>

## REFERENCES

- Substance Abuse and Mental Health Services Administration. Results from the 2012 National Survey on Drug Use and Health: mental health findings. *NSDUH Series H-47, HHS Publication No. (SMA) 13-4805*. Rockville, MD: Office of Applied Studies;2013.
- Centers for Disease Control and Prevention. Vital signs: overdoses of prescription opioid pain relievers—United States, 1999–2008. *MMWR*. 2011;60(43):1487-1492.
- Warner M, Chen LH, Makuc DM, Anderson RN, Minino AM. Drug poisoning deaths in the United States, 1980–2008. *NCHS Data Brief*. 2011;(81):1-8.
- Mack KA. Drug-induced deaths—United States, 1999–2010. *MMWR Surveill Summ*. 2013;62(Suppl 3):161-163.
- Rudd RA, Paulozzi LJ, Bauer MJ, et al. Increases in heroin overdose deaths—28 states, 2010 to 2012. *MMWR*. 2014;63(39):849-854.
- Rudd RA, Aleshire N, Zibbell JE, Gladden M. Increases in drug and opioid overdose deaths—United States, 2000–2014. *Morb Mortal Wkly Rep*. 2016;64(Early Release):1-5.
- Rudd RA, Seth P, David F, Scholl L. Increases in drug and opioid-involved overdose deaths—United States, 2010–2015. *MMWR Morb Mortal Wkly Rep*. 2016;65(5051):1445-1452.
- US Department of Health and Human Services. Statement of the U.S. Surgeon General VADM Jerome M. Adams, MD, MPH, regarding the President's action on opioids; 10/26/2017. <https://www.surgeongeneral.gov/news/2017/10/pr20171026.html>. Accessed 11/27/2017.
- Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain—United States, 2016. *MMWR Recomm Rep*. 2016;65(1):1-49.
- Chou R, Fanciullo GJ, Fine PG, et al. Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. *J Pain*. 2009;10(2):113-130.
- Argoff CE, Kahan M, Sellers EM. Preventing and managing aberrant drug-related behavior in primary care: systematic review of outcomes evidence. *J Opioid Manag*. 2014;10(2):119-134.
- Garcia MM, Angelini MC, Thomas T, Lenz K, Jeffrey P. Implementation of an opioid management initiative by a state Medicaid program. *J Manag Care Pharm*. 2014;20(5):447-454.
- Coplan PM, Kale H, Sandstrom L, Landau C, Chilcoat HD. Changes in oxycodone and heroin exposures in the National Poison Data System after introduction of extended-release oxycodone with abuse-deterrent characteristics. *Pharmacoepidemiol Drug Saf*. 2013;22(12):1274-1282.
- Severtson SG, Bartelson BB, Davis JM, et al. Reduced abuse, therapeutic errors, and diversion following reformulation of extended-release oxycodone in 2010. *J Pain*. 2013;14(10):1122-1130.
- Butler SF, Cassidy TA, Chilcoat H, et al. Abuse rates and routes of administration of reformulated extended-release oxycodone: initial findings from a sentinel surveillance sample of individuals assessed for substance abuse treatment. *J Pain*. 2013;14(4):351-358.
- Nelson LS, Perrone J. Curbing the opioid epidemic in the United States: the risk evaluation and mitigation strategy (REMS). *JAMA*. 2012;308(5):457-458.
- Porada S. REMS: red tape, or a remedy for opioid abuse? *J Fam Pract*. 2011;60(9 Suppl):S55-S62.
- Sloan PA. Opioid risk management: understanding FDA mandated risk evaluation and mitigation strategies (REMS). *J Opioid Manag*. 2009;5(3):131-133.
- Stanos S. Evolution of opioid risk management and review of the classwide REMS for extended-release/long-acting opioids. *Physician Sports Med*. 2012;40(4):12-20.
- Arnold RM, Han PK, Seltzer D. Opioid contracts in chronic nonmalignant pain management: objectives and uncertainties. *Am J Med*. 2006;119(4):292-296.
- Federation of State Medical Boards of the United States I. Model guidelines for the use of controlled substances for the treatment of pain. Euless, TX2000.
- Han H, Kass PH, Wilsey BL, Li CS. Increasing trends in schedule II opioid use and doctor shopping during 1999–2007 in California. *Pharmacoepidemiol Drug Saf*. 2014;23(1):26-35.
- Weiner SG, Griggs CA, Mitchell PM, et al. Clinician impression versus prescription drug monitoring program criteria in the assessment of

- drug-seeking behavior in the emergency department. *Ann Emerg Med.* 2013;62(4):281-289.
24. Chakravarthy B, Shah S, Lotfipour S. Prescription drug monitoring programs and other interventions to combat prescription opioid abuse. *West J Emerg Med.* 2012;13(5):422-425.
25. Dunn KM, Saunders KW, Rutter CM, et al. Opioid prescriptions for chronic pain and overdose: a cohort study. *Ann Intern Med.* 2010;152(2):85-92.
26. Green CA, Perrin NA, Janoff SL, Campbell CI, Chilcoat HD, Coplan PM. Assessing the accuracy of opioid overdose and poisoning codes in diagnostic information from electronic health records, claims data, and death records. *Pharmacoepidemiol Drug Saf.* 2017;26(5):509-517.
27. Rowe C, Vittinghoff E, Santos GM, Behar E, Turner C, Coffin PO. Performance measures of diagnostic codes for detecting opioid overdose in the emergency department. *Acad Emerg Med.* 2017;24(4):475-483.
28. Chung CP, Callahan ST, Cooper WO, et al. Development of an algorithm to identify serious opioid toxicity in children. *BMC Res Notes.* 2015;8(1):293.
29. Efron B, Johnstone I, Hastie T, Tibshirani R. Least angle regression. *Ann Stat.* 2002;32(2):407-499.
30. Tibshirani R. Regression shrinkage and selection via the lasso. *J Royal Stat Soc B.* 1996;58(1):267-288.
31. Rokach L, Maimon O. *Data Mining with Decision Trees: Theory and Applications.* Singapore: Word Scientific Publishing Co., Inc; 2008.
32. Simon GE, Johnson E, Lawrence JM, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *Am J Psychiatry.* 2018: [https://doi.org/10.1176/appi.ajp.2018.17101167;175\(10\):951-960](https://doi.org/10.1176/appi.ajp.2018.17101167;175(10):951-960).
33. Hazlehurst B, Green CA, Perrin NA, et al. Using natural language processing of clinical text to enhance the identification of opioid-related overdoses in electronic health records data. *Pharmacoepidemiol Drug Saf.* 2018;28(8):1143-1151.
34. Green CA, Hazlehurst B, Brandes J, et al. Development of an algorithm to identify inpatient opioid-related overdoses and oversedation using electronic data. *Pharmacoepidemiol Drug Saf.* 2018;28(8):1138-1142.

## SUPPORTING INFORMATION

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

**How to cite this article:** Green CA, Perrin NA, Hazlehurst B, et al. Identifying and classifying opioid-related overdoses: A validation study. *Pharmacoepidemiol Drug Saf.* 2019;28: 1127–1137. <https://doi.org/10.1002/pds.4772>