

ORIGINAL REPORT

Using natural language processing of clinical text to enhance identification of opioid-related overdoses in electronic health records data

Brian Hazlehurst¹  | Carla A. Green¹  | Nancy A. Perrin¹  | John Brandes¹ |
David S. Carrell²  | Andrew Baer³ | Angela DeVeaugh-Geiss⁴ | Paul M. Coplan^{4,5}

¹Center for Health Research, Kaiser Permanente Northwest, Portland, OR

²Health Research Institute, Kaiser Permanente Washington, Seattle, WA

³Group Health Research Institute, Group Health Cooperative, Seattle, WA

⁴Epidemiology, Medical Affairs, Purdue Pharma, LP, Stamford, CT

⁵Adjunct, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Correspondence

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

Present Address

Angela DeVeaugh-Geiss, Epidemiology, Indivior, North Chesterfield, VA.
Paul M. Coplan, Epidemiology, Johnson & Johnson, New Brunswick, NJ.

Funding information

Opioid Postmarketing Consortium (OPC)

Abstract

Purpose: To enhance automated methods for accurately identifying opioid-related overdoses and classifying types of overdose using electronic health record (EHR) databases.

Methods: We developed a natural language processing (NLP) software application to code clinical text documentation of overdose, including identification of intention for self-harm, substances involved, substance abuse, and error in medication usage. Using datasets balanced with cases of suspected overdose and records of individuals at elevated risk for overdose, we developed and validated the application using Kaiser Permanente Northwest data, then tested portability of the application using Kaiser Permanente Washington data. Datasets were chart-reviewed to provide a gold standard for comparison and evaluation of the automated method.

Results: The method performed well in identifying overdose (sensitivity = 0.80, specificity = 0.93), intentional overdose (sensitivity = 0.81, specificity = 0.98), and involvement of opioids (excluding heroin, sensitivity = 0.72, specificity = 0.96) and heroin (sensitivity = 0.84, specificity = 1.0). The method performed poorly at identifying adverse drug reactions and overdose due to patient error and fairly at identifying substance abuse in opioid-related unintentional overdose (sensitivity = 0.67, specificity = 0.96). Evaluation using validation datasets yielded significant reductions, in specificity and negative predictive values only, for many classifications mentioned above. However, these measures remained above 0.80, thus, performance observed during

PRIOR POSTINGS AND PRESENTATIONS: Preliminary versions of this work were presented as a poster at the 32nd International Conference on Pharmacoepidemiology and Therapeutic Risk Management (ICPE, 2016), with the title "Automating opioid overdose surveillance using natural language processing."

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 received prior funding from Purdue Pharma L.P. 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 L.P. to carry out related research.

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

development was largely maintained during validation. Similar results were obtained when evaluating portability, although there was a significant reduction in sensitivity for unintentional overdose that was attributed to missing text clinical notes in the database.

Conclusions: Methods that process text clinical notes show promise for improving accuracy and fidelity at identifying and classifying overdoses according to type using EHR data.

KEYWORDS

electronic health records, methods, natural language processing, opioid overdose, pharmacoepidemiology

1 | INTRODUCTION

The epidemic of opioid-related overdoses¹⁻⁹ demands development and application of methods for public health surveillance. Such methods need to accurately identify opioid-related overdoses so that existing interventions, such as state-based prescription drug monitoring programs,^{10,11} prescribing guidelines,¹²⁻¹⁴ risk evaluation and mitigation strategies (REMS),^{15,16} state and payer dose-related limitations and prescribing requirements,^{17,18} and abuse-deterrent opioid formulations, can be evaluated.^{19,20} In addition, the ability to classify opioid-related overdoses according to type—whether they were unintentional or suicides/attempts, whether or not they involved heroin, and whether or not substance abuse was involved—provides opportunities to design and evaluate preventive interventions for specific kinds of overdoses. To date, validated algorithms for identifying overdoses are based on diagnostic codes and are few.²¹⁻²³ Moreover, to our knowledge, there have been no attempts to classify opioid-related overdoses according to type.

To enhance the capability of diagnostic algorithms to identify and classify opioid-related overdoses using electronic health record (EHR) databases, we developed a natural language processing (NLP) software application to code clinical text documentation of health care services related to medication and drug overdose. The ultimate goal of this work was to determine if processing of clinical narratives documenting patient care can improve upon algorithms using solely administrative and diagnostic codes from EHR and claims data as part of a study to develop and validate an opioid overdose algorithm. In this paper, we report on development and evaluation of the NLP component. In a companion paper, we report on methods that combined this NLP component together with diagnostic codes from EHR data.²⁴

We developed an application of the MediClass (Medical Classifier) system²⁵ to identify overdose in EHR text clinical notes. The application identifies overdose and types of overdose using clinical notes from hospitalizations as well as from outpatient, telephone, ED, and urgent care visits when available. We developed the application using a “development dataset”²⁴ consisting of EHR records of the Kaiser Permanente Northwest (KPNW) patient population. We then validated the processor using a second “validation dataset”²⁴ from KPNW and a third “portability dataset”²⁴ from a distinct health system and patient

KEY POINTS

- Natural language processing technology (NLP) can be used to accurately identify overdose events and circumstances related to overdose in the text clinical documentation of electronic health records.
- Opioid overdoses involving heroin can be accurately identified and differentiated from opioid overdoses not involving heroin using NLP.
- Opioid overdoses that are suicides or suicide attempts can be identified with good accuracy using NLP.
- Opioid overdoses involving substance abuse can be identified with adequate accuracy using NLP.
- Key challenges remain in the data infrastructure of the health system that would enable access to all relevant data for comprehensive public health surveillance of opioid overdose events.

population at Kaiser Permanente Washington (KPW, known as Group Health Cooperative prior to 2017). Each dataset had a companion gold-standard dataset created by standardized chart review for the events of interest.²⁴ The study period was from January 1, 2008, to December 31, 2014. This time frame was chosen to maximize availability and recency of data during the ICD-9 era, which ended during 2015, because the main study was validating an ICD-9 algorithm.²⁴

2 | METHODS

2.1 | Study populations

KPNW and KPW are both nonprofit, group model integrated health systems serving patients in northern Oregon and southern Washington State (KPNW, ~500 000 patients) and greater Washington State (KPW, ~760 000 patients), respectively. The data for this study come from the

EHRs of these two health systems for patients selected into the three datasets during the study period, as described below.

2.2 | Study data

Each study record in the three datasets represents a patient event of interest (for example, an emergency department visit with a diagnosis code of opioid overdose) and typically included multiple encounters in the EHR (eg, ED visit, inpatient stay, outpatient visit, and telephone follow-up). Study records are created based on sampling criteria that are common to the three datasets, with the goal of making the three datasets similar but independent samples of their overall patient populations. The general sampling design was to create a balance of “suspected overdose” cases (ie, a qualifying event based on diagnosis code indicating opioid overdose) and any additional overdoses found by a manual review of records for patients who had no suspected overdose but were deemed “at risk” for an overdose due to specific clinical characteristics related to medical history and drug use. See companion paper for details about the sampling method used.²⁴

Each study record in each dataset entailed, typically, multiple EHR encounters; all within a window of time around the qualification date for the record. Data processed using NLP included all (and only) machine readable electronic text notes associated with each encounter. Trained abstractors used the EHR clinical interface to conduct chart audits that served as gold standard representation of the clinical case based on a set of criteria used to determine and classify overdoses. Additional information about chart audit methods is available in the companion paper.²⁴ It is worth noting that chart auditors using the EHR clinical interface had access to the entire clinical history of the patient, including multiple kinds of data, some of which are unavailable to the automated classifier because the data are not in a usable form (eg, scanned records from hospitals outside of the health system).

2.3 | Study datasets

The three datasets used in this study allowed for development, validation, and portability assessment of the NLP processor. Each dataset included a roughly equal balance of records selected by ICD9-CM diagnosis codes indicating a possible opioid overdose (called “suspected opioid overdose” cases), and records selected by ICD9-CM diagnosis codes indicating patients with elevated likelihood for an overdose due to a history of mental health conditions, chronic pain, substance abuse, and opioid use (called “at risk” cases). The latter includes overdoses found by manual chart review of the 2 years of medical history around the “at risk” qualification date. For each qualifying event (overdose diagnosis code, overdose determination by chart review, or date patient qualified as “at risk”), encounters within the period that began 3 days prior to and concluded 6 days following the qualifying event date were included as part of the study record. Only the machine readable text clinical notes of each encounter were processed by the NLP processor. The final dataset sizes—which included many records with no data available for the NLP processor—were 1006 records for the development dataset, 1696 records for the validation dataset, and

435 records for the portability dataset. Table 1 shows the dataset record counts according to the different sampling criteria as well as the base populations meeting criteria in each health system.

2.4 | Chart review gold standard

Chart audits were conducted on each study record in each dataset and constitute the gold standard for each dataset. The audit included determinations about presence or absence of overdose (or other event), the type of overdose (eg, heroin, opioid or not, polysubstance including opioid), and reasons for overdose (eg, suicide attempt, medication abuse, and medication misuse). See companion paper for details on the chart review process and the resulting gold standards for each dataset²⁴.

2.5 | NLP development and evaluation

MediClass (a “Medical Classifier”) is a general-purpose automated system for identifying study-specific information in electronic clinical records.²⁵ It classifies encounter records by applying a set of logical rules to clinical concepts that are automatically identified in free-text using NLP. MediClass is designed to employ a study-specific “knowledge module,” which defines the terms, concepts, and rules for combining concepts that are of interest. Rules can be chained together into sequences expressing decision logic for complex combinations of concepts that effectively classifies the record on a number of dimensions. MediClass applications have been developed and used for a diverse array of real-world assessments of health and care delivery using electronic medical records, including published studies in smoking cessation,²⁶ vaccine safety surveillance,²⁷ asthma care quality,²⁸ and weight loss counseling in gestational diabetes.²⁹

Details of how MediClass performs classification of encounter records have been published elsewhere.^{25,30} For this study, we built an application (ie, specified a knowledge module) to classify the electronic clinical notes of EHR encounter records according to indicators of overdose and type of overdose as recorded by documenting clinicians. The knowledge module can be visualized as a set of “rule trees” whose nodes represent concepts identified in the clinical text, subject to constraints represented by arcs that connect the nodes (see Figure 1). When applied to an encounter, a rule tree classifies the encounter positively when all of its constraints are satisfied by the clinical text contained in any discrete note of the encounter record. Figure 1 shows the final version of the rule tree that produces the classification Overdose. Conceptually, the overdose rule tree identifies an overdose when the documenting clinician notes it explicitly, a rescue attempt is made with naloxone, or there is language that combines a qualifying drug with poisoning, respiratory failure, or altered mental state.

As a second step, the encounter-level classifications were aggregated (for each patient and event date in the sample) across a period that began 3 days prior to and concluded 6 days following the event date, to generate final opioid overdose classifications that could be compared with chart review as the gold standard. Table 2 shows the full set of final classifications that were produced by this automated method, together with the comparison variables generated by chart audit.

TABLE 1 Sampling description for the three datasets

Sampling Goal	Events Identified by:	KPNW			KPW	
		Base Population	Development Dataset	Validation Dataset	Base Population	Portability Dataset
Suspected overdose cases	Opioid overdose codes	2271	483	848	750	188
	Adverse effects codes	254	78	102	0	0
At risk for overdose	Pain, mental health, substance abuse codes	87 550	222 with ≥ 30 d supply of ER/LA opioids	373 with ≥ 30 d supply of ER/LA opioids	33 438	247
			223 with ≤ 30 d supply of ER/LA opioids	372 with ≤ 30 d supply of ER/LA opioids		0
Total			1006	1695	34 188	435

Abbreviations: KPNW, Kaiser Permanente Northwest; KPW, Kaiser Permanente Washington.

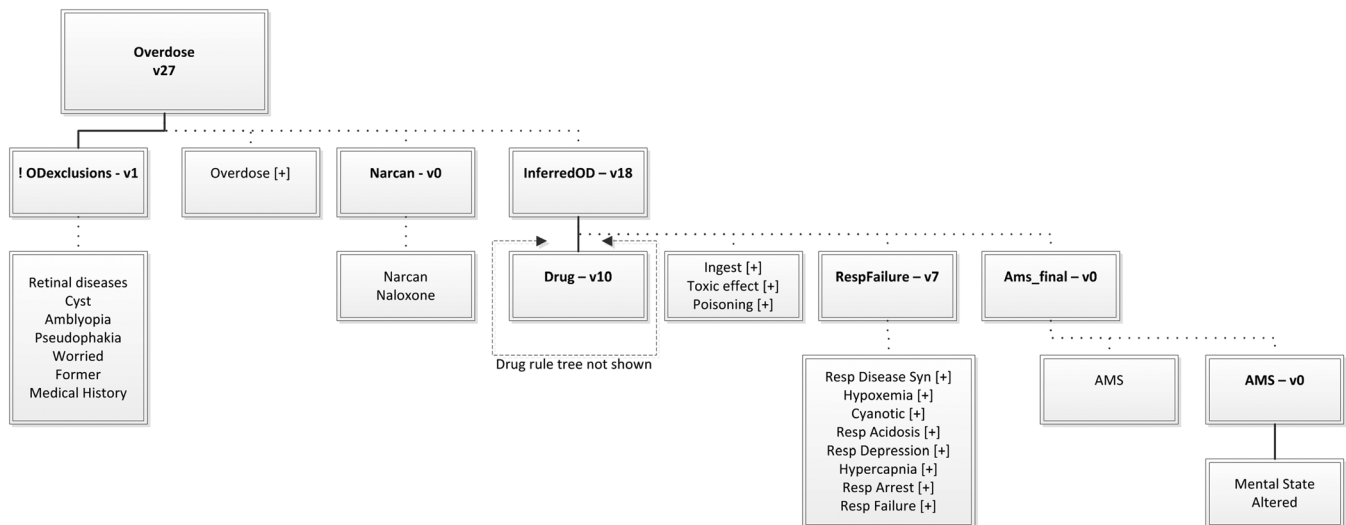


FIGURE 1 The overdose rule tree. A “rule tree” specifies a complex set of constraints that must be matched in the encounter data (text data within a single note, in this case) to generate a positive classification using the MediClass system. A rule tree is rooted by a single rule. A rule tree can be used to define a class alone or in combination with other rule trees. Each node in the tree is either a single rule (marked with a version number and shown in bold font) or it is one or more unified medical language system (UMLS) concepts (shown as plain font labels with no version numbers). Terms (not shown here) are child nodes of concepts, which help define how a concept is matched, using linguistic manipulations, against a sequence of tokens found in the text data. Terms are provided by the UMLS, as well as by custom additions found through trial and error in the development process, and constitute a lexicon of clinical expressions grouped by the concepts that they represent. Also not shown are proximity and ordering constraints, which govern relationships between concepts that are grouped by a rule. For example, a proximity constraint enforces a maximum allowable distance between any tokens (linguistic primitives of the text note) that participate in the identification of concepts within a rule. Every rule has one or more child nodes—each child node is connected to its parent by either the Boolean AND relation (shown with a solid line) or the Boolean OR relation (shown with a dashed line). For the parent node to match the data, all of the AND children and at least one of the OR children must match the data. Rules can take the following modifier: != Boolean NOT (ie, the reported match status of the rule is inverted from what is determined by normal match criteria for the rule). Concepts can take the following modifiers, which define constraints on how terms are matched in the text data: [-] = “negated form only” (ie, only terms of the concept asserted as negative in the text will create a match) [+]= “positive form only” (ie, only terms of the concept not asserted as negative in the text will create a match)

Development of the NLP processor entailed many iterations of the following three steps: (a) processing the entire development dataset, (b) evaluating performance compared to chart audit, and (c) modifying the knowledge module (terms, concepts, and rules) to remove the discordance between automated and chart audit-based classifications of events. When no additional improvements could be made using the development dataset compared to chart review as gold standard, we evaluated performance (sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]). We then ran the processor on the validation dataset and evaluated performance using those data.

Finally, we ran the NLP processor in the KPW data environment, on the portability dataset, and evaluated performance using those data.

3 | RESULTS

3.1 | Development dataset

Final performance of the automated method compared with chart audit for the development dataset is shown in Table 3. The entire dataset included 1006 study records. However, only 627 records were found

TABLE 2 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 Unintentional overdose (excludes intentional overdose) Overdose of any type (combines intentional and unintentional overdose) Adverse drug reaction—ADR (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 ADR	Heroin NLP identifies an overdose or adverse drug event in combination with heroin, regardless of other opioid or non-opioid 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 non-opioid 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 non-opioid prescription or over-the-counter medications, but in the absence of heroin. Any opioid (excludes heroin, includes 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, methamphetamine). Any substance abuse Any of the above types of abuse.	Opioid or non-opioid 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 non-prescribed substances = yes In an opioid overdose event Opioid or non-opioid prescription med or non-prescribed substance abuse or alcohol present = yes In an opioid overdose event
Patient error in opioid-related overdose or ADR	Patient error (excludes all abuse as defined above) NLP identifies an opioid overdose or adverse drug reaction in combination with mention of mistake/accident in taking medications	Opioid or non-opioid medication- taking error = yes AND NOT ABUSE (as defined above)

Abbreviation: NLP, natural language processing.

to have machine-readable clinical notes and thus the evaluation was performed using this 62% subset of the entire dataset. The remaining 38% of records in the development dataset were either marked during gold-standard chart review as having no contributing source data (ie, no text clinical notes found in the EHR) or there were no corresponding EHR encounters found in the KPNW EHR data warehouse.

Overall, the method performed well for identifying overdoses (sensitivity = 0.80, specificity = 0.93) and involvement of opioids excluding heroin (sensitivity = 0.72, specificity = 0.96) and for those involving heroin (sensitivity = 0.84, specificity = 1.0). In addition, the method performed well at identifying intentional overdose (sensitivity = 0.81, specificity = 0.98). The method performed poorly in identifying adverse drug reactions as opposed to overdose (sensitivity = 0.24, specificity = 0.93) and patient error in opioid-related

events (sensitivity = 0.33, specificity = 0.98). Performance was fair for identifying substance abuse involved in opioid-related unintentional overdose (sensitivity = 0.67, specificity = 0.96).

3.2 | Validation dataset

Following the development phase, the method was applied to the validation dataset and compared with chart review of those events (Table 4). The entire dataset included 1696 study records, balanced similarly to the development dataset. However, only 710 records were found to have machine-readable clinical notes and thus the evaluation was performed using this 42% subset of the full dataset. The remaining 58% of records in the validation dataset were either marked during gold-standard chart review as having no

TABLE 3 Evaluation of NLP application using development dataset

Development Dataset (KPNW) N = 1006		n ^a	Performance Compared with Gold Standard Chart Review n ^a = 627			
NLP-only Classification			Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Event type, irrespective of substance involved	Intentional overdose	74	.81 (.70-.89)	.98 (.96-.99)	.83 (.72-.91)	.97 (.96-.99)
	Unintentional overdose	158	.71 (.63-.78)	.94 (.91-.96)	.79 (.71-.85)	.91 (.87-.93)
	Overdose of any type	232	.80 (.74-.85)	.93 (.90-.95)	.87 (.81-.91)	.89 (.85-.92)
	Adverse drug reaction—ADR	79	.24 (.15-.35)	.93 (.90-.95)	.32 (.21-.45)	.89 (.87-.92)
Substance involved in overdose or ADR	Heroin	19	.84 (.60-.96)	1.0 (.99-1.0)	.80 (.56-.93)	1.0 (.99-1.0)
	Opioid only (excludes heroin)	98	.37 (.27-.47)	.94 (.92-.95)	.39 (.29-.50)	.93 (.91-.95)
	Polysubstance including opioid (excludes heroin)	175	.57 (.49-.64)	.96 (.94-.97)	.75 (.67-.82)	.91 (.89-.93)
	Any opioid (excludes heroin, includes polysubstance)	273	.72 (.66-.77)	.96 (.94-.97)	.88 (.83-.92)	.90 (.87-.92)
Substance abuse involved in opioid-related overdose	Prescription med abuse (whether prescribed or not)	56	.38 (.25-.51)	.96 (.94-.98)	.50 (.34-.66)	.94 (.92-.96)
	Substance abuse (including alcohol abuse or presence)	45	.62 (.47-.76)	.97 (.95-.98)	.58 (.43-.72)	.97 (.95-.98)
	Illicit drug abuse	84	.46 (.36-.58)	.97 (.97-.98)	.68 (.55-.80)	.92 (.90-.94)
	Any substance abuse	129	.67 (.59-.75)	.96 (.94-.97)	.81 (.72-.87)	.92 (.89-.94)
Patient error in opioid-related overdose or ADR	Patient error	21	.33 (.15-.57)	.98 (.96-.99)	.33 (.15-.57)	.98 (.96-.99)

Abbreviations: KPNW, Kaiser Permanente Northwest; NLP, natural language processing; NPV, negative predictive value; PPV, positive predictive value.

^an is the count as determined by gold standard chart review, includes only events for which EHR source data were available.

TABLE 4 Evaluation of NLP application using validation dataset

Validation dataset (KPNW), n = 1696		n ^a	Performance Compared with Gold Standard Chart Review, n ^a = 710			
NLP-only Classification			Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Event type, irrespective of substance involved	Intentional overdose	122	.74 (.65-.81)	.97 (.95-.98)	.82 (.73-.88)	.95 (.92-.96)
	Unintentional overdose	212	.66 (.59-.72)	.89 (.86-.92)	.73 (.66-.79)	.86 (.83-.89)
	Overdose of any type	334	.78 (.73-.82)	.89 (.85-.92)	.86 (.82-.90)	.82 (.78-.86)
	Adverse drug reaction—ADR	93	.31 (.22-.42)	.94 (.92-.96)	.44 (.32-.57)	.90 (.87-.92)
Substance involved in overdose or ADR	Heroin	54	.70 (.56-.82)	.99 (.98-1.0)	.86 (.72-.92)	.98 (.96-.99)
	Opioid only (excludes heroin)	166	.27 (.21-.35)	.88 (.85-.90)	.40 (.31-.50)	.80 (.76-.83)
	Polysubstance including opioid (excludes heroin)	199	.62 (.55-.69)	.84 (.81-.87)	.60 (.53-.67)	.85 (.82-.88)
	Any opioid (excludes heroin, includes polysubstance)	365	.75 (.70-.79)	.87 (.83-.90)	.86 (.81-.89)	.76 (.72-.80)
Substance abuse involved in opioid-related overdose	Prescription med abuse (whether prescribed or not)	87	.37 (.27-.48)	.93 (.91-.95)	.43 (.31-.55)	.91 (.88-.93)
	Substance abuse (including alcohol abuse or presence)	42	.67 (.50-.80)	.92 (.90-.94)	.34 (.24-.46)	.98 (.96-.98)
	Illicit drug abuse	107	.49 (.39-.58)	.93 (.90-.95)	.54 (.44-.64)	.91 (.88-.93)
	Any substance abuse	187	.72 (.65-.78)	.89 (.86-.92)	.71 (.64-.77)	.90 (.87-.92)
Patient error in opioid-related overdose or ADR	Patient error	33	.30 (.16-.49)	.97 (.96-.98)	.37 (.20-.58)	.97 (.95-.98)

Note. Bold values indicate significantly different measures ($P < .05$, chi-square test) from results using development dataset (see Table 3).

Abbreviations: KPNW, Kaiser Permanente Northwest; NLP, natural language processing; NPV, negative predictive value; PPV, positive predictive value.

^an is the count as determined by gold standard chart review, includes only events for which EHR source data were available.

contributing source data (ie, no text clinical notes found in the EHR) or there were no corresponding EHR encounters found in the KPNW EHR data warehouse.

Overall, performance was similar to what was observed in the development phase using the development dataset. For several classifications there was significantly reduced specificity and NPVs

(see bolded values in Table 4). Specificity and NPVs generally remained above 0.85 and always at or above 0.80, however. There were also significantly reduced PPV measures for polysubstance involvement and alcohol abuse classifications (see bolded values in Table 4). Overall, the performance observed during development was largely maintained during validation using the formerly untested validation dataset.

3.3 | Portability dataset

The procedures, technology, and process of the method were exported to the KPW data environment, and applied there to process the portability dataset, then compared with results of chart audits (Table 5). The entire dataset included 435 study records, balanced similarly to the development dataset. However, only 305 records were found to have machine-readable clinical notes, and thus the evaluation was performed using this 70% subset of the entire dataset. The remaining 30% of records in the portability dataset were either marked during gold-standard chart audit as having no contributing source data (ie, no text clinical notes found in the EHR) or there were no corresponding EHR encounters found in the KPW EHR data warehouse.

Overall, performance was similar to that observed in the development phase using the development dataset. There was significantly reduced sensitivity for unintentional overdose (and this also leading to reduced sensitivity in overdose, which includes both intentional and unintentional events—see bolded values in Table 5). Further investigations of reasons for this reduced sensitivity indicated that missing text clinical notes

(ie, encounters with notes indicating overdose when viewed in the EHR clinical interface had missing text notes in the extraction database) in just four cases played a substantial role. Reanalyzing the data by removing these incomplete study records from the original analysis yielded improved sensitivity measures of 0.59 (unintentional overdose) and 0.71 (any overdose), respectively. These results were compared with performance seen in the development dataset performance evaluation. They were not significantly different.

There were also significantly reduced sensitivities for heroin overdose and adverse drug event (see bolded values in Table 5). However, there were only five cases for each of these types of events in this small dataset, making evaluation of these classifications very uncertain. The only other significant differences from performance using the development dataset were seen in specificity and NPV for the substances involved in the overdose; however, these measures all remained at or above 0.92 (see bolded values in Table 5). On the whole, performance observed during the development phase was maintained in this test of the portability of the automated method.

4 | DISCUSSION

Automated methods for processing clinical notes documenting care in the EHR show promise for accurately identifying opioid-related overdose and some types of overdose. Although not all overdoses are captured in the EHR of the US health care system, many are, and the method we developed and evaluated could be deployed in

TABLE 5 Evaluation of NLP application using portability dataset

Portability dataset (KPW), n = 435		Performance Compared with Gold Standard Chart Review n ^a = 305				
NLP-only Classification		n ^a	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Event type, irrespective of substance involved	Intentional overdose	31	.65 (.45-.80)	.99 (.96-1.0)	.83 (.62-.95)	.96 (.93-.98)
	Unintentional overdose	43	.53 (.38-.69)	.96 (.92-.98)	.68 (.49-.82)	.93 (.89-.95)
	Overdose of any type	74	.68 (.56-.78)	.97 (.93-.98)	.86 (.74-.93)	.90 (.86-.94)
	Adverse drug reaction—ADR	5	0.0 (0.0-.54)	.97 (.94-.98)	0.0 (0.0-.34)	.98 (.96-.99)
Substance involved in overdose or ADR	Heroin	5	.20 (.01-.70)	.99 (.97-1.0)	.33 (.02-.87)	.99 (.96-1.0)
	Opioid only (excludes heroin)	26	.38 (.21-.59)	.94 (.90-.96)	.37 (.20-.58)	.94 (.91-.97)
	Polysubstance including opioid (excludes heroin)	42	.48 (.32-.63)	.97 (.94-.99)	.71 (.51-.86)	.92 (.88-.95)
	Any opioid (excludes heroin, includes polysubstance)	68	.69 (.57-.79)	.97 (.93-.98)	.85 (.88-.95)	.92 (.88-.95)
Substance abuse involved in opioid-related overdose	Prescription med abuse (whether prescribed or not)	18	.28 (.11-.54)	.98 (.95-.99)	.45 (.18-.75)	.96 (.92-.98)
	Substance abuse (including alcohol abuse or presence)	18	.44 (.22-.69)	.98 (.95-.99)	.53 (.27-.78)	.97 (.94-.98)
	Illicit drug abuse	25	.28 (.13-.50)	.99 (.97-1.0)	.70 (.35-.92)	.94 (.90-.96)
	Any substance abuse	44	.52 (.37-.67)	.98 (.95-.99)	.82 (.62-.93)	.92 (.88-.95)
Patient error in opioid-related overdose or ADR	Patient error	5	0.0 (0.0-.54)	.99 (.97-1.0)	0.0 (0.0-.80)	.98 (.96-.99)

Note. Bold values indicate measures that are significantly different ($P < .05$, chi-square test) from results using the development dataset (see Table 3).

Abbreviations: KPWN, Kaiser Permanente Northwest; NLP, natural language processing; NPV, negative predictive value; PPV, positive predictive value.

^an is the count as determined by gold standard chart review, includes only events for which EHR source data were available.

public health surveillance to aid in identifying the extent and types of opioid overdose events and evaluating efforts to curb those events. While efforts to use NLP to extract “phenotype” and outcomes data from the EHR have been steadily increasing,³¹ only recently has attention turned to using these methods to address the opioid epidemic. Carrell and colleagues have published a series of papers using NLP to identify problem opioid use,^{32,33} and others have evaluated similar methods for their potential to assist clinical improvement efforts in combatting abuse.³⁴ We are not aware of any published studies using NLP to identify opioid overdose and types of opioid overdose.

A large percentage of records in each dataset (30% to 58%) could not be used in the evaluations because no EHR encounter data with machine-readable text clinical notes were found during chart audit or during EHR data warehouse extraction. Instead, these records, even though they were reviewable within the human interface to the EHR, were often scanned faxes or pdf documents sourced from hospitals outside the health plan, legacy inpatient system data within the integrated delivery network that do not migrate to the EHR data warehouse, or claims data that could not provide the necessary data for text processing. While this large dropout of data in our evaluation effort was somewhat surprising and a bit disappointing, it faithfully represents the incomplete state of data systems for identifying opioid overdose events from clinical text documentation, particularly looking back in time, within the current infrastructure of the health care system.

There are limitations to the NLP method we used. For example, the sub-classification of overdose titled “opioid only” (defined in the gold standard as a single-opioid overdose event, see Table 2) had low sensitivity and PPV due to (a) many overdose events involved no mention of specific opioid drugs within the same context as the overdose discussion, or chart reviewers had access to medication records that were not available to the NLP method indicating patient was on a single specific opioid at the time of the overdose; (b) although the classification rules we developed identify opioid medications using a wide array of distinct drug classes and named brands and can separate opioid events from non-opioid events, the rules are unable to accurately separate single-opioid events from multiple-opioid events.

In this paper, we did not compare the NLP-based method with methods that rely on diagnosis or claims codes alone or methods that combine such codes with NLP processing of clinical text. However, see the companion paper for a report on these methods²⁴. Clearly, any inclusion of NLP methods will require infrastructure to gain access to the contents of EHR clinical notes, and this will add complexity and cost to any large scale public health surveillance effort. Therefore, it will be important to assess the tradeoff in implementation complexity and cost with any gains in measurement quality provided. This work provides important content for that discussion.

We anticipated that including variables extracted from text clinical notes might increase the accuracy and “fidelity” in methods to identify opioid overdose that use only diagnostic codes—some of those results are reported in the companion paper²⁴. By “fidelity,” we mean to highlight measurement accuracy for a range of overdose classifications (eg, identifying intentionality, polysubstance involvement, and substance

abuse) that are of great importance to public health but which may not be accurately captured in coded administrative data alone. In other words, although NLP methods for opioid overdose surveillance may come with increased complexity (and cost), they also promise to provide important information and public health insights not otherwise available. In conclusion, we believe that methods that process text clinical notes show promise for improving accuracy and fidelity at identifying and classifying overdoses according to type using EHR data.

ETHICS STATEMENT

The Institutional Review Boards (IRB) at KPNW and KPW approved the use of the data and procedures for this study.

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

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 comprised of companies that hold NDAs of extended-release and long-acting opioid analgesics and at the time of study conduct 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.; Mallinckrodt Inc.; Pernix Therapeutics Holdings, Inc.; Pfizer, Inc.; Purdue Pharma, LP.

ORCID

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

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

Nancy A. Perrin  <https://orcid.org/0000-0002-2025-6543>

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

REFERENCES

1. Okie S. A flood of opioids, a rising tide of deaths. *N Engl J Med*. 2010;363(21):1981-1985.
2. Johnson EM, Lanier WA, Merrill RM, et al. Unintentional prescription opioid-related overdose deaths: description of decedents by next of kin or best contact, Utah, 2008-2009. *J Gen Intern Med*. 2013;28(4):522-529.

3. Centers for Disease Control and Prevention. Overdose deaths involving prescription opioids among Medicaid enrollees—Washington, 2004-2007. *MMWR*. 2009;58(42):1171-1175.
4. Centers for Disease Control and Prevention. Vital signs: overdoses of prescription opioid pain relievers—United States, 199-2008. *MMWR*. 2011;60(43):1487-1492.
5. Warner M, Chen LH, Makuc DM, Anderson RN, Minino AM. Drug poisoning deaths in the United States, 1980-2008. *NCHS Data Brief*. 2011 Dec;(81):1-8.
6. Warner M, Chen LH, Makuc DM. Increase in fatal poisonings involving opioid analgesics in the United States, 1999-2006. *NCHS Data Brief*. 2009;22:1-8.
7. Unick GJ, Rosenblum D, Mars S, Ciccarone D. Intertwined epidemics: national demographic trends in hospitalizations for heroin- and opioid-related overdoses, 1993-2009. *PLoS ONE*. 2013;8(2):e54496.
8. Rudd RA, Paulozzi LJ, Bauer MJ, et al. Increases in heroin overdose deaths—28 states, 2010 to 2012. *MMWR*. 2014;63(39):849-854.
9. 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.
10. 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.
11. McCarty D, Bovett R, Burns T, et al. Oregon's strategy to confront prescription opioid misuse: a case study. *J Subst Abuse Treat*. 2015; 48(1):91-95.
12. 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.
13. Washington State Agency Medical Directors' Group (AMDG). Interagency Guideline on Prescribing Opioids for Pain 06/2015; 3rd. Available at <http://www.agencymeddirectors.wa.gov/Files/2015AMDGOpioidGuideline.pdf>. Accessed 11/17/2016.
14. 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.
15. 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.
16. U. S. Department of Health & Human Services. Abuse-Deterrent Opioids - Evaluation and Labeling: Guidance for Industry. U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER); 2015, Washington DC. Available Online at: <https://www.fda.gov/regulatory-information/search-fdaguidance-documents/abuse-deterrent-opioids-evaluation-and-labeling>
17. Franklin GM, Mai J, Turner J, Sullivan M, Wickizer T, Fulton-Kehoe D. Bending the prescription opioid dosing and mortality curves: impact of the Washington state opioid dosing guideline. *Am J Ind Med*. 2012;55(4):325-331.
18. Keast SL, Nesser N, Farmer K. Strategies aimed at controlling misuse and abuse of opioid prescription medications in a state Medicaid program: a policymaker's perspective. *Am J Drug Alcohol Abuse*. 2015;41(1):1-6.
19. Bannwarth B. Will abuse-deterrent formulations of opioid analgesics be successful in achieving their purpose? *Drugs*. 2012;72(13): 1713-1723.
20. Moorman-Li R, Motycka CA, Inge LD, Congdon JM, Hobson S, Pokropski B. A review of abuse-deterrent opioids for chronic nonmalignant pain. *Pharm Ther*. 2012;37(7):412-418.
21. 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.
22. 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.
23. 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.
24. Green CA, Perrin NA, Hazlehurst B, et al. Identifying and classifying opioid-related overdoses: a validation study. *Pharmacoepidemiol Drug Saf*. 2019. in press
25. Hazlehurst B, Frost HR, Sittig DF, Stevens VJ. MediClass: a system for detecting and classifying encounter-based clinical events in any electronic medical record. *J Am Med Inform Assoc*. 2005;12(5):517-529.
26. Hazlehurst B, Sittig DF, Stevens VJ, et al. Natural language processing in the electronic medical record: assessing clinician adherence to tobacco treatment guidelines. *Am J Prev Med*. 2005;29(5):434-439.
27. Hazlehurst B, Naleway A, Mullooly J. Detecting possible vaccine adverse events in clinical notes of the electronic medical record. *Vaccine*. 2009;27(14):2077-2083.
28. Hazlehurst B, McBurnie MA, Mularski RA, Puro JE, Chauvie SL. Automating care quality measurement with health information technology. *Am J Manag Care*. 2012;18(6):313-319.
29. Hazlehurst BL, Lawrence JM, Donahoo WT, et al. Automating assessment of lifestyle counseling in electronic health records. *Am J Prev Med*. 2014;46(5):457-464.
30. Hazlehurst BL, Kurtz SE, Masica A, et al. CER hub: an informatics platform for conducting comparative effectiveness research using multi-institutional, heterogeneous, electronic clinical data. *Int J Med Inform*. 2015;84(10):763-773.
31. Shivade C, Raghavan P, Fosler-Lussier E, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. *J Am Med Inform Assoc*. 2014;21(2):221-230.
32. Carrell DS, Cronkite D, Palmer RE, et al. Using natural language processing to identify problem usage of prescription opioids. *Int J Med Inform*. 2015 Dec;84(12):1057-1064.
33. Palmer RE, Carrell DS, Cronkite D, et al. The prevalence of problem opioid use in patients receiving chronic opioid therapy: computer-assisted review of electronic health record clinical notes. *Pain*. 2015 Jul;156(7): 1208-1214.
34. Haller IV, Renier CM, Juusola M, et al. Enhancing risk assessment in patients receiving chronic opioid analgesic therapy using natural language processing. *Pain Med*. 2016 Dec 29;pnw283.

How to cite this article: Hazlehurst B, Green CA, Perrin NA, et al. Using natural language processing of clinical text to enhance identification of opioid-related overdoses in electronic health records data. *Pharmacoepidemiol Drug Saf*. 2019;28:1143-1151. <https://doi.org/10.1002/pds.4810>