

Original Paper

Detecting Uncoded Self-Harm in Veterans' Electronic Health Records Using Positive and Unlabeled Learning: Retrospective Cohort Study

Praveen Kumar¹, PhD; Alexandria D Vizsolay¹, BS; Rajesh Upadhayaya¹, BS; Fariha Moomtaheen¹, MS; Donald R Greer², BBA; Cristian G Bologa^{1,3}, PhD; Kristan A Schneider¹, PhD; Sharon E Davis^{4,5}, PhD; Michael E Matheny^{4,5}, MD, MS, MPH; David van der Goes^{3,6}, PhD; Gerardo Villarreal^{3,7}, MD; Yiliang Zhu¹, PhD; Mauricio Tohen⁷, MBA, MD, DrPH; Scott A Malec^{1,3}, PhD; Jeremy J Yang^{1,3}, PhD; Elliot M Fielstein^{4,8}, PhD; Christophe Gerard Lambert^{1,3}, PhD

¹Department of Internal Medicine, School of Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM, United States

²Greer Black Company, Bozeman, MT, United States

³Raymond G. Murphy VA Medical Center, Albuquerque, NM, United States

⁴Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

⁵VA Tennessee Valley Healthcare System, Nashville, TN, United States

⁶Department of Economics, University of New Mexico, Albuquerque, NM, United States

⁷Department of Psychiatry and Behavioral Sciences, School of Medicine, University of New Mexico Health Sciences Center, Albuquerque, NM, United States

⁸Office of Mental Health, United States Department of Veterans Affairs, Washington, DC, United States

Corresponding Author:

Christophe Gerard Lambert, PhD
Department of Internal Medicine
School of Medicine, University of New Mexico Health Sciences Center
1 University of New Mexico, MSC10 5550
Albuquerque, NM 87131
United States
Phone: 1 505-272-9709
Email: cglambert@salud.unm.edu

Abstract

Background: Underdiagnosis and undercoding are common across mental health conditions, particularly suicide and self-harm. This leaves health care datasets lacking reliable negative examples needed for predictive modeling, phenotype prevalence estimation, and identification of individuals at elevated risk. We use positive and unlabeled (PU) learning to address this challenge.

Objective: This study aims to identify US Veterans whose self-harm events were not explicitly captured through diagnostic codes in electronic health records (EHRs) and estimate the underlying prevalence using a novel PU learning algorithm.

Methods: We performed a retrospective cohort study using Veterans Health Administration EHRs (from October 1, 1999, to August 31, 2019), selecting a random 25% sample of 1,329,120 Veterans out of 5,316,480 (1,193,563 males and 135,557 females) with at least 2 years of observation. The study cohort comprised 24,625 Veterans with coded self-harm and 1,304,495 uncoded, with the mean ages of 38.39 (SD 12.17) and 48.76 (SD 15.04) years, respectively. We applied our PULSNAR (positive unlabeled learning selected not at random) algorithm to estimate the proportion of individuals with uncoded self-harm. Covariates included age, medical conditions, procedures, and clinical observations. Four experts (raters) independently reviewed charts of 97 uncoded Veterans, each selected from 1% intervals of calibrated PULSNAR probabilities from 0.01 to 0.97. Agreement was assessed among raters, PULSNAR classifications, and consensus review decisions. Post hoc calibration was used to refine prevalence estimates.

Results: Of the 159,049 covariates in the dataset, PULSNAR's Extreme Gradient Boosting (XGBoost) model identified 1302 (0.82%) as informative for classification. Only 1.85% (24,625/1,329,120) of Veterans had diagnostic codes indicating self-harm events, while PULSNAR estimated an overall prevalence of 10.46% (139,026/1,329,120) by identifying an additional $\alpha=8.77\%$ (114,404/1,304,495) of self-harm cases among the uncoded population. Of the 97 chart-reviewed patients,

39 had documented but uncoded self-harm. PULSNAR probabilities were post hoc calibrated such that their sum over the 97 cases equaled 39, which adjusted the combined coded and imputed prevalence downward from 10.46% to 7.91% (105,133/1,329,120). By applying this calibration to shift the probabilities of all uncoded Veterans, with bootstrapping for confidence intervals, PULSNAR estimates that coded self-harm represents only 23.4% (95% CI 17.76% to 31.51%) of all documented (coded+notes) self-harm.

Conclusions: Under the “selected not at random” assumption, PULSNAR provides an innovative and scalable framework for estimating the clinically documented prevalence of mental health conditions and identifying the uncoded individuals with calibrated prediction, without requiring confirmed negative labels. This method offers an alternative to time-consuming chart reviews for detecting likely cases missing structured coding capture. By addressing diagnostic undercoding of mental health conditions in EHRs, this approach has the potential to enhance the estimation of mental health prevalence and support screening, activation of automated clinical decision support, targeted intervention, better resource allocation, and research to improve outcomes in real-world settings.

J Med Internet Res 2026;28:e89071; doi: [10.2196/89071](https://doi.org/10.2196/89071)

Keywords: self-injurious behavior; machine learning; Veterans' health; PULSNAR; electronic health record; positive unlabeled learning selected not at random

Introduction

Suicide and self-harm remain significant public health concerns in the United States, consistently ranking among the leading causes of death. In 2023, suicide was among the top 8 causes of death for individuals aged 10-64 years and the second leading cause for those aged 10-34 years [1,2]. According to the Centers for Disease Control and Prevention, suicide accounted for 49,316 deaths in 2023 [1], 1.6% of all deaths. The prevalence of suicidal thoughts and attempts is even higher than suicide deaths. In 2023, an estimated 12.8 million American adults aged 18 years or older experienced serious thoughts of suicide, 3.7 million made a suicide plan, and 1.5 million attempted suicide [1,3]. These data suggest that for each suicide death, there are hundreds of individuals experiencing suicidal ideation or engaging in self-harm.

Veterans are at a disproportionately higher risk of experiencing trauma and self-destructive behavior compared to the general population, accounting for nearly 14% of adult suicide deaths despite representing only 7.6% of the population [4]. The overall unadjusted suicide rate among Veterans was roughly double that of non-Veteran adults (34.7 per 100,000 vs 17.1 per 100,000) [5]. Both male (37.3 vs 28.7 per 100,000) and female Veterans (13.5 vs 7.2 per 100,000) had markedly higher suicide rates than their non-Veteran counterparts. In addition, younger Veterans (ages 18-34 years) faced the highest risk (47.6 per 100,000)—far exceeding that of any other age group [5]. Co-occurring conditions, such as posttraumatic stress disorder, depression, bipolar disorder, substance use disorders, traumatic brain injury, and prior self-injurious behavior, significantly contribute to this elevated risk [4,6].

Suicide rarely occurs in isolation; it is often preceded by identifiable risk factors, with trauma, past self-harm, and suicidal ideation being among the strongest predictors of future suicide risk [7,8]. Early identification of individuals exhibiting these behaviors is critical, as timely intervention can significantly reduce the risk of suicide. Studying self-harm, therefore, is essential for comprehensive suicide prevention efforts. Within the Veterans Health Administration

(VHA), policy and quality metrics primarily focus on suicidal self-directed violence within the past 12 months, rather than self-harm broadly defined. Although our study centers on ever self-harm as a phenotype, it is important to distinguish this analytic focus from operational VHA surveillance, which prioritizes identification and follow-up of suicidal behaviors in the near-term versus capturing all self-harm behaviors with and without intent.

The widespread adoption of electronic health records (EHRs) in the United States has generated large repositories of patient health care data, comprising both structured data (eg, demographics, diagnoses, procedures, and prescriptions) and unstructured data (eg, clinical notes, imaging, and pathology reports) [9-11]. While these data are invaluable for observational studies and health care analytics, they also reveal inconsistencies in data quality across different clinical settings and heterogeneous data structures that span structured codes, free-text notes, and temporal measurements [11]. Additionally, these data are affected by missingness due to multiple mechanisms, including the incomplete capture of out-of-network care, underreporting of sensitive conditions, and not documenting or undercoding of clinical diagnoses. Undercoding refers to instances in which events or conditions, such as self-harm behaviors and mental illnesses, are not recorded in structured diagnosis codes in the EHR, leading to physicians overlooking past diagnoses, inaccurate prevalence estimates, and hindering effective intervention strategies. Importantly, undercoding of mental health conditions, including suicidality and self-harm, is common in EHRs, limiting research that relies on accurate reporting of these conditions [12-15]. VHA's suicide risk surveillance relies heavily on policy-driven, standardized national note templates such as the Suicide Behavior and Overdose Report (SBOR), which capture suicidal and other self-directed violence events in structured clinical documentation. These templates function in parallel to diagnostic coding and are a primary mechanism by which VHA monitors suicide behaviors. Undercoding of ICD (*International Classification of Diseases*) self-harm diagnoses reflects partially the extent of VHA suicide risk monitoring efforts.

The increased availability of EHRs and advancements in machine learning (ML) methodologies have led to increased application of ML techniques to identify and predict instances of self-harm and suicidal ideation using EHRs and insurance claims data [12,13,15-22]. Kumar et al [12] and Nestsiarovich et al [13] used an Extreme Gradient Boosting (XGBoost) [23] model on visit-level data to estimate the uncoded self-harm events among individuals with major mental illness. Simon et al [16] developed random forest-based models aimed at predicting fatal or nonfatal self-harm events within 90 days following a sampled encounter. Simon et al [17] used LASSO (least absolute shrinkage and selection operator)-based logistic regression models to predict suicide attempts and suicide deaths post-outpatient visits. Rozova et al [18] used natural language processing (NLP) supervised learning techniques on free-text triage notes from emergency department (ED) visits to detect self-harm and suicidal ideation among ED patients. Walsh et al [19] used random forest and nonregularized logistic regression models on longitudinal clinical data to detect the risk of nonfatal suicide attempts in adolescents. Tsui et al [20] leveraged NLP models on clinical notes along with 4 ML techniques, including Naïve Bayes, LASSO regression, random forest, and an ensemble of extreme gradient boosting, for the prediction of first-time suicide attempts. Su et al [21] used LASSO logistic regression to predict suicidal behavior among children and adolescents based on their longitudinal clinical records, identifying both short- and long-term risk factors. Barak-Corren et al [22] developed Bayesian models using a retrospective cohort approach to predict future documented suicidal behavior.

Previous studies used traditional positive-negative ML classifiers to identify/predict instances of self-harm in health records. However, due to undercoding common in mental health data, unlabeled instances contain both positive (diagnosed and undiagnosed) and negative (unaffected) cases, leading to biased classification and predictions. To address this, we applied a novel positive and unlabeled (PU) learning algorithm, PULSNAR (positive unlabeled learning selected not at random) [24], to Veterans' EHR data to estimate the proportion of Veterans with ever self-harm. Notably, we are not predicting future self-harm, but rather classifying whether patients had experienced self-harm at any point during the study period. PU learning is a semisupervised approach that uses labeled (coded) positive examples and unlabeled (uncoded) examples containing an unknown mixture of positives and negatives [24].

To our knowledge, this is the first study to use PU learning algorithms to estimate the proportion of Veterans ever with self-harm imputing uncoded individuals in Veterans' EHR data. By applying PULSNAR, we aim to improve the detection and estimation of self-harm prevalence among US Veterans, demonstrating a framework applicable to detecting other undetected mental health diagnoses. This approach supports earlier screening as well as novel intervention strategies to reduce self-harm and suicide rates in this high-risk population. Our findings are likely to improve awareness of risk factors of self-destructive behaviors among

Veterans and highlight the broader use of PU learning in mental health informatics.

Methods

PU Learning Background

In many real-world applications, annotating all records can be challenging, expensive, or even impossible due to the volume of data. Often, only positive instances are labeled, leaving a considerable portion of the data unlabeled [25]. Notably, an unlabeled instance does not necessarily indicate a negative case because the absence of a diagnosis code does not confirm the absence of a condition. Given that only a fraction of records are labeled, learning from PU data has emerged as an active area of research [24-32]. The majority of current PU learning methods are based on the "selected completely at random" (SCAR) assumption [25-30], which posits that labeled positives are independent and identically distributed (i.i.d.) random samples from the positive distribution, meaning that the probability of an instance being labeled as positive is independent of its attributes [26]. However, in real-world applications, this assumption often does not hold due to labeling bias; for example, patients with more severe or specific symptoms are more likely to be labeled with a disease than those with milder or unspecific symptoms.

Despite the prevalence of real-world non-SCAR PU data, only a few studies have focused on PU learning under the non-SCAR assumption [24,31,32]. In this study, we have applied PULSNAR, a PU algorithm based on the SNAR (selected not at random) assumption. PULSNAR uses a divide-and-conquer approach to cluster SNAR positives into subtypes and estimate the proportion of each subtype among the unlabeled using the PULSCAR (positive unlabeled learning selected completely at random) algorithm. Under the SNAR assumption, the probability that a positive example is labeled depends on its attributes [24,32], which is more appropriate for health care data for which labeling bias is prevalent.

PU Learning Algorithms to Estimate the Proportion of Self-Harm Among Uncoded Individuals

We found that existing state-of-the-art PU learning methods suffer from scalability issues and fail to execute on large datasets [24]. As a result, in earlier work, we developed 2 novel PU learning algorithms aimed at estimating the proportion (α) of positive instances among unlabeled examples and subsequently imputing these instances: PULSCAR for when the SCAR assumption holds, and PULSNAR for when it does not. What sets our methods apart from other PU methods is that they not only estimate α but also calculate calibrated probabilities using the estimated α , leading to markedly improved classification performance. We focus our analysis on the application of the PULSNAR framework to universally available structured EHR data; extensive methodological comparisons demonstrating

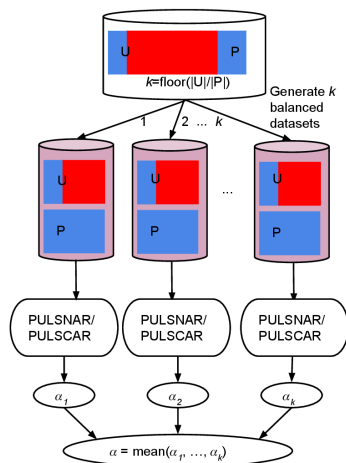
its superiority over standard supervised baselines have been detailed in our prior work [24,33].

EHR data may contain various types of self-harm events, such as hanging, poisoning, cutting, etc, and the likelihood of coding these different types of self-harm may differ due to the severity of the condition and the source and nature of the underlying trauma. Therefore, it becomes evident that the SCAR assumption may not hold for self-harm data, and SCAR-based PU methods may not be suitable for such data. Recognizing this limitation, we applied the PULSNAR algorithm to estimate the proportion of self-harm-uncoded patients in an EHR dataset of US Veterans.

When dealing with highly class-imbalanced training datasets, ML algorithms often exhibit a bias toward the majority class, struggling to effectively generalize patterns from the minority class, which yields biased predictions [34]. Since only a tiny fraction of records were coded for self-harm in EHR data, we adopted a random undersampling approach [35], where we transformed the imbalanced dataset into k balanced datasets (as illustrated in Figure 1), where $k = \text{floor}(\text{unlabeled}/\text{labeled})$. Each balanced dataset comprised all labeled self-harm records along with a similar

number of unlabeled records. Subsequently, we applied PULSCAR and PULSNAR algorithms (with XGBoost as the classifier) to each balanced dataset to estimate the proportion (α) of self-harm among the unlabeled records. The final α was determined by computing the mean of the k estimates of α . We performed hyperparameter tuning to determine optimal values for the XGBoost parameters. To better explore the high-dimensional covariate space, we used aggressive feature subsampling combined with a low learning rate and a large ensemble of deeper trees. This configuration limits early dominance by a small set of predictors and promotes broader exploration of the covariate space. With this sampling scheme, deeper trees provide more opportunities for additional variables to enter each tree and enable the model to capture complex interactions among covariates. Although deeper trees may increase the risk of overfitting, the large sample size and cross-validation procedures mitigate this risk. The final XGBoost parameter values were $\text{max_depth}=12$, $\text{n_estimators}=400$, $\text{learning_rate}=0.05$, $\text{min_child_weight}=1$, $\text{colsample_bytree}=0.3$, $\text{colsample_bylevel}=0.6$, $\text{n_jobs}=32$, $\text{random_state}=0$, and $\text{objective}='binary:logistic'$. Other parameters were kept at their default values.

Figure 1. Steps for α estimation when the dataset has high class imbalance. The mix of blue and red represents the unlabeled set (U), and blue represents the positive set (P). To address class imbalance, k balanced datasets are generated, where $k = \text{floor}(\text{unlabeled}/\text{labeled})$. Each balanced dataset includes all labeled records and a randomly selected subset of unlabeled records of comparable size. The PU algorithm is applied to each of the k balanced datasets to estimate the proportion (α) of positives within the unlabeled set. The final α estimate is obtained by averaging the k individual α estimates. PULSNAR: positive unlabeled learning selected not at random; PULSCAR: positive unlabeled learning selected completely at random.



Data Source

For this study, we used the Veterans Health Administration (VHA) EHR data (from October 1, 1999, to August 31, 2019) available in the OMOP CDMv5 (Observational Medical Outcomes Partnership Common Data Model) [36] format, selecting a random 25% sample of 1,329,120 Veterans (out of 5,316,480). The only inclusion criterion was that Veterans had at least 2 years of enrollment. Ethical approvals and data use agreements were obtained from the appropriate institutional review boards to ensure compliance with privacy and confidentiality regulations.

Phenotyping and Covariate Selection

A self-harm phenotype was defined by the presence of one or more ICD-10-CM (International Classification of Diseases, 10th Revision, Clinical Modification) or ICD-9-CM codes (Table S1 in Multimedia Appendix 1). These codes encompass all instances of intentional self-harm or suicide attempts by any means, including a history of self-harm. Patients with any of these codes were labeled as positive cases, while all others remained unlabeled. Accurate identification of self-harm events in structured data is complicated by self-directed violence nomenclature and coding rules that require documentation of clinical intent. Injury and poisoning codes must distinguish accidental events, intentional self-harm, assault, or undetermined causes, and in the absence

of clear documentation, intent is often coded as accidental. These challenges are not unique to VHA; they affect coding practices in non-VHA systems and research cohorts as well.

Covariates included patient age at enrollment and the presence/absence of medical conditions, procedures, and clinical observations over the duration of patient observation. *ICD-9-CM* and *ICD-10-CM* diagnosis codes were mapped to their Systematized Nomenclature of Medicine (SNOMED) equivalents (and all ancestors thereof) using the OMOP vocabulary [37]. Procedure codes from *ICD-9-CM Volume 3 (ICD-9-CM V3)*, the *ICD-10 Procedure Coding System (ICD-10-PCS)*, and *Current Procedural Terminology, Fourth Edition*, were mapped to *ICD-10-PCS* concepts (and all ancestors thereof). Overall, from a dataset of 1,329,120 Veterans, we selected 159,049 covariates for use in the

PU learning algorithm. Codes used to define the self-harm phenotype (Table S1 in Multimedia Appendix 1) were excluded from the covariate list to prevent data leakage. Since each patient had an average of only 1203 nonzero features, a compressed sparse row (CSR) matrix with 1,329,120 rows (patients) and 159,049 columns (covariates) was created as input for the PU models. Covariates were encoded as binary values (0/1) in the CSR matrix; if a covariate was not present in an individual’s data, it was set to 0, and if it was present, it was set to 1. Therefore, no covariate included in the modeling framework contained missing values. As a result, tests for missingness (eg, missing completely at random testing) and multiple imputation procedures were not applicable for our study. Figure 2 illustrates the complete schema of our study.

Figure 2. Study schema. First, k balanced datasets were generated, and the PU learning method was applied to each of them. The parameter α_j was estimated for each balanced dataset, and the final α for the complete dataset was computed by taking the mean of the k estimated values $\alpha_1, \dots, \alpha_k$. Using the estimated α , the PU method calculated the calibrated probability of being labeled as positive for each uncoded individual. These calibrated classifications were then used to determine probable positive (PP) and probable negative (PN) individuals among the uncoded individuals. CP: coded positive; OMOP: Observational Medical Outcomes Partnership; PULSNAR: positive unlabeled learning selected not at random; VHA: Veterans Health Administration.

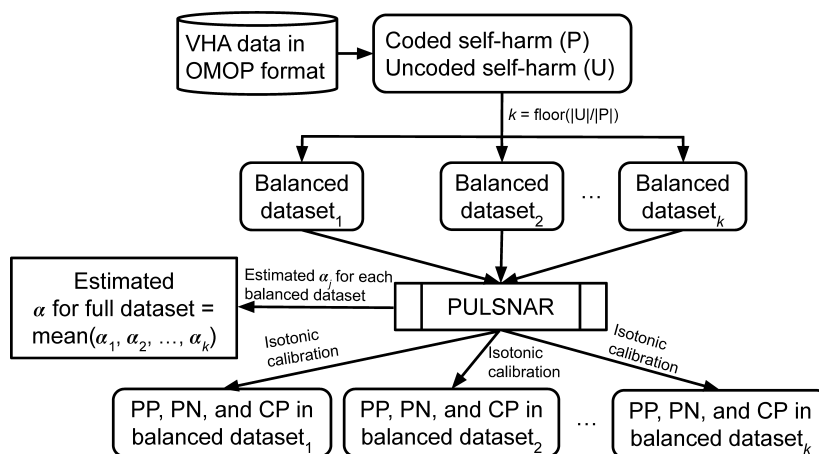


Chart Review Process

To validate our model’s classifications, we conducted a chart review of a random sample of unlabeled individuals whose calibrated probabilities fell into each 1% probability bin. Due to the absence of unlabeled individuals in the calibrated probability bins in the first and last two percentiles, the selection was limited to 97 individuals. To ensure rigorous clinical validity, a comprehensive set of search keywords was established a priori in direct collaboration with our clinical psychiatrist coauthor (GV). Four informaticists (coauthors CGB, DvdG, JJY, and CGL) used these clinically vetted keywords and search utilities to identify and extract potential evidence of suicidal or nonsuicidal self-harm from combined charts, which often exceeded 500,000 lines per patient. Interrater reliability was assessed at this stage to ensure consistency among the informatics reviewers. Crucially, the extracted clinical evidence for each of the 97 cases was subsequently evaluated and discussed in detail with the psychiatrist. Rather than limiting clinical oversight to disagreement resolution among informaticists, this joint case-by-case review ensured that definitional and semantic complexities (eg, distinguishing

when chronic substance abuse strictly constitutes intentional self-harm) were uniformly adjudicated by a mental health clinician. Through this expert-guided process, individuals were manually classified as positive (class 1) or negative (class 0) for ever self-harm. Finally, we compared the sum of the PULSNAR-calibrated probabilities of the 97 individuals with the sum of their consensus labels identified through expert chart review.

Calibrated Self-Harm Prevalence Estimates for Veterans

To calibrate the PULSNAR-estimated fraction of patients with self-harm against consensus chart review results, we applied a bias-only logit shift—a logistic transformation that adds a uniform constant offset (c) to the logits of the original probabilities without altering their ordering. Mathematically, for each probability p_i , the logit is computed as $l_i = \log\left(\frac{p_i}{1-p_i}\right)$; c is solved numerically (via root-finding) such that $\sum_{i=0}^{|U|} \frac{1}{1 + \exp(-l_i + c)}$ is equal to the number of positives confirmed by chart review.

The adjusted probabilities are $\hat{p}_i = \frac{1}{1 + \exp(-(l_i + \hat{c}))}$. This method preserves relative ordering and can shift probabilities lower when correcting overestimation, providing exact sum alignment to the gold standard. We quantified calibration uncertainty using a 100,000-sample bootstrap, sampling chart-reviewed observations with replacement, solving for c each time to produce an empirical distribution. The 2.5th and 97.5th percentiles formed the 95% CI for \hat{c} , which we then applied to the unlabeled population to bound the total estimated positives.

PULSNAR Classification vs Expert Chart Review

The probability of a self-harm diagnosis was stratified into low (probability < 25%), intermediate (25% ≤ probability ≤ 75%), or high (probability > 75%) categories. These stratified probability categories (which also serve as a proxy measure for the underlying behavioral risk or detection and identification of such) were compared with the outcome of the chart reviews stratified into unanimous decision against self-harm (ie, all reviewers agreed that no evidence of self-harm was found in the patient chart), dissenting evidence for self-harm (ie, some but not all reviewers agreed on self-harm), and unanimous decision for self-harm (ie, all reviewers agreed on self-harm).

Ethical Considerations

Ethical approval for this study was obtained from the institutional review board and research and development committees of the VA New Mexico Healthcare System (H2947), University of New Mexico Health Sciences Center (20-477), and VA Tennessee Valley Healthcare System (#1576576). The requirement for informed consent was waived because the study involved secondary analysis of existing electronic health record data and posed minimal risk to participants. All data were deidentified prior to analysis, and access to the data was restricted to authorized study personnel and maintained on secure, password-protected VA servers to ensure privacy and confidentiality. No participants were contacted for this study, and no compensation was provided. No identifiable individual-level information or images are included in this manuscript or supplementary materials.

Results

Patient Characteristics

Applying our inclusion and exclusion criteria, we identified 1,329,120 individuals (1,193,563 males and 135,557 females) for the study. Table 1 summarizes the key demographic and clinical characteristics of the study population.

Table 1. Demographic and clinical characteristics of patients with and without coded self-harm selected from Veterans Health Administration electronic health record data (October 1, 1999, to August 31, 2019). The listed comorbidities represent key covariates identified as important features by the XGBoost^a model within the PULSNAR^b algorithm.

Patient characteristics (n=1,329,120)	Coded for self-harm (n=24,625, 1.85%)		Uncoded for self-harm (n=1,304,495, 98.15%)		OR ^c (95% CI)
	Patients, n (%)	95% CI	Patients, n (%)	95% CI	
Male	20,723 (84.15)	83.69-84.61	1,172,840 (89.91)	89.86-89.96	0.60 (0.58-0.62)
Female	3902 (15.85)	15.39-16.31	131,655 (10.09)	10.04-10.14	1.68 (1.62-1.74)
Age (years), mean (SD)	38.39 (12.17)	— ^d	48.76 (15.04)	—	—
Age (years)					
0-19	1301 (5.28)	5.01-5.57	32,128 (2.46)	2.44-2.49	2.21 (2.09-2.34)
20-29	6126 (24.88)	24.34-25.42	168,060 (12.88)	12.83-12.94	2.24 (2.17-2.31)
30-39	5424 (22.03)	21.51-22.55	164,091 (12.58)	12.52-12.64	1.96 (1.90-2.02)
40-49	7261 (29.49)	28.92-30.06	257,077 (19.71)	19.64-19.78	1.70 (1.66-1.75)
50-59	3712 (15.07)	14.63-15.53	365,610 (28.03)	27.95-28.10	0.46 (0.44-0.47)
≥60	801 (3.25)	3.03-3.48	317,529 (24.34)	24.27-24.41	0.10 (0.10-0.11)
Comorbidities					
Personality disorder	2536 (10.30)	9.92-10.68	5627 (0.43)	0.42-0.44	26.48 (25.22-27.82)
Bipolar disorder	2374 (9.64)	9.27-10.02	5887 (0.45)	0.44-0.46	23.53 (22.38-24.72)
Schizophrenia	944 (3.83)	3.60-4.08	2277 (0.17)	0.17-0.18	22.79 (21.09-24.64)
Major depressive disorder	4412 (17.92)	17.44-18.40	13,572 (1.04)	1.02-1.06	20.76 (20.00-21.53)
Psychoactive substance use disorder	3783 (15.36)	14.91-15.82	10,933 (0.84)	0.82-0.85	21.47 (20.63-22.34)
Posttraumatic stress disorder	3563 (14.47)	14.03-14.91	10,369 (0.79)	0.78-0.81	21.10 (20.27-22.00)
Attention-deficit/hyperactivity disorder	509 (2.07)	1.89-2.25	1266 (0.10)	0.09-0.10	21.72 (19.56-24.11)
Chronic pain	2538 (10.31)	9.93-10.69	6977 (0.53)	0.52-0.55	21.37 (20.37-22.41)
Adjustment disorder	2748 (11.16)	10.77-11.56	7763 (0.60)	0.58-0.61	20.98 (20.03-21.97)
Sleep disorder	3570 (14.50)	14.06-14.94	10,744 (0.82)	0.81-0.84	20.41 (19.59-21.25)

Patient characteristics (n=1,329,120)	Coded for self-harm (n=24,625, 1.85%)		Uncoded for self-harm (n=1,304,495, 98.15%)		OR ^c (95% CI)
	Patients, n (%)	95% CI	Patients, n (%)	95% CI	
Traumatic brain injury	56 (0.23)	0.17-0.30	101 (0.01)	0.01-0.01	29.45 (20.84-41.21)
Anxiety disorder	4424 (17.97)	17.49-18.45	13,855 (1.06)	1.04-1.08	20.40 (19.66-21.16)
Induced psychotic disorder	21 (0.09)	0.05-0.13	34 (0.00)	0.00-0.00	32.75 (18.05-58.10)

^aXGBoost: Extreme Gradient Boosting.

^bPULSNAR: positive unlabeled learning selected not at random.

^cOR: odds ratio.

^dNot applicable.

Performance of PU Models

Table 2 presents the estimated proportion of self-harm cases by the PULSCAR and PULSNAR algorithms. The proportion of individuals coded for self-harm was only 1.85% (24,625/1,329,120). As expected, PULSCAR provided a lower α estimate (21,524/1,304,495, 1.65%) compared to

PULSNAR (114,404/1,304,495, 8.77%), because the EHR self-harm data of Veterans did not satisfy the SCAR assumption. PULSCAR estimated 3.47% (46,120/1,329,120) of the individuals with coded and imputed self-harm, while PULSNAR provided a higher estimate of 10.46% (139,026/1,329,120).

Table 2. Performance of the PULSCAR^a and PULSNAR^b algorithms in imputing uncoded self-harm cases from Veterans Health Administration (VHA) electronic health record data. The parameter α denotes the estimated proportion of uncoded self-harm cases identified by each method.

	VHA (ever self-harm)
Records, n	1,329,120
Coded self-harm	24,625 (1.85%, 95% CI 1.83%-1.88%)
Covariates, n	159,049
Covariate types	3 (condition, procedure, and observation)
Important covariates, n	1302
α using PULSCAR	1.65%
α using PULSNAR	8.77%
Coded+imputed self-harm using PULSCAR	3.47%
Coded+imputed self-harm using PULSNAR	10.46% (14 clusters)

^aPULSCAR: positive unlabeled learning selected completely at random.

^bPULSNAR: positive unlabeled learning selected not at random.

Important Covariates Used by the XGBoost Model

Our covariate selection approach identified 159,049 covariates for the XGBoost model used in the PULSCAR and PULSNAR methods, but XGBoost identified only 1302

(0.82%) as informative for classification. Feature (covariate) importance was measured by the average gain (mean reduction in the loss function) contributed by each covariate. The top 15 covariates with the highest gain scores are shown in Table 3.

Table 3. Top 15 covariates identified by the XGBoost^a model, along with their corresponding gain scores. Of the 159,049 covariates included in the positive and unlabeled models based on XGBoost, only 1302 (0.82%) were identified by the model as contributing to training in our cohort. Feature (covariate) importance was quantified using average gain, defined as the mean reduction in the model's loss function attributable to splits on a given covariate across all trees; thus, higher gain values indicate a greater contribution to improving the model's predictive performance.

OMOP ^b concept ID	Concept name	Gain score
438028	Poisoning by a drug AND/OR a medicinal substance	21.61
4009713	Thoughts of self-harm	14.62
440270	Poisoning by antidepressants	11.18
2840354	Medical and Surgical @ Anatomical Regions, Upper Extremities @ Supplement @ Axilla, Left @ Open @ Autologous Tissue Substitute	10.68
44782421	Acute deep venous thrombosis of the upper extremity	10.50
4133169	Operative procedure on the pelvis	10.44
434626	Borderline personality disorder	9.87
4306645	Finding of thought content	9.73

OMOP ^b concept ID	Concept name	Gain score
2887059	Medical and Surgical @ Gastrointestinal System @ Bypass @ Descending Colon @ Open	9.59
4181019	Cluster B personality disorder	9.49
2832323	Medical and Surgical @ Anatomical Regions, General @ Repair @ Abdominal Wall @ Percutaneous @ No Device	9.47
2841968	Imaging @ Lower Arteries @ Fluoroscopy @ Abdominal Aorta @ Low Osmolar	9.30
2762368	Supplement Thoracic Vertebra with Nonautologous Tissue Substitute, Percutaneous Approach	9.27
437456	Poisoning by an anticonvulsant	8.90
2895851	Medical and Surgical @ Upper Bones @ Excision @ Clavicle, Left @ Open	8.87

^aXGBoost: Extreme Gradient Boosting.

^bOMOP: Observational Medical Outcomes Partnership.

Chart Review

Table 4 presents the interreviewer agreement among 4 reviewers across 97 uncoded self-harm cases, as well as the agreement between individual reviewers and PULSNAR, and between reviewers and the overall reviewer consensus. Table S2 in Multimedia Appendix 1 reports the pairwise Cohen κ coefficients for these agreements. Interreviewer reliability

among 4 reviewers across 97 uncoded self-harm cases was substantial (Fleiss $\kappa=0.668$, $z=16.1$, $P<.001$). Notably, there were 11 patients of 39 (28.2%) where the consensus review was positive, but at least one reviewer missed the relevant evidence. Of $4 \times 39=156$ evaluations of positive charts, 17 (10.9%) were false negatives.

Table 4. Agreement between PULSNAR^a and each reviewer, between individual reviewers, between each reviewer and their consensus, and between the reviewers' consensus and PULSNAR for 97 uncoded self-harm cases. Agreement is calculated as the percentage agreeing out of the 97 charts. Overall interreviewer reliability was substantial (Fleiss $\kappa=0.668$, $z=16.1$, $P<.001$). Among 39 consensus-positive cases, 11 (28.2%) had at least one reviewer miss relevant evidence.

	Reviewer 1	Reviewer 2	Reviewer 3	Reviewer 4	PULSNAR	Reviewer consensus
Reviewer 1	100.0%	89.7%	94.9%	82.5%	58.8%	88.7%
Reviewer 2	89.7%	100.0%	88.7%	76.3%	54.6%	82.5%
Reviewer 3	94.9%	88.7%	100.0%	79.4%	59.8%	87.6%
Reviewer 4	82.5%	76.3%	79.4%	100.0%	63.9%	85.6%
PULSNAR	58.8%	54.6%	59.8%	63.9%	100.0%	59.8%
Reviewer consensus	88.7%	82.5%	87.6%	85.6%	59.8%	100.0%

^aPULSNAR: positive unlabeled learning selected not at random.

For completeness, we also calculated standard classification metrics on the 97 chart-reviewed cases using the PULSNAR probabilities versus expert consensus as the ground truth. The model demonstrated an area under the curve (AUC) of 0.6813, an F_1 -score of 0.5517, a precision of 0.5, and a recall of 0.6154.

Post Hoc Calibration

After applying the bias-only logit shift on the PULSNAR probabilities for the 97 reviewed charts, using their consensus chart-reviewed self-harm labels, the sum of transformed PULSNAR probabilities equaled 39 (the expert count), with a logit shift parameter of $c=-0.54$ (bootstrap 95% CI -1.1420 to -0.0252). Applying this transformation to the 1,304,495 unlabeled patients yielded a sum of adjusted probabilities of 80,574.7, corresponding to 6.18% positives among the uncoded patients, which is lower than the PULSNAR-estimated α of 8.77%. This suggests that, if all notes for all Veterans without coded self-harm were chart-reviewed, 6.18% (95% CI 4.1%-8.74%) would reveal a documented history of self-harm. Thus, coded self-harm represents approximately $24,625/(80,574.7+24,625)=23.4\%$ (95% CI 17.76%-31.51%) of all documented (coded+notes) self-harm.

Efficiency/Scalability Quantification

To quantify the operational utility and scalability of the PULSNAR approach, we evaluated both the manual effort required for expert chart review and the computational resources consumed by our model inference across the full cohort. Expert chart review time ranged from approximately 25 minutes to 2 hours per patient chart, depending on the complexity of the record and the volume of clinical notes. Across the 97 charts reviewed in this study, this corresponds to an estimated total effort of approximately 40-200 person-hours per reviewer. In contrast, running the PULSNAR model for inference on the entire cohort of 1.3 million records was computationally tractable, requiring approximately 63 hours of wall-clock time on a machine with 16 vCPUs, 128 GB of RAM, and 128 GB of disk space.

PULSNAR Classification vs Expert Chart Review

Table 5 shows the comparison of PULSNAR-classified self-harm risk categories with chart review outcomes. Among those Veterans for whom expert reviewers could not find evidence of self-harm in some patient charts, PULSNAR classified the probability of self-harm as low in 35.42% of the

cases and found it to be intermediate or high in the remaining 64.58% of the cases. Classifying patients as intermediate or high risk despite the absence of self-harm coding is a desirable feature, as the patient chart might also not document self-harm. Among those Veterans for whom some, but not all, expert reviewers agreed on recorded self-harm behavior in the patient chart, PULSNAR would classify the probability of self-harm to be low only in 14.81% of the cases,

and intermediate to high in 85.19% of the cases. Among those Veterans for whom the expert reviewers unanimously agreed on evidence of self-harm behavior in the patient chart, PULSNAR estimated the probability of self-harming behavior as low only in 13.64% of the cases, and intermediate to high in 86.36% of the cases. Stratified as such, there was an association between expert chart reviewers and PULSNAR (Fisher exact test: $P=.02$).

Table 5. Comparison of PULSNAR^a-classified self-harm risk categories (low, intermediate, and high probability) with expert chart review outcomes (unanimous against self-harm, dissenting evidence for self-harm, and unanimous for self-harm) among 97 chart-reviewed uncoded cases. There was a statistically significant association between PULSNAR risk categories and chart review outcomes (Fisher exact test, $P=.02$).

Expert review	PULSNAR classification		
	Low	Intermediate	High
Unanimous no self-harm	35.42	50.00	14.58
Dissenting self-harm	14.81	40.74	44.44
Unanimous self-harm	13.64	68.18	18.18

^aPULSNAR: positive unlabeled learning selected not at random.

Discussion

Principal Results

In this study, we successfully applied the novel PULSNAR algorithm to a large representative cohort of US Veterans' electronic health records to estimate the burden of self-harm beyond what is captured in structured diagnostic codes. Our primary finding is that relying solely on structured diagnostic codes dramatically underestimates the clinically documented prevalence of self-harm. By using only structured data within a PU learning framework and calibrating predictions against expert chart review of clinical notes, our approach provides a more comprehensive estimate of self-harm prevalence and highlights the extent of undercoding in VA electronic health records. These results support our central hypothesis that PU learning can recover hidden disease burden in routinely collected health care data. Critically, our results suggest that only about 1 in 4 patients with clinical notes documenting self-harm, or a history thereof, have the condition captured in structured diagnostic codes. While the discrimination metrics (AUC=0.68) appear modest, this is primarily because the model relies solely on structured data for prediction, while the validation ground truth is derived from a comprehensive review of free-text clinical notes. Furthermore, the nonrandom, uncoded test set, which lacks confirmed negatives, is optimized not for discrimination assessment, but rather prevalence calibration, which remains the study's primary, successful objective.

Self-harm, like many sensitive mental health conditions, is frequently underreported in EHRs due to confidentiality concerns, stigma, limited help-seeking, barriers to care, and inconsistent screening and documentation practices across health care settings [38]. Additionally, self-harm is likely underdocumented in ICD codes in VHA records because VHA facilities receive federal funding on a per-enrolled-patient basis rather than through per-service billing, reducing incentives for exhaustive diagnostic coding [39]. It should

be noted that VHA suicide risk surveillance does not rely solely on ICD codes. Policy and standardized templates such as SBOR are used to systematically document suicidal behaviors and overdoses [40], meaning that ICD undercoding does not present a complete picture of VHA's operational monitoring of suicide risk. This underreporting hinders accurate prevalence estimation, risk identification, resource allocation, risk modeling, and intervention design [12,13,41, 42]. To address these gaps, we applied PULSNAR to estimate the proportion of Veterans with both coded and uncoded self-harm and to identify likely cases at scale, serving as a case study for broader phenotyping of undercoded mental illnesses.

Veterans' health records often contain hundreds of thousands of lines of clinical notes, making it challenging and time-consuming for clinicians to manually identify individuals at risk of self-harm, suicidality, or other mental health phenotypes through chart review [43,44]. More broadly, the self-directed violence nomenclature itself is often confusing to frontline clinicians, and the expectation of uniform, intent-based coding is difficult to meet in routine care [45]. Because injury and poisoning codes must encode intent, incomplete or ambiguous documentation pushes events toward "accidental" categories, hindering differentiation between nonsuicidal self-harm and suicidal self-directed violence [46]. This is a fundamental limitation of current surveillance approaches across health systems, and it constrains the accuracy of both administrative data and research phenotypes. Further, while one might hope that a patient's problem list would contain a history of self-harm if it was ever noted, we found that only 22.6% (5556/24,625) of patients with coded self-harm ever had self-harm or self-harm history recorded in their VHA problem list. Given the high workload and time constraints faced by health care providers, thoroughly reviewing lengthy patient records can significantly slow the assessment process, potentially delaying risk assessment and timely intervention [47]. Another structural barrier is the absence of required, standardized training

for VHA or non-VHA clinicians on self-directed violence nomenclature and the use of structured diagnostic codes for suicide risk surveillance. In practice, documentation and coding are shaped by local culture, individual comfort, and time constraints rather than uniform training, which contributes to inconsistent recognition, labeling, and coding of self-harm and suicidal behaviors across settings [48,49]. The PULSNAR method addresses this limitation by efficiently estimating the probability of self-harm through analysis of patient data, thereby assisting clinicians in prioritizing high-risk individuals for further evaluation.

Our expert chart review of 97 unlabeled Veterans yielded 3 key insights relevant to validating algorithmic phenotypes. First, probability calibration is essential and can be achieved with limited chart review. Applying post hoc calibration to the full cohort resulted in a more conservative estimate of self-harm prevalence at 7.91%, in contrast to PULSNAR's estimate of 10.46%. Both estimates, however, are consistent with ranges reported in prior studies of Veteran populations, supporting the plausibility of our findings [50-52]. Still, here we privilege the calibrated estimate tied to observed human review.

Second, the main human-human discrepancies were both definitional and false negatives from information overload. The largest contributor to divergence was the treatment of "punching objects" (eg, walls/doors) without stated self-harm intent. One reviewer initially counted many such episodes as nonsuicidal self-injury (NSSI) (preconsensus count 45), whereas other reviewers generally did not—applying an intent or foreseeability standard. Borderline cases (eg, punching a window with tendon laceration) were debated under a reasonable foreseeability framework, absent self-harm intent. Similarly, statements about long-horizon self-destruction (eg, "drinking myself to death") were generally classified as substance use disorder unless the episode reflected acute, explicit self-harm intent (eg, deliberate overdose). Under an inclusive NSSI sensitivity definition that includes "punching objects" [53,54], the consensus chart review identified 39 positives. This is 9 fewer than PULSNAR's estimate of 48, which is expected and desirable, as PULSNAR is designed to provide an upper bound on the proportion of positives among unlabeled cases. Notably, 28.2% of positives had at least one reviewer missing chart evidence—a real concern, as physicians have limited time to comb through notes, unlike our reviewers who used systematic search tools and spent more time per chart than a typical visit duration.

Third, information asymmetries between data modalities, for example, notes versus codes, explain much of the remaining gap [55]. For example, manual reviewers who examined unstructured clinical notes identified several cases of self-harm history (typically past suicide attempts prior to Veteran enrollment) that were documented only in narrative notes but not captured in structured data fields. In contrast, PULSNAR relied exclusively on structured data and would have missed these cases. Conversely, some PULSNAR "high-risk" cases lacked explicit self-harm documentation in notes but exhibited risk constellations in coded data

(injuries/poisonings, major mental illness, and substance use disorders). This pattern is expected given the information asymmetry between notes and codes, reinforcing that neither source is complete [55,56].

Contribution

We emphasize two contributions. (1) Epidemiologic impact: After calibration, PULSNAR yields a population-scale estimate of ever self-harm that materially exceeds code-based prevalence. Notably, prior Veteran studies report ever NSSI between ~6% and 16%, including samples with rates of 14%-16% [50-52], providing an external range against which our calibrated estimate can be interpreted. (2) Operational utility: PULSNAR enables triage of >1.3 million records, focusing human effort (and/or targeted NLP) where the marginal value of chart review is highest. This offers a blueprint for identifying other undercoded mental health diagnoses. In our experience, expert review of 97 charts required substantial person-hours, whereas model inference over the full cohort is computationally tractable; thus, even moderate accuracy can generate large efficiency gains by shrinking the manual search space.

Our findings support a hybrid workflow generalizable to other undetected conditions: (1) run PULSNAR on structured data to estimate the probability of being a positive case of self-harm (or other target phenotype) for uncoded individuals, (2) select a limited number of uncoded individuals from each probability bin of PULSNAR-estimated probabilities, (3) perform chart review for those selected uncoded individuals, and (4) apply post hoc calibration using chart-reviewed labels and PULSNAR-estimated probabilities of uncoded individuals to obtain calibrated cohort estimates with uncertainty. These steps, collectively, convert a complex validation into actionable epidemiology and a scalable detection pathway for underdocumented self-harm or other mental illnesses.

Limitations

First, this study used a single data source, VHA EHRs, whose patient population, coding practices, and care patterns differ systematically from those of other US and international health systems. Accordingly, the generalizability of our findings to Veterans treated exclusively outside the VHA or to non-Veteran populations needs validation. Second, since the true α is generally unidentifiable [57], PULSNAR estimates an upper bound on α across different positive subtypes. Thus, the corresponding predicted probabilities may be overestimated. Third, the prevalence estimate relies on the assumption that the post hoc calibration factor derived from a small, stratified sample of 97 chart-reviewed cases accurately transfers to the entire unlabeled population of 1.3 million Veterans, a potential source of error we partially addressed with bootstrap resampling to quantify uncertainty. Finally, while we validated this method using chart review for self-harm, which is often documented in notes but uncoded, this approach cannot confirm cases absent from both notes and codes. Furthermore, other conditions may follow different recording mechanisms; for instance, posttraumatic stress disorder in the VHA is typically coded if documented due to disability benefit incentives.

Consequently, extending this framework to such diagnoses may require validating against the incidence of future coded diagnoses rather than concurrent notes. Because EHR data do not contain reliable true negative labels for uncoded self-harm, conventional supervised classification metrics such as AUC-ROC, precision-recall, and F_1 -scores cannot be unbiasedly estimated [58]. This limitation is common in PU learning settings where only confirmed positives are available [24]. The chart-reviewed subset comprised only previously unlabeled cases and was not selected as a representative validation cohort with confirmed positives and negatives. Accordingly, model evaluation focused on prevalence estimation and agreement with expert review rather than traditional supervised classification metrics.

Finally, our findings should be interpreted in the context of broader limitations of suicide risk surveillance. VHA policy and clinical operations prioritize suicidal self-directed violence within the past 12 months and rely on policy-driven documentation tools such as the SBOR [40,59], whereas our phenotyping targets ever self-harm based primarily on ICD coding. Self-directed violence nomenclature and intent-based coding rules are complex [60], and there is no required, standardized training for VHA or non-VHA clinicians on their application [61]; as a result, documentation and coding of self-harm and suicidal behavior are inconsistent across settings and clinical providers. These surveillance and training constraints likely contribute to undercoding and misclassification and should be considered when interpreting our prevalence estimates and model performance.

Future Directions

Future research could validate our PU learning algorithms in other populations and health care settings and across

a broader range of mental health diagnoses. Incorporating unstructured data, such as clinical notes, through NLP techniques, may further enhance the detection of uncoded self-harm instances and develop valid and reliable scales to measure self-harm. Additionally, integrating our approach into clinical workflows could facilitate real-time identification of at-risk individuals, enabling timely intervention.

Conclusions

Our study demonstrates the effectiveness of PU learning algorithms under the SNAR assumption in identifying uncoded instances of self-harm among US Veterans. PULSNAR can support both population-level prevalence estimation and individual-level risk stratification using structured health data, although differences between structured billing codes and clinical notes may affect concordance with manual chart reviews. Our findings reveal a significantly higher prevalence of self-harm than what is captured in diagnostic codes, emphasizing the urgent need for more accurate detection and imputation methods. Our approach offers a scalable and efficient adjunct to manual chart reviews for detecting undetected mental illness diagnoses, with the potential to enhance clinical practice, inform policy decisions, support comparative effectiveness studies with imputed phenotypes [13], improve predictive modeling of self-harm and other conditions, and ultimately contribute to reducing suicide rates among Veterans and improving mental health. We encourage the adoption of similar methods in other health care systems to address undercoding challenges, improve patient outcomes, and advance the application of ML in health care analytics.

Acknowledgments

We gratefully acknowledge Edgar J Villarreal, PhD, and his team from the VA Office of Suicide Prevention for their invaluable guidance on the nuances of Veterans Health Administration policy and operational practices, which substantially strengthened the contextual clarity of this work. Generative artificial intelligence tools were not used at any stage of this study, including study design, methodology development, literature review, code generation, or manuscript writing and editing. However, Grammarly was used as a spell and grammar checker to improve the clarity of the manuscript.

Funding

This research was supported by funding from the US National Institutes of Health, specifically, the National Institute of Mental Health grant R01MH129764, the National Library of Medicine grant R00LM013367, and infrastructure support from the National Center for Advancing Translational Sciences grants UL1TR001449 and UM1TR005466. The views expressed in this paper are those of the authors and do not necessarily reflect those of the National Institutes of Health. Access to VA data and computational environments provided by the Department of Veterans Affairs (VA) Informatics and Computing Infrastructure (VINCI), funded under the research priority to Put VA Data to Work for Veterans (VA ORD 22-D4V).

Conflicts of Interest

None declared.

Multimedia Appendix 1

ICD-CM codes and Cohen κ coefficients. ICD-CM: *International Classification of Diseases, Clinical Modification*. [DOCX File (Microsoft Word File), 1754 KB-Multimedia Appendix 1]

References

1. Facts about suicide. Centers for Disease Control and Prevention. Mar 26, 2025. URL: <https://www.cdc.gov/suicide/facts/index.html> [Accessed 2026-05-09]

2. 10 leading causes of death, United States, 2024. Centers for Disease Control and Prevention. URL: <https://wisqars.cdc.gov/lcd/> [Accessed 2026-05-09]
3. 2022 National Survey on Drug Use and Health (NSDUH) Annual National Report. Substance Abuse and Mental Health Services Administration. URL: <https://www.samhsa.gov/data/report/2023-nsduh-annual-national-report> [Accessed 2026-05-09]
4. Ruiz F, Burgo-Black L, Hunt SC, Miller M, Spelman JF. A practical review of suicide among veterans: preventive and proactive measures for health care institutions and providers. *Public Health Rep.* 2023;138(2):223-231. [doi: [10.1177/00333549221085240](https://doi.org/10.1177/00333549221085240)] [Medline: [35403486](https://pubmed.ncbi.nlm.nih.gov/35403486/)]
5. 2024 National Veteran Suicide Prevention Annual Report. U.S. Department of Veterans Affairs; Dec 2024. URL: https://www.mentalhealth.va.gov/docs/data-sheets/2024/2024-Annual-Report-Part-2-of-2_508.pdf [Accessed 2026-05-09]
6. Schafer KM, Duffy M, Kennedy G, et al. Suicidal ideation, suicide attempts, and suicide death among veterans and service members: a comprehensive meta-analysis of risk factors. *Mil Psychol.* 2022;34(2):129-146. [doi: [10.1080/08995605.2021.1976544](https://doi.org/10.1080/08995605.2021.1976544)] [Medline: [38536290](https://pubmed.ncbi.nlm.nih.gov/38536290/)]
7. Ribeiro JD, Franklin JC, Fox KR, et al. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. *Psychol Med.* Jan 2016;46(2):225-236. [doi: [10.1017/S0033291715001804](https://doi.org/10.1017/S0033291715001804)] [Medline: [26370729](https://pubmed.ncbi.nlm.nih.gov/26370729/)]
8. Predescu E, Sipos R. Self-harm behaviors, suicide attempts, and suicidal ideation in a clinical sample of children and adolescents with psychiatric disorders. *Children (Basel).* Apr 14, 2023;10(4):725. [doi: [10.3390/children10040725](https://doi.org/10.3390/children10040725)] [Medline: [37189974](https://pubmed.ncbi.nlm.nih.gov/37189974/)]
9. Banda JM, Seneviratne M, Hernandez-Boussard T, Shah NH. Advances in electronic phenotyping: from rule-based definitions to machine learning models. *Annu Rev Biomed Data Sci.* Jul 2018;1:53-68. [doi: [10.1146/annurev-biodatasci-080917-013315](https://doi.org/10.1146/annurev-biodatasci-080917-013315)] [Medline: [31218278](https://pubmed.ncbi.nlm.nih.gov/31218278/)]
10. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. *J Am Med Inform Assoc.* Dec 2013;20(e2):e206-11. [doi: [10.1136/amiajnl-2013-002428](https://doi.org/10.1136/amiajnl-2013-002428)] [Medline: [24302669](https://pubmed.ncbi.nlm.nih.gov/24302669/)]
11. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J Am Med Inform Assoc.* Oct 1, 2018;25(10):1419-1428. [doi: [10.1093/jamia/ocy068](https://doi.org/10.1093/jamia/ocy068)] [Medline: [29893864](https://pubmed.ncbi.nlm.nih.gov/29893864/)]
12. Kumar P, Nestsiarovich A, Nelson SJ, Kerner B, Perkins DJ, Lambert CG. Imputation and characterization of uncoded self-harm in major mental illness using machine learning. *J Am Med Inform Assoc.* Jan 1, 2020;27(1):136-146. [doi: [10.1093/jamia/ocz173](https://doi.org/10.1093/jamia/ocz173)] [Medline: [31651956](https://pubmed.ncbi.nlm.nih.gov/31651956/)]
13. Nestsiarovich A, Kumar P, Lauve NR, et al. Using machine learning imputed outcomes to assess drug-dependent risk of self-harm in patients with bipolar disorder: a comparative effectiveness study. *JMIR Ment Health.* Apr 21, 2021;8(4):e24522. [doi: [10.2196/24522](https://doi.org/10.2196/24522)] [Medline: [33688834](https://pubmed.ncbi.nlm.nih.gov/33688834/)]
14. Hedegaard H, Schoenbaum M, Claassen C, Crosby A, Holland K, Proescholdbell S. Issues in developing a surveillance case definition for nonfatal suicide attempt and intentional self-harm using International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) coded data. *Natl Health Stat Report.* Feb 2018(108):1-19. [Medline: [29616901](https://pubmed.ncbi.nlm.nih.gov/29616901/)]
15. Edgcomb JB, Thiruvalluru R, Pathak J, Brooks JO. Machine learning to differentiate risk of suicide attempt and self-harm after general medical hospitalization of women with mental illness. *Med Care.* Feb 1, 2021;59:S58-S64. [doi: [10.1097/MLR.0000000000001467](https://doi.org/10.1097/MLR.0000000000001467)] [Medline: [33438884](https://pubmed.ncbi.nlm.nih.gov/33438884/)]
16. Simon GE, Shortreed SM, Johnson E, et al. Predicting risk of suicidal behavior from insurance claims data vs. linked data from insurance claims and electronic health records. *Pharmacoepidemiol Drug Saf.* Jan 2024;33(1):e5734. [doi: [10.1002/pds.5734](https://doi.org/10.1002/pds.5734)] [Medline: [38112287](https://pubmed.ncbi.nlm.nih.gov/38112287/)]
17. Simon GE, Johnson E, Lawrence JM, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *Am J Psychiatry.* Oct 1, 2018;175(10):951-960. [doi: [10.1176/appi.ajp.2018.17101167](https://doi.org/10.1176/appi.ajp.2018.17101167)] [Medline: [29792051](https://pubmed.ncbi.nlm.nih.gov/29792051/)]
18. Rozova V, Witt K, Robinson J, Li Y, Verspoor K. Detection of self-harm and suicidal ideation in emergency department triage notes. *J Am Med Inform Assoc.* Jan 29, 2022;29(3):472-480. [doi: [10.1093/jamia/ocab261](https://doi.org/10.1093/jamia/ocab261)] [Medline: [34897466](https://pubmed.ncbi.nlm.nih.gov/34897466/)]
19. Walsh CG, Ribeiro JD, Franklin JC. Predicting suicide attempts in adolescents with longitudinal clinical data and machine learning. *J Child Psychol Psychiatry.* Dec 2018;59(12):1261-1270. [doi: [10.1111/jcpp.12916](https://doi.org/10.1111/jcpp.12916)] [Medline: [29709069](https://pubmed.ncbi.nlm.nih.gov/29709069/)]
20. Tsui FR, Shi L, Ruiz V, et al. Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts. *JAMIA Open.* Jan 2021;4(1):ooab011. [doi: [10.1093/jamiaopen/ooab011](https://doi.org/10.1093/jamiaopen/ooab011)] [Medline: [33758800](https://pubmed.ncbi.nlm.nih.gov/33758800/)]

21. Su C, Aseltine R, Doshi R, Chen K, Rogers SC, Wang F. Machine learning for suicide risk prediction in children and adolescents with electronic health records. *Transl Psychiatry*. Nov 26, 2020;10(1):413. [doi: [10.1038/s41398-020-01100-0](https://doi.org/10.1038/s41398-020-01100-0)] [Medline: [33243979](https://pubmed.ncbi.nlm.nih.gov/33243979/)]
22. Barak-Corren Y, Castro VM, Javitt S, et al. Predicting suicidal behavior from longitudinal electronic health records. *Am J Psychiatry*. Feb 1, 2017;174(2):154-162. [doi: [10.1176/appi.ajp.2016.16010077](https://doi.org/10.1176/appi.ajp.2016.16010077)] [Medline: [27609239](https://pubmed.ncbi.nlm.nih.gov/27609239/)]
23. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery; 2016:785-794. [doi: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)]
24. Kumar P, Lambert CG. Positive unlabeled learning selected not at random (PULSNAR): class proportion estimation without the selected completely at random assumption. *PeerJ Comput Sci*. 2024;10:e2451. [doi: [10.7717/peerj-cs.2451](https://doi.org/10.7717/peerj-cs.2451)] [Medline: [39650456](https://pubmed.ncbi.nlm.nih.gov/39650456/)]
25. Jaskie K, Elkan C, Spanias A. A modified logistic regression for positive and unlabeled learning. In: *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE; 2019. [doi: [10.1109/IEEECONF44664.2019.9048765](https://doi.org/10.1109/IEEECONF44664.2019.9048765)]
26. Elkan C, Noto K. Learning classifiers from only positive and unlabeled data. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery; 2008:213-220. [doi: [10.1145/1401890.1401920](https://doi.org/10.1145/1401890.1401920)]
27. Du Plessis MC, Sugiyama M. Class prior estimation from positive and unlabeled data. *IEICE Trans Inf Syst*. 2014;E97.D(5):1358-1362. [doi: [10.1587/transinf.E97.D.1358](https://doi.org/10.1587/transinf.E97.D.1358)]
28. Ramaswamy H, Scott C, Tewari A. Mixture proportion estimation via kernel embeddings of distributions. In: *Proceedings of the 33rd International Conference on Machine Learning*. JMLR; 2016:2052-2060. [doi: [10.5555/3045390.3045607](https://doi.org/10.5555/3045390.3045607)]
29. Bekker J, Davis J. Estimating the class prior in positive and unlabeled data through decision tree induction. *Proc AAAI Conf Artif Intell*. 2018;32(1). [doi: [10.1609/aaai.v32i1.11715](https://doi.org/10.1609/aaai.v32i1.11715)]
30. Ivanov D. DEDPUL: difference-of-estimated-densities-based positive-unlabeled learning. In: *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE; 2020:782-790. [doi: [10.1109/ICMLA51294.2020.00128](https://doi.org/10.1109/ICMLA51294.2020.00128)]
31. Gerych W, Hartvigsen T, Buquicchio L, Agu E, Rundensteiner E. Recovering the propensity score from biased positive unlabeled data. *Proc AAAI Conf Artif Intell*. 2022;36(6):6694-6702. [doi: [10.1609/aaai.v36i6.20624](https://doi.org/10.1609/aaai.v36i6.20624)]
32. Bekker J, Robberechts P, Davis J. Beyond the selected completely at random assumption for learning from positive and unlabeled data. In: *Machine Learning and Knowledge Discovery in Databases*. Springer; 2020:71-85. [doi: [10.1007/978-3-030-46147-8_5](https://doi.org/10.1007/978-3-030-46147-8_5)]
33. Kumar P, Moomtaheen F, Malec SA, et al. Detecting opioid use disorder in health claims data with positive unlabeled learning. *IEEE J Biomed Health Inform*. Feb 2025;29(2):750-757. [doi: [10.1109/JBHI.2024.3515805](https://doi.org/10.1109/JBHI.2024.3515805)] [Medline: [40030473](https://pubmed.ncbi.nlm.nih.gov/40030473/)]
34. Pozzolo AD, Caelen O, Johnson RA, Bontempi G. Calibrating probability with undersampling for unbalanced classification. In: *2015 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE; 2015:159-166. [doi: [10.1109/SSCI.2015.33](https://doi.org/10.1109/SSCI.2015.33)]
35. Philip K, Chan SJ. Toward scalable learning with non-uniform class and cost distributions: a case study in credit card fraud detection. In: *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*. AAAI Press; 1998:164-168. [doi: [10.5555/3000292.3000320](https://doi.org/10.5555/3000292.3000320)]
36. Voss EA, Makadia R, Matcho A, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. *J Am Med Inform Assoc*. May 2015;22(3):553-564. [doi: [10.1093/jamia/ocu023](https://doi.org/10.1093/jamia/ocu023)] [Medline: [25670757](https://pubmed.ncbi.nlm.nih.gov/25670757/)]
37. Lynch KE, Deppen SA, DuVall SL, et al. Incrementally transforming electronic medical records into the observational medical outcomes partnership common data model: a multidimensional quality assurance approach. *Appl Clin Inform*. Oct 2019;10(5):794-803. [doi: [10.1055/s-0039-1697598](https://doi.org/10.1055/s-0039-1697598)] [Medline: [31645076](https://pubmed.ncbi.nlm.nih.gov/31645076/)]
38. Waller G, Newbury-Birch D, Simpson D, et al. The barriers and facilitators to the reporting and recording of self-harm in young people aged 18 and under: a systematic review. *BMC Public Health*. Jan 24, 2023;23(1):158. [doi: [10.1186/s12889-023-15046-7](https://doi.org/10.1186/s12889-023-15046-7)] [Medline: [36694149](https://pubmed.ncbi.nlm.nih.gov/36694149/)]
39. Dizon MP, Chow A, Ong MK, et al. Lower comorbidity scores and severity levels in Veterans Health Administration hospitals: a cross-sectional study. *BMC Health Serv Res*. May 8, 2024;24(1):601. [doi: [10.1186/s12913-024-11063-3](https://doi.org/10.1186/s12913-024-11063-3)] [Medline: [38714970](https://pubmed.ncbi.nlm.nih.gov/38714970/)]
40. Gujral K, Bahraini N, Brenner LA, et al. VA's implementation of universal screening and evaluation for the suicide risk identification program in November 2020 -Implications for Veterans with prior mental health needs. *PLoS ONE*. 2023;18(4):e0283633. [doi: [10.1371/journal.pone.0283633](https://doi.org/10.1371/journal.pone.0283633)] [Medline: [37040367](https://pubmed.ncbi.nlm.nih.gov/37040367/)]

41. Meda N, Angelozzi L, Poletto M, et al. How many people die by suicide each year? Not 727,000: a systematic review and meta-analysis of suicide underreporting across 71 countries over 122 years. *Front Psychiatry*. 2025;16:1609580. [doi: [10.3389/fpsyt.2025.1609580](https://doi.org/10.3389/fpsyt.2025.1609580)] [Medline: [40873674](https://pubmed.ncbi.nlm.nih.gov/40873674/)]
42. Getzen E, Ungar L, Mowery D, Jiang X, Long Q. Mining for equitable health: assessing the impact of missing data in electronic health records. *J Biomed Inform*. Mar 2023;139:104269. [doi: [10.1016/j.jbi.2022.104269](https://doi.org/10.1016/j.jbi.2022.104269)] [Medline: [36621750](https://pubmed.ncbi.nlm.nih.gov/36621750/)]
43. McKenzie J, Rajapakshe R, Shen H, Rajapakshe S, Lin A. A semiautomated chart review for assessing the development of radiation pneumonitis using natural language processing: diagnostic accuracy and feasibility study. *JMIR Med Inform*. Nov 12, 2021;9(11):e29241. [doi: [10.2196/29241](https://doi.org/10.2196/29241)] [Medline: [34766919](https://pubmed.ncbi.nlm.nih.gov/34766919/)]
44. Xie F, Ling Grant DS, Chang J, Amundsen BI, Hechter RC. Identifying suicidal ideation and attempt from clinical notes within a large integrated health care system. *Perm J*. Apr 5, 2022;26(1):85-93. [doi: [10.7812/TPP/21.102](https://doi.org/10.7812/TPP/21.102)] [Medline: [35609162](https://pubmed.ncbi.nlm.nih.gov/35609162/)]
45. Teismann T, Eimen JJ, Cwik JC. Misclassification of self-directed violence. *Crisis*. Nov 2023;44(6):525-528. [doi: [10.1027/0227-5910/a000897](https://doi.org/10.1027/0227-5910/a000897)] [Medline: [36636794](https://pubmed.ncbi.nlm.nih.gov/36636794/)]
46. Simon GE, Shortreed SM, Boggs JM, et al. Accuracy of ICD-10-CM encounter diagnoses from health records for identifying self-harm events. *J Am Med Inform Assoc*. Nov 14, 2022;29(12):2023-2031. [doi: [10.1093/jamia/ocac144](https://doi.org/10.1093/jamia/ocac144)] [Medline: [36018725](https://pubmed.ncbi.nlm.nih.gov/36018725/)]
47. Arndt BG, Micek MA, Rule A, Shafer CM, Baltus JJ, Sinsky CA. More tethered to the EHR: EHR workload trends among academic primary care physicians, 2019-2023. *Ann Fam Med*. 2024;22(1):12-18. [doi: [10.1370/afm.3047](https://doi.org/10.1370/afm.3047)] [Medline: [38253499](https://pubmed.ncbi.nlm.nih.gov/38253499/)]
48. Davis M, Siegel J, Becker-Haimes EM, et al. Identifying common and unique barriers and facilitators to implementing evidence-based practices for suicide prevention across primary care and specialty mental health settings. *Arch Suicide Res*. 2023;27(2):192-214. [doi: [10.1080/13811118.2021.1982094](https://doi.org/10.1080/13811118.2021.1982094)] [Medline: [34651544](https://pubmed.ncbi.nlm.nih.gov/34651544/)]
49. LeCloux M, Aguinaldo LD, Lanzillo EC, Horowitz LM. PCP opinions of universal suicide risk screening in rural primary care: current challenges and strategies for successful implementation. *J Rural Health*. Jun 2021;37(3):554-564. [doi: [10.1111/jrh.12508](https://doi.org/10.1111/jrh.12508)] [Medline: [32845543](https://pubmed.ncbi.nlm.nih.gov/32845543/)]
50. Halverson TF, Mann AJD, Zerkow RL, et al. Nonsuicidal self-injury in veterans: prevalence, clinical characteristics, and gender differences from a national cohort. *Psychiatry Res*. Sep 2022;315:114708. [doi: [10.1016/j.psychres.2022.114708](https://doi.org/10.1016/j.psychres.2022.114708)] [Medline: [35868073](https://pubmed.ncbi.nlm.nih.gov/35868073/)]
51. Gromatsky M, Halverson TF, Dillon KH, et al. The prevalence of nonsuicidal self-injury in military personnel: a systematic review and meta-analysis. *Trauma Violence Abuse*. Dec 2023;24(5):2936-2952. [doi: [10.1177/15248380221119513](https://doi.org/10.1177/15248380221119513)] [Medline: [36062896](https://pubmed.ncbi.nlm.nih.gov/36062896/)]
52. Bryan C, Bryan A. Nonsuicidal self-injury among a sample of United States military personnel and veterans enrolled in college classes. *J Clin Psychol*. Sep 2014;70(9):874-885. [doi: [10.1002/jclp.22075](https://doi.org/10.1002/jclp.22075)] [Medline: [24619940](https://pubmed.ncbi.nlm.nih.gov/24619940/)]
53. Kimbrel NA, Thomas SP, Hicks TA, et al. Wall/object punching: an important but under-recognized form of nonsuicidal self-injury. *Suicide Life Threat Behav*. Oct 2018;48(5):501-511. [doi: [10.1111/sltb.12371](https://doi.org/10.1111/sltb.12371)] [Medline: [28925016](https://pubmed.ncbi.nlm.nih.gov/28925016/)]
54. Patel TA, Dillon KH, Cassiello-Robbins C, Calhoun PS, Beckham JC, Kimbrel NA. Anger, impulsivity and wall/object punching in a sample of U.S. veterans with psychiatric disorders. *J Psychiatr Res*. Mar 2022;147:269-273. [doi: [10.1016/j.jpsychires.2022.01.036](https://doi.org/10.1016/j.jpsychires.2022.01.036)] [Medline: [35074743](https://pubmed.ncbi.nlm.nih.gov/35074743/)]
55. Nguyen CT, Richman JS, Chiles JW, Singh JA, McDonald MLN. Osteoarthritis case identification in the Million Veteran Program cohort: comparison of diagnostic codes versus clinical notes. *Osteoarthritis Cartilage*. Jun 2026;8(2):100758. [doi: [10.1016/j.ocarto.2026.100758](https://doi.org/10.1016/j.ocarto.2026.100758)] [Medline: [41799924](https://pubmed.ncbi.nlm.nih.gov/41799924/)]
56. Moldwin A, Demner-Fushman D, Goodwin TR. Empirical findings on the role of structured data, unstructured data, and their combination for automatic clinical phenotyping. *AMIA Jt Summits Transl Sci Proc*. 2021;2021:445-454. [Medline: [34457160](https://pubmed.ncbi.nlm.nih.gov/34457160/)]
57. Blanchard G, Lee G, Scott C. Semi-supervised novelty detection. *J Mach Learn Res*. Dec 1, 2010;11:2973-3009. [doi: [10.5555/1756006.1953028](https://doi.org/10.5555/1756006.1953028)]
58. Zou R, D Williamson B, M Shortreed S, Coley RY. Validation of a risk-prediction model in the presence of outcome misclassification. *Stat Med*. Apr 2026;45(8-9):e70377. [doi: [10.1002/sim.70377](https://doi.org/10.1002/sim.70377)]
59. Felleman BI, Doran NM, Asamsama OH, Oliva EM, Han BH. New methodology to improve tracking of veteran overdose deaths and characterization of a population of veteran overdose decedents in San Diego County. *Drug Alcohol Depend Rep*. Dec 2025;17:100392. [doi: [10.1016/j.dadr.2025.100392](https://doi.org/10.1016/j.dadr.2025.100392)] [Medline: [41322677](https://pubmed.ncbi.nlm.nih.gov/41322677/)]
60. Brenner LA, Breshears RE, Betthausen LM, et al. Implementation of a suicide nomenclature within two VA healthcare settings. *J Clin Psychol Med Settings*. Jun 2011;18(2):116-128. [doi: [10.1007/s10880-011-9240-9](https://doi.org/10.1007/s10880-011-9240-9)] [Medline: [21626353](https://pubmed.ncbi.nlm.nih.gov/21626353/)]

61. Inadequate staff training and lack of oversight contribute to the Veterans Health Administration's suicide risk screening and evaluation deficiencies. United States Department of Veterans Affairs; Dec 18, 2024. URL: <https://www.vaog.gov/sites/default/files/reports/2024-12/vaog-23-02939-13.pdf> [Accessed 2026-05-11]

Abbreviations

AUC: area under the curve
CSR: compressed sparse row
ED: emergency department
EHR: electronic health record
ICD: *International Classification of Diseases*
LASSO: least absolute shrinkage and selection operator
ML: machine learning
NLP: natural language processing
NSSI: nonsuicidal self-injury
OMOP: Observational Medical Outcomes Partnership
PU: positive and unlabeled
PULSCAR: positive unlabeled learning selected completely at random
PULSNAR: positive unlabeled learning selected not at random
SBOR: Suicide Behavior and Overdose Report
SCAR: selected completely at random
SNAR: selected not at random
SNOMED: Systematized Nomenclature of Medicine
VHA: Veterans Health Administration
XGBoost: Extreme Gradient Boosting

Edited by Stefano Brini; peer-reviewed by Andy Tai, Stuart Nelson; submitted 05.Dec.2025; final revised version received 17.Apr.2026; accepted 19.Apr.2026; published 04.Jun.2026

Please cite as:

Kumar P, Vizsolay AD, Upadhayaya R, Moomtaheen F, Greer DR, Bologa CG, Schneider KA, Davis SE, Matheny ME, van der Goes D, Villarreal G, Zhu Y, Tohen M, Malec SA, Yang JJ, Fielstein EM, Lambert CG

Detecting Uncoded Self-Harm in Veterans' Electronic Health Records Using Positive and Unlabeled Learning: Retrospective Cohort Study

J Med Internet Res 2026;28:e89071

URL: <https://www.jmir.org/2026/1/e89071>

doi: [10.2196/89071](https://doi.org/10.2196/89071)

© Praveen Kumar, Alexandria D Vizsolay, Rajesh Upadhayaya, Fariha Moomtaheen, Donald R Greer, Cristian G Bologa, Kristan A Schneider, Sharon E Davis, Michael E Matheny, David van der Goes, Gerardo Villarreal, Yiliang Zhu, Mauricio Tohen, Scott A Malec, Jeremy J Yang, Elliot M Fielstein, Christophe Gerard Lambert. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 04.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.