

Original Paper

# Natural Language Processing of Clinical Notes to Identify Mental Illness and Substance Use Among People Living with HIV: Retrospective Cohort Study

Jessica P Ridgway<sup>1</sup>, MD, MSc; Arno Uvin<sup>1</sup>, BSc; Jessica Schmitt<sup>1</sup>, LCSW; Tomasz Oliwa<sup>2</sup>, PhD; Ellen Almirol<sup>1</sup>, MSc; Samantha Devlin<sup>1</sup>, MSc; John Schneider<sup>1</sup>, MD, MPH

<sup>1</sup>Department of Medicine, University of Chicago, Chicago, IL, United States

<sup>2</sup>Center for Research Informatics, University of Chicago, Chicago, IL, United States

**Corresponding Author:**

Jessica P Ridgway, MD, MSc

Department of Medicine

University of Chicago

5841 S Maryland Ave

MC 5065

Chicago, IL, 60637

United States

Phone: 1 7737029185

Email: [jessica.ridgway@uchospitals.edu](mailto:jessica.ridgway@uchospitals.edu)

## Abstract

**Background:** Mental illness and substance use are prevalent among people living with HIV and often lead to poor health outcomes. Electronic medical record (EMR) data are increasingly being utilized for HIV-related clinical research and care, but mental illness and substance use are often underdocumented in structured EMR fields. Natural language processing (NLP) of unstructured text of clinical notes in the EMR may more accurately identify mental illness and substance use among people living with HIV than structured EMR fields alone.

**Objective:** The aim of this study was to utilize NLP of clinical notes to detect mental illness and substance use among people living with HIV and to determine how often these factors are documented in structured EMR fields.

**Methods:** We collected both structured EMR data (diagnosis codes, social history, Problem List) as well as the unstructured text of clinical HIV care notes for adults living with HIV. We developed NLP algorithms to identify words and phrases associated with mental illness and substance use in the clinical notes. The algorithms were validated based on chart review. We compared numbers of patients with documentation of mental illness or substance use identified by structured EMR fields with those identified by the NLP algorithms.

**Results:** The NLP algorithm for detecting mental illness had a positive predictive value (PPV) of 98% and a negative predictive value (NPV) of 98%. The NLP algorithm for detecting substance use had a PPV of 92% and an NPV of 98%. The NLP algorithm for mental illness identified 54.0% (420/778) of patients as having documentation of mental illness in the text of clinical notes. Among the patients with mental illness detected by NLP, 58.6% (246/420) had documentation of mental illness in at least one structured EMR field. Sixty-three patients had documentation of mental illness in structured EMR fields that was not detected by NLP of clinical notes. The NLP algorithm for substance use detected substance use in the text of clinical notes in 18.1% (141/778) of patients. Among patients with substance use detected by NLP, 73.8% (104/141) had documentation of substance use in at least one structured EMR field. Seventy-six patients had documentation of substance use in structured EMR fields that was not detected by NLP of clinical notes.

**Conclusions:** Among patients in an urban HIV care clinic, NLP of clinical notes identified high rates of mental illness and substance use that were often not documented in structured EMR fields. This finding has important implications for epidemiologic research and clinical care for people living with HIV.

(*JMIR Med Inform* 2021;9(3):e23456) doi: [10.2196/23456](https://doi.org/10.2196/23456)

**KEYWORDS**

natural language processing; HIV; substance use; mental illness; electronic medical records

## Introduction

Behavioral health disorders are highly prevalent among people living with HIV [1,2], who have a 2 to 4-fold higher risk of depression than the general population, with prevalence rates ranging from 24% to 63% [3-9]. A recent study among over 10,000 people living with HIV at seven HIV care sites across the United States found the prevalence of substance use disorder to be 48%, with 20% of patients having polysubstance use disorder [10]. This is higher than the rate of the general US population, in which the prevalence of substance use disorder is 7.7% [11].

In addition to being common among people living with HIV, mental illness and substance use often lead to poor health outcomes for this population. People living with HIV who have mental illness and substance use disorder have lower rates of engagement in HIV care and are less likely to adhere to antiretroviral therapy than those without behavioral health disorders [12-18]. Depression has been independently associated with mortality among several large cohorts of people living with HIV [12,19-21]. Besides poor individual health outcomes, people living with HIV with mental illness or substance use disorder are more likely to transmit HIV to others, because behavioral health disorders are associated with elevated HIV viral loads and behaviors that increase the risk of HIV transmission [22-24]. Many people living with HIV have co-occurring mental health disorders and substance use disorders, further exacerbating these adverse health outcomes [5,14].

To improve understanding of mental illness and substance use among people living with HIV, electronic medical record (EMR)-based behavioral health data are increasingly being utilized in HIV-related clinical research and medical care [25-27]. For example, Tolson et al [25] used an electronic reporting tool within the EMR to identify people living with HIV with substance use disorders to determine the association of substance use with hospitalization and virologic suppression. Other researchers used electronic billing codes to identify risk factors for suicidal ideation among people living with HIV [27]. However, mental illness and substance use are often underdocumented in structured EMR fields (eg, diagnosis codes, Problem List) [26,28,29], potentially leading to the exclusion of people living with HIV with behavioral health disorders from these studies if only discrete EMR data are used.

Natural language processing (NLP) of unstructured text of clinical notes in the EMR may identify behavioral health disorders beyond those identified using structured EMR fields alone [30,31]. Afshar et al [31] used NLP of clinical notes to identify patients with alcohol misuse, demonstrating greater

accuracy than EMR-based billing codes; however, this study was performed among hospitalized trauma patients rather than with outpatients living with HIV. Oliwa et al [32] used NLP of clinical notes to identify phrases associated with improved engagement in HIV care. Their study identified NLP phrases related to substance use and mental health among people living with HIV, but they did not compare their findings with documentation in structured EMR fields.

To fill these gaps, the aim of this study was to utilize NLP of clinical notes to detect mental illness and substance use among people living with HIV, and to determine how often these factors were documented in structured EMR fields.

## Methods

We performed a retrospective cohort study among people living with HIV at the University of Chicago Medicine (UCM) in Chicago, Illinois. Participants were included in the study if they were HIV-positive, 18 years of age or older, and attended at least one outpatient HIV care encounter at UCM between May 1, 2011 and May 30, 2016. This study was approved by the University of Chicago Institutional Review Board.

For eligible participants, we collected both structured EMR data as well as the unstructured text of clinical HIV care notes during the study time period. Structured EMR data collected included demographics, diagnosis codes (International Classification of Disease [ICD]-9 and ICD-10), Problem List (a list of physician-assigned diagnoses in the EMR), and social history. Unstructured data included the text of notes written by physicians, advanced practice providers, nurses, and social workers in the Department of Infectious Diseases. Data were extracted from the University of Chicago Clinical Research Data Warehouse, which stores data from the EMR (EPIC, Verona, WI) as well as administrative databases.

To develop the NLP algorithms for detecting mental illness and substance use, subject matter experts (physicians at the Department of Infectious Diseases and HIV care social workers) defined potential indicative words and crafted regular expressions to search for these key words and phrases related to mental illness and substance use (see [Textbox 1](#)). NegEx with augmented negation terms was applied to the key words and phrases found in clinical notes [33]. Those that were identified as negated occurrences by NegEx were excluded for the subsequent NLP steps. The Lucene Porter stemmer was used as a stemming algorithm to provide matching generalization between the tokens and the words/phrases from [Textbox 1](#) [34]. Stanford CoreNLP with additional domain-specific split patterns was employed as a tokenizer and sentence splitter to provide the NegEx input sentences [35].

**Textbox 1.** Words and phrases detected by natural language processing algorithms.

- Words/phrases for mental illness  
Depression, Depressed, Bipolar, Anxiety, Panic, Psychiatry, Schizophrenia, Bipolar, Psychosis, Care2Prevent (mental health program), Anxious, Therapist (excluding physical therapist), Behavioral health, C2P, Psychotic  
*Note: Stemmed forms, regular expression word boundaries, and a negative lookbehind in the case of “therapist” are excluded from this list for readability purposes.*
- Words/phrases for substance abuse  
IVDU (intravenous drug user), Cocaine, Heroin, Crack, Alcohol abuse, AA (Alcoholics Anonymous) meeting, Haymarket (drug treatment program), NA (Narcotics Anonymous) meeting, Drug treatment program

**Textboxes 2-4** list the diagnosis codes and Problem List phrases used to identify mental illness and substance use. Structured data from the Social History EMR section was considered to identify substance use if there was documentation of any illegal drug use (with the exception of marijuana) or if there was specific documentation of abuse of substances, including both legal and illegal substances.

To validate the NLP algorithm for mental illness, a random sample of 100 clinical notes flagged as positive for mental illness and 100 clinical notes not flagged for mental illness were manually reviewed to determine if the note documented that the patient had a mental illness. Two reviewers examined each note, and any discrepancies were resolved based on discussion and mutual agreement between reviewers. Using the determination from the manual chart review as the gold standard, we calculated the positive predictive value (PPV) of the

algorithm (ie, the number of notes in which mental illness was present based on chart review divided by the number of reviewed notes that were flagged as positive for mental illness). We also calculated the negative predictive value (NPV) of the algorithm (ie, the number of notes in which mental illness was not present based on chart review divided by the number of reviewed notes not flagged by the mental illness algorithm). Similarly, to validate the NLP algorithm for substance use, a random sample of 100 clinical notes in which the algorithm detected substance use and 100 clinical notes where substance use was not detected were manually reviewed. Subsequently, the PPV and NPV for the substance use algorithm were also calculated.

We compared numbers of patients with mental illness or substance use identified by structured EMR fields with those identified by the NLP algorithms.

**Textbox 2.** International Classification of Diseases (ICD) diagnosis codes used to identify mental illness.

- ICD-9 codes  
291.9, 293.81, 293.82, 293.83, 293.84, 294.9, 295.3, 295.31, 295.32, 295.33, 295.34, 295.35, 295.42, 295.44, 295.6, 295.7, 295.71, 295.72, 295.75, 295.8, 295.9, 295.92, 296, 296.01, 296.02, 296.1, 296.15, 296.2, 296.21, 296.22, 296.23, 296.24, 296.25, 296.26, 296.3, 296.31, 296.32, 296.33, 296.34, 296.35, 296.36, 296.4, 296.41, 296.42, 296.44, 296.5, 296.51, 296.52, 296.53, 296.54, 296.55, 296.6, 296.64, 296.7, 296.8, 296.9, 297.1, 297.9, 298.9, 300, 300.01, 300.21, 300.3, 300.4, 300.81, 301.7, 301.82, 301.83, 301.9, 309, 309.24, 309.28, 309.3, 309.4, 309.81, 310.8, 311, 312.81, 312.82, 313.81, 314, 314.01, 648.41, 648.44, E950.0, E950.2, E950.3, E950.4, E950.9, E953.0, V11.0, V40.0, V40.9
- ICD-10 codes  
F0630, F09, F203, F2089, F209, F23, F250, F251, F259, F3110, F3111, F312, F3130, F3132, F3170, F3181, F319, F321, F323, F329, F330, F331, F332, F333, F3340, F3341, F3342, F339, F4001, F4010, F411, F419, F4323, F458, F509, F603, F900, F901, F913, F919, Z915

**Textbox 3.** International Classification of Diseases (ICD) diagnosis codes used to identify substance use.

- ICD-9 codes  
291, 291.2, 291.3, 291.81, 291.9, 304, 304.01, 304.02, 304.2, 304.22, 304.23, 304.3, 304.31, 304.7, 304.71, 304.72, 304.8, 304.83, 305, 305.01, 305.02, 305.03, 305.2, 305.21, 305.22, 305.23, 305.4, 305.5, 305.51, 305.52, 305.53, 305.6, 305.61, 305.62, 305.63, 305.7, 305.91, 305.93, 425.5, 535.3, 571, 571.2, 648.33, 965.01, 970.81, E850.0, E850.1, E850.2, E854.8, E860.0, E860.9, E935.0
- ICD-10 codes  
F1010, F10120, F10129, F10188, F1020, F1021, F10239, F1029, F1099, F1110, F1120, F1121, F1123, F1190, F11959, F1210, F1220, F1290, F12929, F1410, F14129, F1414, F14188, F1420, F1421, F14259, F1490, F14929, F1494, F1510, F1520, I426, K7031, K852, K860, O99313

**Textbox 4.** Problem List phrases.

- Mental illness
 

ADD (attention deficit disorder); ADHD (attention deficit hyperactivity disorder); ADHD (attention deficit hyperactivity disorder), inattentive type; ADHD, predominantly inattentive type; Adjustment disorder with depressed mood; Adjustment disorder with mixed anxiety and depressed mood; Agoraphobia with panic disorder; Anxiety; Anxiety and depression; Anxiety disorder; Anxiety disorder in conditions classified elsewhere; Anxiety state, unspecified; Anxiety, mild; Attention deficit disorder without mention of hyperactivity; Bipolar 1 disorder; Bipolar 2 disorder; Bipolar affective; Bipolar affective disorder; Bipolar affective disorder, currently depressed, moderate; Bipolar depression; Bipolar disorder; Bipolar disorder, currently in remission of unspecified degree, most recent episode type unspecified; Bipolar disorder, unspecified; Bipolar I disorder, most recent episode depressed; Bipolar I disorder, most recent episode depressed, severe with psychosis; Bipolar I disorder, most recent episode manic; Bipolar I disorder, most recent episode manic, mild; Borderline personality disorder; Bulimia nervosa; Depressed mood; Depression; Depression (disease); Depression with anxiety; Depression, major; Depression, major, recurrent, severe with psychosis; Depression, recurrent; Depressive disorder; Depressive disorder, not elsewhere classified; Depressive episode; H/O attempted suicide; History of depression; Major depression; Major depression, recurrent; Major depression, recurrent, chronic; Major depressive disorder; Major depressive disorder, recurrent episode, in full remission; Major depressive disorder, recurrent episode, in partial or unspecified remission; Major depressive disorder, recurrent episode, mild; Major depressive disorder, recurrent episode, moderate; Major depressive disorder, recurrent episode, severe, without mention of psychotic behavior; Major depressive disorder, recurrent episode, unspecified; Major depressive disorder, severe; Major depressive disorder, single episode in full remission; Major depressive disorder, single episode, moderate; Manic depression; MDD (major depressive disorder), recurrent episode; MDD, recurrent episode, moderate; Mechanical complication of other vascular device, implant, and graft; Mood disorder; Mood disorder due to known physiological condition; Mood disorder in conditions classified elsewhere; Panic attack; Panic attacks; Panic disorder without agoraphobia; Paranoia (psychosis); Paranoid schizophrenia, chronic condition; Paranoid schizophrenia, chronic condition with acute exacerbation; Paranoid schizophrenia, unspecified condition; Postpartum depression; Posttraumatic stress; Posttraumatic stress disorder; Psychiatric illness; Psychiatric pseudoseizure; Psychosis; Psychosis, organic; Psychotic disorder with delusions in conditions classified elsewhere; PTSD (posttraumatic stress disorder); Schizoaffective disorder; Schizoaffective disorder, unspecified condition; Schizophrenia; Schizophrenia, disorganized, chronic with acute exacerbation; Schizophrenia, paranoid type; Schizophrenia, unspecified type; Suicidal ideation; Suicide attempt by drug ingestion; Suicide ideation; Unspecified schizophrenia, unspecified condition
- Substance use disorders
 

Addiction, marijuana; Alcohol abuse; Alcohol abuse, continuous drinking behavior; Alcohol abuse, daily use; Alcohol abuse, episodic; Alcohol dependence; Alcohol dependence in remission; Alcohol dependence with acute alcoholic intoxication; Alcohol use; Alcohol withdrawal; Alcoholic cirrhosis of liver; Alcoholism with alcohol dependence; Cannabis use disorder, mild, abuse; Cocaine abuse; Cocaine abuse, in remission; Cocaine addiction; Cocaine dependence, continuous; Cocaine substance abuse; Cocaine use; Cocaine withdrawal; Dementia associated with alcoholism; ETOH abuse; Excessive blood level of alcohol; H/O alcohol abuse; H/O drug abuse; H/O substance abuse; Habitual alcohol use; History of alcohol abuse; History of alcohol use; History of cocaine abuse; History of cocaine use; History of drug abuse; History of heroin abuse; History of opioid abuse; Hx of cocaine abuse; IV (intravenous) drug abuse; IVDU (intravenous drug user); Marijuana abuse; Methadone dependence; Methadone use; Methamphetamine abuse; Opioid abuse, unspecified; Pancreatitis, alcoholic, acute; Polysubstance abuse; Psychoactive substance-induced organic mood disorder

## Results

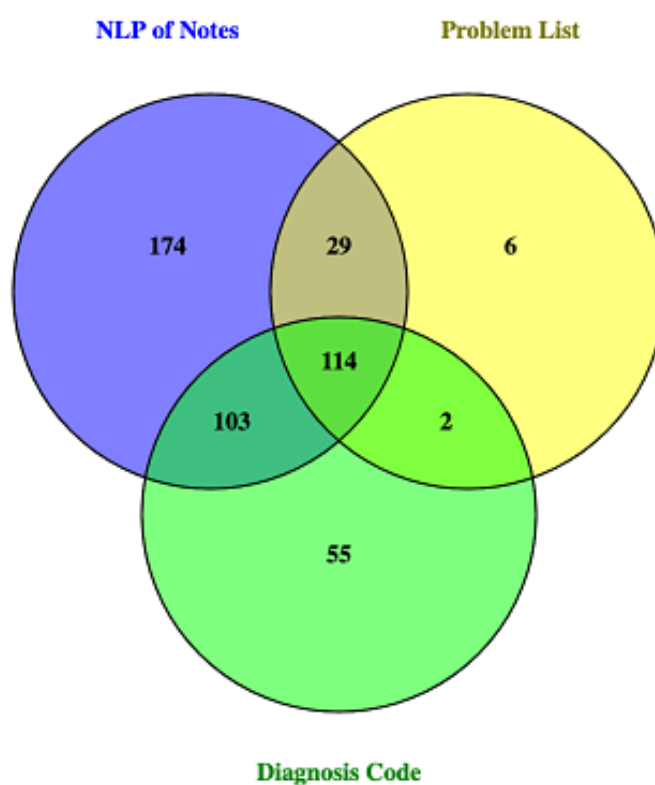
During the study period, 778 people living with HIV attended at least one HIV care appointment (Table 1). A total of 13,905 clinical notes were included, with a mean of 13 notes per patient (range 1-109). Based on manual review of clinical notes as described above, the NLP algorithm for detecting mental illness had a PPV of 98% and an NPV of 98%. The NLP algorithm for detecting substance use had a PPV of 92% and an NPV of 98%.

The NLP algorithm for mental illness identified 54.0% (420/778) of patients as having documentation of mental illness in the text

of clinical notes (Figure 1). With the PPV of the algorithm of 98%, this would suggest that 412 patients truly had mental illness. Among the patients with mental illness detected by NLP, 58.6% (246/420) had documentation of mental illness in at least one structured EMR field (ie, Problem List or diagnosis code), including 34.0% (143/420) with a mental illness listed in the Problem List and 51.7% (217/420) with a diagnosis code related to mental illness. Sixty-three patients had documentation of mental illness in structured EMR fields that was not detected by NLP of clinical notes.

**Table 1.** Demographic characteristics of participants (N=778).

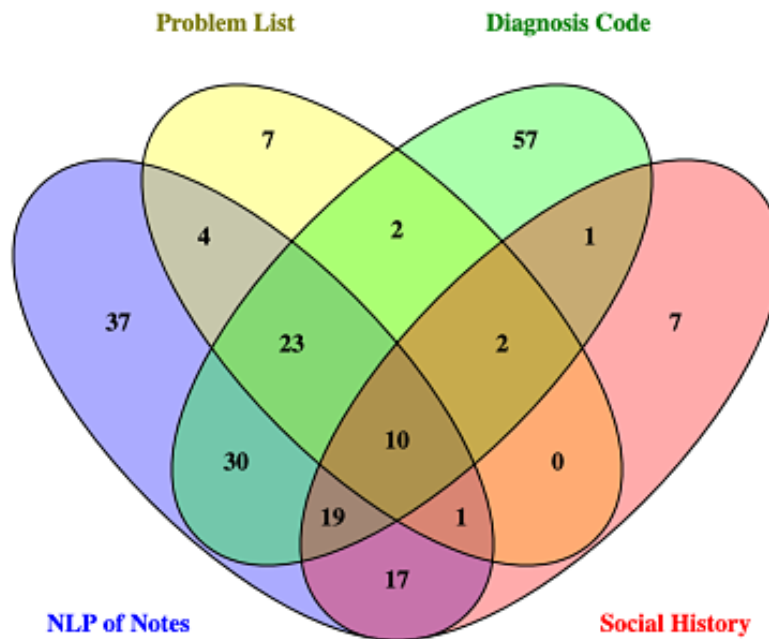
Characteristic	Value
Age (years), mean (SD)	43.1 (13.5)
Female, n (%)	287 (36.9)
<b>Race/ethnicity, n (%)</b>	
Black	620 (79.7)
White	107 (13.8)
Latinx	27 (3.5)
Asian	8 (1.0)
Other	16 (2.1)
<b>Insurance, n (%)</b>	
Medicaid	272 (35.0)
Medicare	228 (29.3)
Private	257 (33.0)
Other/self-pay	21 (2.7)

**Figure 1.** Electronic medical record documentation of mental illness among people living with HIV. NLP: natural language processing.

The NLP algorithm for substance use detected substance use in the text of clinical notes in 18.1% (141/778) of participants (Figure 2). Based on the PPV of the algorithm of 92%, it is likely that 130 patients truly had substance use. Among patients with substance use detected by NLP, 73.8% (104/141) had documentation of substance use in at least one structured EMR field, including 27.0% (38/141) with documentation of substance

use in the Problem List, 58.2% (82/141) with a diagnosis code related to substance use, and 33.3% (47/141) with substance use documented in the Social History section of the EMR. Seventy-six patients had documentation of substance use in structured EMR fields that was not detected by NLP of clinical notes.

**Figure 2.** Electronic medical record documentation of substance use among people living with HIV. NLP: natural language processing.



## Discussion

Among patients in an urban HIV care clinic, NLP of clinical notes identified high rates of mental illness and substance use that were often not documented in structured EMR fields. This finding has important implications for clinical care and epidemiologic research among people living with HIV. Namely, relying on structured EMR fields alone to identify people living with HIV with behavioral health disorders may miss a substantial number of patients. Given the high PPV of our algorithms, addition of such NLP algorithms to current tools for identifying behavioral health disorders could augment detection of these disorders among people living with HIV.

To our knowledge, this is the first study to utilize NLP of EMR notes to detect mental illness and substance use among people living with HIV. Other studies have used NLP to detect depression and substance misuse in non-HIV care settings [30,31,36]. Adekanattu et al [36] used NLP to identify depression from EMR notes among patients prescribed antidepressants, and found that 31% of patients with depression detected by NLP were missing a diagnosis code for depression. Zhou et al [37] similarly used NLP of hospital discharge summaries to identify depression among hospitalized patients, and found that 20% of patients with depression detected by NLP did not have a depression diagnosis code. These rates of discordant documentation are lower than that obtained in this study, in which nearly half of patients with mental illness detected by NLP did not have a diagnosis code for mental illness. This discrepancy may be explained by differences in the patient populations studied. Our patients are from a general

HIV clinic, rather than inpatients or outpatients already prescribed antidepressants, populations in which medical providers may be more likely to enter a diagnosis code for mental illness.

Our NLP algorithm for mental illness identified 54% of people living with HIV in our clinical population as having mental illness. This is similar to other studies among people living with HIV, which have shown prevalence rates as high as 63% based on validated depression screening tools (eg, Patient Health Questionnaire-9) [3-9]. The NLP algorithm detected substance use in 18% of our clinical population. This rate is within the lower end of what has previously been reported. Hartzler et al [10] found that the prevalence of substance use disorders among people living with HIV at 7 HIV care sites ranged from 21% to 71% based on substance use disorder screening tools. Of note, for both mental illness and substance use, the NLP algorithms failed to flag a substantial number of patients who had mental illness or substance use documented in structured EMR fields, suggesting that NLP algorithms should be used in combination with structured fields rather than as a replacement for structured fields for detecting these characteristics.

As EMR data are increasingly being used for clinical care and research among people living with HIV, extracting accurate behavioral health data from the EMR is essential. EMRs have been used to provide electronic feedback to providers to alert them that patients may have untreated depression [38,39]. Results from NLP of clinical notes could potentially augment such electronic alerts. Recent studies have used structured EMR fields, including documentation of substance use and mental illness, to create predictive models of HIV appointment

adherence [12,40]. However, if mental illness and substance use are not adequately documented in structured EMR fields, inclusion of NLP of clinician notes may improve such predictive models by identifying additional risk factors for appointment nonadherence.

Our study has several limitations. We did not review all clinical notes for the presence or absence of behavioral health disorder documentation, and some of the NLP-detected cases may be false positives. Although we adjusted for negation in the text, we may have falsely detected mental illness in some instances where providers wrote in a nonstandard format that patients did *not* have mental illness or where they documented that a family member and not the patient themselves had a behavioral health disorder. In addition, certain phrases (eg, Alcoholics Anonymous meeting) may have detected patients with past substance use disorder rather than active substance use disorder. However, in the review of a random sample of 400 notes, we found a high PPV for the NLP algorithms. The NLP algorithms may have also failed to flag notes that documented behavioral health disorders (ie, false negatives). Moreover, the NLP algorithms

do not necessarily detect patients with mental illness or substance use, but only detect documentation in the clinical notes of mental illness or substance use. If providers did not ask patients about these topics or did not document regarding their conversations, then people living with HIV with behavioral health disorders may have been missed by our algorithms. Inclusion of validated behavioral health screening tools within the EMR would likely improve detection of mental illness and substance use. These screening tools were not routinely in place in our clinic at the time of the study, and therefore we were unable to assess how they would have affected the results.

In conclusion, we performed the first study of NLP of unstructured clinical notes for mental illness and substance use among people living with HIV. Although these behavioral health disorders were commonly detected by NLP, they were often undocumented in structured fields of the EMR. More research is needed to understand how to best utilize both structured and unstructured EMR data for clinical and epidemiologic research among people living with HIV.

## Acknowledgments

This work was supported by National Institutes of Health (NIH; 1K23MH121190-01) and the NIH-funded Third Coast Center for AIDS Research (CFAR; P30 AI117943). Data from this study were provided by the Clinical Research Data Warehouse maintained by the Center for Research Informatics at University of Chicago. The Center for Research Informatics is funded by the Biological Sciences Division, Institute for Translational Medicine/CTSA (NIH UL1 TR000430) at the University of Chicago. The funders had no role in review or approval of the manuscript for publication.

## Authors' Contributions

JR and JAS conceived of and designed the study. TO, EA, AU, and SD collected and analyzed the data. JR, JAS, and JS interpreted the data. JR drafted the manuscript, and JAS, TO, EA, AU, SD, and JS critically revised the manuscript. JR obtained funding and supervised the study.

## Conflicts of Interest

None declared.

## References

1. Remien R, Stirratt M, Nguyen N, Robbins R, Pala A, Mellins C. Mental health and HIV/AIDS: the need for an integrated response. *AIDS* 2019 Jul 15;33(9):1411-1420 [FREE Full text] [doi: [10.1097/QAD.0000000000002227](https://doi.org/10.1097/QAD.0000000000002227)] [Medline: [30950883](https://pubmed.ncbi.nlm.nih.gov/30950883/)]
2. Rezaei S, Ahmadi S, Rahmati J, Hosseinifard H, Dehnad A, Aryankhesal A, et al. Global prevalence of depression in HIV/AIDS: a systematic review and meta-analysis. *BMJ Support Palliat Care* 2019 Dec 19;9(4):404-412. [doi: [10.1136/bmjspcare-2019-001952](https://doi.org/10.1136/bmjspcare-2019-001952)] [Medline: [31537580](https://pubmed.ncbi.nlm.nih.gov/31537580/)]
3. Bing EG, Burnam MA, Longshore D, Fleishman JA, Sherbourne CD, London AS, et al. Psychiatric disorders and drug use among human immunodeficiency virus-infected adults in the United States. *Arch Gen Psychiatry* 2001 Aug 01;58(8):721-728. [doi: [10.1001/archpsyc.58.8.721](https://doi.org/10.1001/archpsyc.58.8.721)] [Medline: [11483137](https://pubmed.ncbi.nlm.nih.gov/11483137/)]
4. Schumacher JE, McCullumsmith C, Mugavero MJ, Ingle-Pang PE, Raper JL, Willig JH, et al. Routine depression screening in an HIV clinic cohort identifies patients with complex psychiatric co-morbidities who show significant response to treatment. *AIDS Behav* 2013 Oct 20;17(8):2781-2791 [FREE Full text] [doi: [10.1007/s10461-012-0342-7](https://doi.org/10.1007/s10461-012-0342-7)] [Medline: [23086427](https://pubmed.ncbi.nlm.nih.gov/23086427/)]
5. Tegger MK, Crane HM, Tapia KA, Uldall KK, Holte SE, Kitahata MM. The effect of mental illness, substance use, and treatment for depression on the initiation of highly active antiretroviral therapy among HIV-infected individuals. *AIDS Patient Care STDs* 2008 Mar;22(3):233-243. [doi: [10.1089/apc.2007.0092](https://doi.org/10.1089/apc.2007.0092)] [Medline: [18290749](https://pubmed.ncbi.nlm.nih.gov/18290749/)]
6. Ciesla JA, Roberts JE. Meta-analysis of the relationship between HIV infection and risk for depressive disorders. *Am J Psychiatry* 2001 May;158(5):725-730. [doi: [10.1176/appi.ajp.158.5.725](https://doi.org/10.1176/appi.ajp.158.5.725)] [Medline: [11329393](https://pubmed.ncbi.nlm.nih.gov/11329393/)]
7. Chenneville T, Gabbidon K, Drake H, Rodriguez C. Comparison of the utility of the PHQ and CES-D for depression screening among youth with HIV in an integrated care setting. *J Affect Disord* 2019 May 01;250:140-144. [doi: [10.1016/j.jad.2019.03.023](https://doi.org/10.1016/j.jad.2019.03.023)] [Medline: [30852366](https://pubmed.ncbi.nlm.nih.gov/30852366/)]

8. Shacham E, Nurutdinova D, Satyanarayana V, Stamm K, Overton E. Routine screening for depression: identifying a challenge for successful HIV care. *AIDS Patient Care STDs* 2009 Nov;23(11):949-955 [FREE Full text] [doi: [10.1089/apc.2009.0064](https://doi.org/10.1089/apc.2009.0064)] [Medline: [19925308](https://pubmed.ncbi.nlm.nih.gov/19925308/)]
9. Brandt C, Zvolensky MJ, Woods SP, Gonzalez A, Safren SA, O'Cleirigh CM. Anxiety symptoms and disorders among adults living with HIV and AIDS: A critical review and integrative synthesis of the empirical literature. *Clin Psychol Rev* 2017 Feb;51:164-184 [FREE Full text] [doi: [10.1016/j.cpr.2016.11.005](https://doi.org/10.1016/j.cpr.2016.11.005)] [Medline: [27939443](https://pubmed.ncbi.nlm.nih.gov/27939443/)]
10. Hartzler B, Dombrowski JC, Crane HM, Eron JJ, Geng EH, Christopher Mathews W, et al. Prevalence and predictors of substance use disorders among HIV care enrollees in the United States. *AIDS Behav* 2017 Apr 13;21(4):1138-1148 [FREE Full text] [doi: [10.1007/s10461-016-1584-6](https://doi.org/10.1007/s10461-016-1584-6)] [Medline: [27738780](https://pubmed.ncbi.nlm.nih.gov/27738780/)]
11. McCance-Katz EF. The National Survey on Drug Use and Health: 2019. Substance Abuse and Mental Health Services Administration. 2020 Sep. URL: [https://www.samhsa.gov/data/sites/default/files/reports/rpt29392/Assistant-Secretary-nsduh2019\\_presentation/Assistant-Secretary-nsduh2019\\_presentation.pdf](https://www.samhsa.gov/data/sites/default/files/reports/rpt29392/Assistant-Secretary-nsduh2019_presentation/Assistant-Secretary-nsduh2019_presentation.pdf) [accessed 2021-01-24]
12. Pence BW, Mills JC, Bengtson AM, Gaynes BN, Breger TL, Cook RL, et al. Association of increased chronicity of depression with HIV appointment attendance, treatment failure, and mortality among HIV-infected adults in the United States. *JAMA Psychiatry* 2018 Apr 01;75(4):379-385 [FREE Full text] [doi: [10.1001/jamapsychiatry.2017.4726](https://doi.org/10.1001/jamapsychiatry.2017.4726)] [Medline: [29466531](https://pubmed.ncbi.nlm.nih.gov/29466531/)]
13. Gonzalez J, Batchelder A, Psaros C, Safren S. Depression and HIV/AIDS treatment nonadherence: a review and meta-analysis. *J Acquir Immune Defic Syndr* 2011 Oct 01;58(2):181-187 [FREE Full text] [doi: [10.1097/QAI.0b013e31822d490a](https://doi.org/10.1097/QAI.0b013e31822d490a)] [Medline: [21857529](https://pubmed.ncbi.nlm.nih.gov/21857529/)]
14. Chander G, Himelhoch S, Moore RD. Substance abuse and psychiatric disorders in HIV-positive patients: epidemiology and impact on antiretroviral therapy. *Drugs* 2006;66(6):769-789. [doi: [10.2165/00003495-200666060-00004](https://doi.org/10.2165/00003495-200666060-00004)] [Medline: [16706551](https://pubmed.ncbi.nlm.nih.gov/16706551/)]
15. Nanni MG, Caruso R, Mitchell AJ, Meggiolaro E, Grassi L. Depression in HIV infected patients: a review. *Curr Psychiatry Rep* 2015 Jan 21;17(1):530. [doi: [10.1007/s11920-014-0530-4](https://doi.org/10.1007/s11920-014-0530-4)] [Medline: [25413636](https://pubmed.ncbi.nlm.nih.gov/25413636/)]
16. Tucker JS, Burnam M, Sherbourne CD, Kung F, Gifford AL. Substance use and mental health correlates of nonadherence to antiretroviral medications in a sample of patients with human immunodeficiency virus infection. *Am J Med* 2003 May;114(7):573-580. [doi: [10.1016/s0002-9343\(03\)00093-7](https://doi.org/10.1016/s0002-9343(03)00093-7)] [Medline: [12753881](https://pubmed.ncbi.nlm.nih.gov/12753881/)]
17. Azar MM, Springer SA, Meyer JP, Altice FL. A systematic review of the impact of alcohol use disorders on HIV treatment outcomes, adherence to antiretroviral therapy and health care utilization. *Drug Alcohol Depend* 2010 Dec 01;112(3):178-193 [FREE Full text] [doi: [10.1016/j.drugalcdep.2010.06.014](https://doi.org/10.1016/j.drugalcdep.2010.06.014)] [Medline: [20705402](https://pubmed.ncbi.nlm.nih.gov/20705402/)]
18. Zuniga JA, Yoo-Jeong M, Dai T, Guo Y, Waldrop-Valverde D. The role of depression in retention in care for persons living with HIV. *AIDS Patient Care STDs* 2016 Jan;30(1):34-38 [FREE Full text] [doi: [10.1089/apc.2015.0214](https://doi.org/10.1089/apc.2015.0214)] [Medline: [26544915](https://pubmed.ncbi.nlm.nih.gov/26544915/)]
19. Ickovics JR, Hamburger ME, Vlahov D, Schoenbaum EE, Schuman P, Boland RJ, HIV Epidemiology Research Study Group. Mortality, CD4 cell count decline, and depressive symptoms among HIV-seropositive women: longitudinal analysis from the HIV Epidemiology Research Study. *JAMA* 2001 Mar 21;285(11):1466-1474. [doi: [10.1001/jama.285.11.1466](https://doi.org/10.1001/jama.285.11.1466)] [Medline: [11255423](https://pubmed.ncbi.nlm.nih.gov/11255423/)]
20. Cook JA, Grey D, Burke J, Cohen MH, Gurtman AC, Richardson JL, et al. Depressive symptoms and AIDS-related mortality among a multisite cohort of HIV-positive women. *Am J Public Health* 2004 Jul;94(7):1133-1140. [doi: [10.2105/ajph.94.7.1133](https://doi.org/10.2105/ajph.94.7.1133)] [Medline: [15226133](https://pubmed.ncbi.nlm.nih.gov/15226133/)]
21. Todd J, Cole S, Pence B, Lesko CR, Bacchetti P, Cohen MH, et al. Effects of antiretroviral therapy and depressive symptoms on all-cause mortality among HIV-infected women. *Am J Epidemiol* 2017 May 15;185(10):869-878 [FREE Full text] [doi: [10.1093/aje/kww192](https://doi.org/10.1093/aje/kww192)] [Medline: [28430844](https://pubmed.ncbi.nlm.nih.gov/28430844/)]
22. Sikkema KJ, Watt MH, Drabkin AS, Meade CS, Hansen NB, Pence BW. Mental health treatment to reduce HIV transmission risk behavior: a positive prevention model. *AIDS Behav* 2010 Apr 15;14(2):252-262 [FREE Full text] [doi: [10.1007/s10461-009-9650-y](https://doi.org/10.1007/s10461-009-9650-y)] [Medline: [20013043](https://pubmed.ncbi.nlm.nih.gov/20013043/)]
23. Hutton HE, Lyketos CG, Zenilman JM, Thompson RE, Erbeling EJ. Depression and HIV risk behaviors among patients in a sexually transmitted disease clinic. *Am J Psychiatry* 2004 May;161(5):912-914. [doi: [10.1176/appi.ajp.161.5.912](https://doi.org/10.1176/appi.ajp.161.5.912)] [Medline: [15121659](https://pubmed.ncbi.nlm.nih.gov/15121659/)]
24. Liang J, Nosova E, Reddon H, Nolan S, Socías E, Barrios R, et al. Longitudinal patterns of illicit drug use, antiretroviral therapy exposure and plasma HIV-1 RNA viral load among HIV-positive people who use illicit drugs. *AIDS* 2020 Jul 15;34(9):1389-1396. [doi: [10.1097/QAD.0000000000002551](https://doi.org/10.1097/QAD.0000000000002551)] [Medline: [32590435](https://pubmed.ncbi.nlm.nih.gov/32590435/)]
25. Tolson C, Richey LE, Zhao Y, Korte JE, Brady K, Haynes L, et al. Association of substance use with hospitalization and virologic suppression in a southern academic HIV clinic. *Am J Med Sci* 2018 Jun;355(6):553-558 [FREE Full text] [doi: [10.1016/j.amjms.2018.03.002](https://doi.org/10.1016/j.amjms.2018.03.002)] [Medline: [29891038](https://pubmed.ncbi.nlm.nih.gov/29891038/)]
26. O'Cleirigh C, Magidson JF, Skeer MR, Mayer KH, Safren SA. Prevalence of psychiatric and substance abuse symptomatology among HIV-infected gay and bisexual men in HIV primary care. *Psychosomatics* 2015 Sep;56(5):470-478 [FREE Full text] [doi: [10.1016/j.psych.2014.08.004](https://doi.org/10.1016/j.psych.2014.08.004)] [Medline: [25656425](https://pubmed.ncbi.nlm.nih.gov/25656425/)]



27. Brown LA, Majeed I, Mu W, McCann J, Durborow S, Chen S, et al. Suicide risk among persons living with HIV. *AIDS Care* 2020 Aug 03:online ahead of print. [doi: [10.1080/09540121.2020.1801982](https://doi.org/10.1080/09540121.2020.1801982)] [Medline: [32741212](https://pubmed.ncbi.nlm.nih.gov/32741212/)]
28. Machado IK, Luz PM, Lake JE, Castro R, Velasque L, Clark JL, et al. Substance use among HIV-infected patients in Rio de Janeiro, Brazil: Agreement between medical records and the ASSIST questionnaire. *Drug Alcohol Depend* 2017 Sep 01;178:115-118 [FREE Full text] [doi: [10.1016/j.drugalcdep.2017.04.033](https://doi.org/10.1016/j.drugalcdep.2017.04.033)] [Medline: [28646713](https://pubmed.ncbi.nlm.nih.gov/28646713/)]
29. Brown LA, Mu W, McCann J, Durborow S, Blank MB. Under-documentation of psychiatric diagnoses among persons living with HIV in electronic medical records. *AIDS Care* 2021 Mar 13;33(3):311-315. [doi: [10.1080/09540121.2020.1713974](https://doi.org/10.1080/09540121.2020.1713974)] [Medline: [31931621](https://pubmed.ncbi.nlm.nih.gov/31931621/)]
30. Topaz M, Murga L, Bar-Bachar O, Cato K, Collins S. Extracting alcohol and substance abuse status from clinical notes: the added value of nursing data. *Stud Health Technol Inform* 2019 Aug 21;264:1056-1060. [doi: [10.3233/SHTI190386](https://doi.org/10.3233/SHTI190386)] [Medline: [31438086](https://pubmed.ncbi.nlm.nih.gov/31438086/)]
31. Afshar M, Phillips A, Karnik N, Mueller J, To D, Gonzalez R, et al. Natural language processing and machine learning to identify alcohol misuse from the electronic health record in trauma patients: development and internal validation. *J Am Med Inform Assoc* 2019 Mar 01;26(3):254-261 [FREE Full text] [doi: [10.1093/jamia/ocy166](https://doi.org/10.1093/jamia/ocy166)] [Medline: [30602031](https://pubmed.ncbi.nlm.nih.gov/30602031/)]
32. Oliwa T, Furner B, Schmitt J, Schneider J, Ridgway J. Development of a predictive model for retention in HIV care using natural language processing of clinical notes. *J Am Med Inform Assoc* 2021 Jan 15;28(1):104-112. [doi: [10.1093/jamia/ocaa220](https://doi.org/10.1093/jamia/ocaa220)] [Medline: [33150369](https://pubmed.ncbi.nlm.nih.gov/33150369/)]
33. Chapman W, Hilert D, Velupillai S. Extending the NegEx Lexicon for Multiple Languages. 2013 Presented at: 14th World Congress on Medical & Health Informatics; 2013; Copenhagen.
34. Apache Lucene. URL: <https://lucene.apache.org/> [accessed 2021-01-26]
35. Manning C, Sureadnu M, Bauer J, Finkel J, Bethard S, McClosky D. The Stanford CoreNLP Natural Language Processing Toolkit. 2014 Presented at: 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations; 2014; Baltimore p. 55-60. [doi: [10.3115/v1/p14-5010](https://doi.org/10.3115/v1/p14-5010)]
36. Adekkanattu P, Sholle ET, DeFerio J, Pathak J, Johnson SB, Champion TR. Ascertaining Depression Severity by Extracting Patient Health Questionnaire-9 (PHQ-9) Scores from Clinical Notes. 2018 Nov Presented at: AMIA Annual Symposium; 2018; San Francisco p. 147-156.
37. Zhou L, Baughman AW, Lei VJ, Lai KH, Navathe AS, Chang F, et al. Identifying patients with depression using free-text clinical documents. *Stud Health Technol Inform* 2015;216:629-633. [Medline: [26262127](https://pubmed.ncbi.nlm.nih.gov/26262127/)]
38. Rollman BL, Hanusa BH, Gilbert T, Lowe HJ, Kapoor WN, Schulberg HC. The electronic medical record. A randomized trial of its impact on primary care physicians' initial management of major depression [corrected]. *Arch Intern Med* 2001 Jan 22;161(2):189-197. [doi: [10.1001/archinte.161.2.189](https://doi.org/10.1001/archinte.161.2.189)] [Medline: [11176732](https://pubmed.ncbi.nlm.nih.gov/11176732/)]
39. Frame A, LaMantia M, Reddy Bynagari B, Dexter P, Boustani M. Development and implementation of an electronic decision support to manage the health of a high-risk Population: the enhanced Electronic Medical Record Aging Brain Care software (eMR-ABC). *EGEMS (Wash DC)* 2013 Mar 11;1(1):1009 [FREE Full text] [doi: [10.13063/2327-9214.1009](https://doi.org/10.13063/2327-9214.1009)] [Medline: [25848560](https://pubmed.ncbi.nlm.nih.gov/25848560/)]
40. Ramachandran A, Kumar A, Koenig H, De Unanue A, Sung C, Walsh J, et al. Predictive analytics for retention in care in an urban HIV clinic. *Sci Rep* 2020 Apr 14;10(1):6421. [doi: [10.1038/s41598-020-62729-x](https://doi.org/10.1038/s41598-020-62729-x)] [Medline: [32286333](https://pubmed.ncbi.nlm.nih.gov/32286333/)]

## Abbreviations

- EMR:** electronic medical record
- ICD:** International Classification of Diseases
- NLP:** natural language processing
- NPV:** negative predictive value
- PPV:** positive predictive value
- UCM:** University of Chicago Medicine

*Edited by G Eysenbach; submitted 12.08.20; peer-reviewed by J Jain, J Walsh, M Torii; comments to author 07.12.20; revised version received 30.01.21; accepted 31.01.21; published 10.03.21*

### *Please cite as:*

Ridgway JP, Uvin A, Schmitt J, Oliwa T, Almirol E, Devlin S, Schneider J  
*Natural Language Processing of Clinical Notes to Identify Mental Illness and Substance Use Among People Living with HIV: Retrospective Cohort Study*  
*JMIR Med Inform* 2021;9(3):e23456  
URL: <https://medinform.jmir.org/2021/3/e23456>  
doi: [10.2196/23456](https://doi.org/10.2196/23456)  
PMID: [33688848](https://pubmed.ncbi.nlm.nih.gov/33688848/)

©Jessica P Ridgway, Arno Uvin, Jessica Schmitt, Tomasz Oliwa, Ellen Almirol, Samantha Devlin, John Schneider. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org>), 10.03.2021. 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 JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.