Natural Language Processing For Identification of Hospitalized People with Substance Use Disorder

Taisuke Sato, Emily D Grussing, Ruchi Patel, Jessica Ridgway, Joji Suzuki, Benjamin Sweigart, Robert Miller, Alysse G Wurcel

Submitted to: Journal of Medical Internet Research
on: June 11, 2024

Table of Contents

Natural Language Processing For Identification of Hospitalized People with Substance Use Disorder

Taisuke Sato[1] BA; Emily D Grussing[1] MD; Ruchi Patel[1] BDS; Jessica Ridgway[2] MD, MS; Joji Suzuki[3] MD; Benjamin Sweigart[1] MA; Robert Miller[1] MS; Alysse G Wurcel[1] MD, MS

[1]Tufts Medical Center Boston US
[2]University of Chicago School of Medicine Chicago US
[3]Brigham and Women's Hospital Boston US

Corresponding Author:
Alysse G Wurcel MD, MS
Tufts Medical Center
Tufts Medical Center, Tupper Building 4F, 800 Washington St
Boston
US

Abstract

Background: People Who Use Drugs (PWUD) are at heightened risk for severe injection-related infections. Current clinical practices and research mostly rely on biomarkers, medication records, ICD codes, and self-screening forms for patients to identify PWUD; the combination of these tools still often fails to identify hospitalized SUD patients, missing crucial intervention opportunities for Serious Injection Related Infections (SIRI).

Objective: This study explores using Natural Language Processing (NLP) to enhance the equitable and comprehensive identification of PWUD in electronic medical records (EMR).

Methods: We retrospectively compiled a cohort of hospitalizations that involved PWUD at Tufts Medical Center (2020-2022). Criteria for entering the cohort included ICD10 codes for SUD, positive drug toxicology, SUD treatment prescriptions, and specific NLP keywords. We conducted human review of clinical notes in Electronic Health Records (EHR) to calculate the positive and negative predictive value of two subcohorts: admissions associated with a diagnosis code of substance use disorder only (D-only) and admissions associated with NLP identification of drug use only (N-only). We also conducted a regression analysis to evaluate the impact of race, ethnicity, and Social Vulnerability Index (SVI) on the outcomes of highly documented drug use versus drug use only documented with NLP.

Results: The study identified 4548 hospitalizations with broad heterogeneity in how people entered the cohort and subcohorts. 288 hospitalizations entered the cohort through NLP presence alone. NLP demonstrated a 54% positive predictive value (PPV), outperforming biomarkers, medication records, and ICD codes in identifying hospitalizations of PWUD. Additionally, NLP significantly enhanced these methods when integrated into the identification algorithm. The study also found that people from racially and ethnically minoritized communities and lower socioeconomic patients were significantly more likely to have SUD not documented in EMRs.

Conclusions: NLP proved effective in identifying hospitalizations of PWUD, surpassing traditional methods. While further refinement is needed, NLP shows a promising capability in minimizing healthcare disparities, particularly in infectious disease care for SUD patients, highlighting a crucial step towards more equitable healthcare.

(JMIR Preprints 11/06/2024:63147)
DOI: https://doi.org/10.2196/preprints.63147

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Natural Language Processing For Identification of Hospitalized People with Substance Use Disorder

Taisuke Sato BA [1], Emily D. Grussing MD [1], Ruchi Patel BDS [1] Jessica P. Ridgway MD MS[3] Joji Suzuki MD [4] Benjamin Sweigart MA [2] Robert Miller MS [2] Alysse Wurcel MD MS [1]

[1] Tufts Medical Center
[2] Tufts Clinical and Translational Science Institute
[3] University of Chicago School of Medicine
[4] Brigham and Women's Hospital

Corresponding Author:
Name: Alysse Wurcel MD MS
Address: Tufts Medical Center, Tupper Building 4F, 800 Washington St, Boston, MA 02111
Tell: 617 636 4605
Email: alysse.wurcel@tuftsmedicine.org

Short Summary:
Hospitalization is a golden opportunity to treat drug-use related Infections. Current metrics are inadequate to identify people who use drugs. We can use AI Natural Language Processing to augment them –especially to identify people who are marginalized and provide vital care resources.

Disclosures/Conflict of interest: None

Funding: Tufts CTSI Small Grants to Advance Translational Science (S-GATS) Program

Keywords: SIRI, Natural Language Processing, Persons Who Use Drugs, Substance use disorder, Hepatitis C, HIV

Abstract:

Introduction: People Who Use Drugs (PWUD) are at heightened risk for severe injection-related infections. Current clinical practices and research mostly rely on biomarkers, medication records, ICD codes, and self-screening forms for patients to identify PWUD; the combination of these tools still often fails to identify hospitalized SUD patients, missing crucial intervention opportunities for Serious Injection Related Infections (SIRI). This study explores using Natural Language Processing (NLP) to enhance the equitable and comprehensive identification of PWUD in electronic medical records (EMR). *Methods*: We retrospectively compiled a cohort of hospitalizations that involved PWUD at Tufts Medical Center (2020-2022). Criteria for entering the cohort included ICD10 codes for SUD, positive drug toxicology, SUD treatment prescriptions, and specific NLP keywords. We conducted human review of clinical notes in Electronic Health Records (EHR) to calculate the positive and negative predictive value of two subcohorts: admissions associated with a diagnosis code of substance use disorder only (D-only) and admissions associated with NLP identification of drug use only (N-only). We also conducted a regression analysis to evaluate the impact of race, ethnicity, and Social Vulnerability Index (SVI) on the outcomes of highly documented drug use versus drug use only documented with NLP. *Results:* The study identified 4548 hospitalizations with broad heterogeneity in how people entered the cohort and subcohorts. 288 hospitalizations entered the cohort through NLP presence alone. NLP demonstrated a 54% positive predictive value (PPV), outperforming biomarkers, medication records, and ICD codes in identifying hospitalizations of PWUD. Additionally, NLP significantly enhanced these methods when integrated into the identification algorithm. The study also found that people from racially and ethnically minoritized communities and lower socioeconomic patients were significantly more likely to have SUD not documented in EMRs. *Conclusion:* NLP proved effective in identifying hospitalizations of PWUD,

surpassing traditional methods. While further refinement is needed, NLP shows a promising capability in minimizing healthcare disparities, particularly in infectious disease care for SUD patients, highlighting a crucial step towards more equitable healthcare.

Introduction:

In the absence of harm reduction tools, people who use drugs (PWUD) are at increased risk of disease, hospitalization, and death.[1-3] Gaps in the provision of guideline-concordant hospital care to PWUD care occur, especially in the care provision to PWUD who are from racially and ethnically minoritized communities.[4-6] Barriers to optimization of health for hospitalized PWUD include under-treated pain and substance use disorder which has been linked to discharges occurring before medical optimization and increased rates of re-admission and mortality.[7-9]

The science of identifying hospitalized PWUD, for epidemiological tracking, resource allocation, and outcomes assessment following implementation strategies aimed to improve outcomes is stymied by the inability to identify these patients. The “gold standard” to identify the hospitalization of a PWUD is human chart review, a highly-regulated, time-intensive process with potential consequences for breaches of confidentiality.[10,11] Administrative billing codes (also known as International Classification of Disease codes, or ICD codes) have been used for PWUD identification. Unlike several other common disease processes such as cardiovascular diseases,[12,13] ICD10 codes for PWUD hospitalization are less sensitive and specific.[14] A systematic review found that for injection drug use, ICD9/10 codes had high specificity, but the sensitivity ranged from 47%-83%.[15] In addition, comments about substance use tend to be noted in the social history section of the electronic medical record (EMR); while substance use is a critical social determinant of health, this practice makes substance use more prone to oversight as a medical issue to be addressed. Researchers sometimes use the hepatitis C virus code as a marker of drug use, although there are a substantial number of people with HCV who are either not currently using drugs or have ever used drugs.[16,17]

The barrier to identifying PWUD can potentially be addressed with natural language processing (NLP) technology. NLP leverages artificial intelligence algorithms to interpret the written text in a context-relevant manner.[18] NLP has been effectively applied to medical examiners' reports to increase the accuracy of identifying substance use disorder-related deaths,[19] identify substance use disorders in outpatients with HIV,[20] and improve preventive care for hospitalized patients with HIV.[17]

There is only a handful of published studies about applying NLP to the identification of hospitalized PWUD admitted for bloodstream infections; however, this effort was a single-center evaluation focused only on injection drug use.[21-24] Despite its innovative capacity to identify PWUD, the science of NLP methodology is nascent. The goal of this study is to evaluate the impact of Natural Language Processing on creating a cohort of hospitalized PWUD.

Methods:

Overview of Cohort Creation: The cohort of PWUD patients was pulled from hospitalizations at Tufts Medical Center, a tertiary health care center located in Boston, Massachusetts. As "drug use" is a broad term it is worth explicitly defining drug use for this study including the use of cocaine, methamphetamine, fentanyl, and heroin. These drugs are known to either be intravenously consumed, or prone to contaminations, counterfeiting, and adulterants that expose infectious risks. Alcohol and cannabis are also often classified as drugs, but for the purposes of this study, the use of this did not qualify for entry into the cohort. Notably, we are using the term "PWUD" to discuss people in the cohort, rather than another term used to describe this population, people who inject drugs (PWID). We believe that most people hospitalized as PWUD are also PWID. However, as we are unable to be completely confident, we decided to use the broader term of PWUD.

Operationalization of Cohort Creation: Analysts pulled hospital encounters from January 1, 2020, to April 1, 2022 if any of the following domains were present:

- Biomarkers (B): In line with previously reported work, positive urine toxicology for drugs or the medication for substance use disorders (Cocaine, amphetamine, methadone, suboxone, fentanyl, opiate, oxycodone), positive hepatitis C virus (HCV) antibody and positive/quantifiable HCV viral load;[25]
- Diagnoses (D): ICD 9 and 10 codes for overdose, substance use disorders, substance-related disorders, and Hepatitis C (Supplemental Table 1);

Supplementaty Table 1: List of ICD9/10 Codes for Inclusion into PWUD Cohort

Parent Code	Description
F11	Opioid related disorders
F14	Cocaine related disorder
F15	Other stimulant related disorders
T400-T406; T436	Poisoning by opium, heroin, other opioids, methadone, synthetic narcotics, cocaine, unspecified narcotics and psychostimulants.
0.70.41, 070.44, 070.51, 070.54, 070.70, 070.71	ICD9 Hepatitis Code
B18.2	Chronic viral hepatitis C

- Medications (M): Sublingual buprenorphine (Suboxone or Subutex) or oral methadone listed as medications on the outpatient medication reconciliation, given during hospitalization, or prescribed on discharge. We note that methadone for OUD would not be a discharge medication that is prescribed, but rather a medication listed in the medication discharge reconciliation.[26]
- Natural Language Processing (N): The study team iteratively produced and refined lists of keywords that are commonly used to describe PWUD in EMR (Supplemental Table

2). Natural language Processing codes were built by the data scientists at Tufts Clinical and Translational Science Institute using the software R. The algorithm is capable of detecting words in a context-specific manner, such as detecting simple misspellings or aggregated words, and also excluding negated words. The algorithm was run on the entire EMR, including but not limited to nursing notes, physician notes, discharge summaries, and emergency room records.

Supplement 2: List of words programmed into NLP to detect SUD

IVDU, FENTANYL, Methadone, heroin, suboxone, IVDA, drug abuse, SUD, Substance use disorder, opioid use disorder, opioid abuse, OUD, opioid overdose, illicit drugs, addicted, addict, drug addict, injection drug use, intravenous drug use, uses fentanyl, Uses heroin, PWID, abuses drugs, injects heroin, injects drugs, injects fentanyl.

Analysis:

(1) *Descriptive Analysis:* We depicted how the people entered the cohort visually using a modified venn diagram incorporating all of the different cohort entry definitions (B, D, M, N). (2) *Positive Predictive value (PPV) calculation:* From subcohorts “D-only” and “N-only”, we randomly selected charts to manually validate the PWUD designation through chart review. The chart review involved assessing three types of notes in each chart: emergency department admission note, history of present illnesses, and discharge summary. If clear indications of a substance use disorder, excluding alcohol, cannabis, and prescribed medications, the chart was marked as “true positive” for PWUD. Chart review was completed by two reviewers (EDG, TS) (3) *Logistic Regression:* The outcome of interest is high-level of documentation (brought into the cohort by the presence of all of the 4 domains (B, D, M, N) vs. NLP only; comparing these two groups allowed to assess factors that correlate with encounters where SUD goes unnoticed or undercharted versus encounters where SUD is well-documented. In addition to the above data, each encounter also had linked demographics data (age, race, ethnicity, gender), length of hospital stay, and Social

Vulnerability Index (SVI). The SVI is a tool developed by the CDC that is used to assess the community's susceptibility to disasters and emergencies; it uses 16 census-based data points to help assess local communities' need for aid before and after disaster.[27] It evaluates factors like socioeconomic status, disability, minority status, and areas that may need additional support during crises. It is a holistic way to represent the social and economic stability of neighborhoods. The SVI was provided as a quartile (1,2,3,4), with 1 representing the highest level of structural vulnerability. Utilizing Stata, we examined the association of key indicators (race and SVI) and the level of documentation for substance use disorders. IRB Approval: The study has been approved by the Health Sciences Institutional Review Board of TuftsMC.

Results:

In Figure 1, the distribution of how 4548 hospitalizations entered the cohort by domains (BDMN) is depicted in a Venn diagram. The subcohorts with the highest number of hospitalizations were those with diagnoses only (n=958), biomarkers only (n=734), and all four criteria: biomarkers, diagnoses, medications, and NLP (n=726). The PPV of the "D-only" cohort, where the encounter was determined to be PWUD based only on ICD codes, was 43% and the PPV of the "N-only" cohort, where the encounter was determined to be PWUD based only on NLP algorithm, was 54% (Table 1). In the multivariable regression (Table 2), people who identify as White Non-Hispanic had higher odds of entering the cohort through NLP alone. (Adjusted OR=2.07, 95% CI=[1.54, 2.79]). People who were in the lower 1st and 2nd quartiles of SVI (the socio-economically disadvantaged group) were also more likely to be in the NLP alone group (Adjusted OR=1.41 95% CI=[1.06, 1.88]).

Table 1: Positive Predictive Values of NLP-only cohort and ICD-only cohorts

Cohort	Number of hospitalizations in the cohort	Number of charts reviewed	Number of charts confirmed as true PWUD by chart review	Positive Predictive Value
D: diagnostic	958	93	40	43%

codes present				
N: NLP present	288	99	53	54%

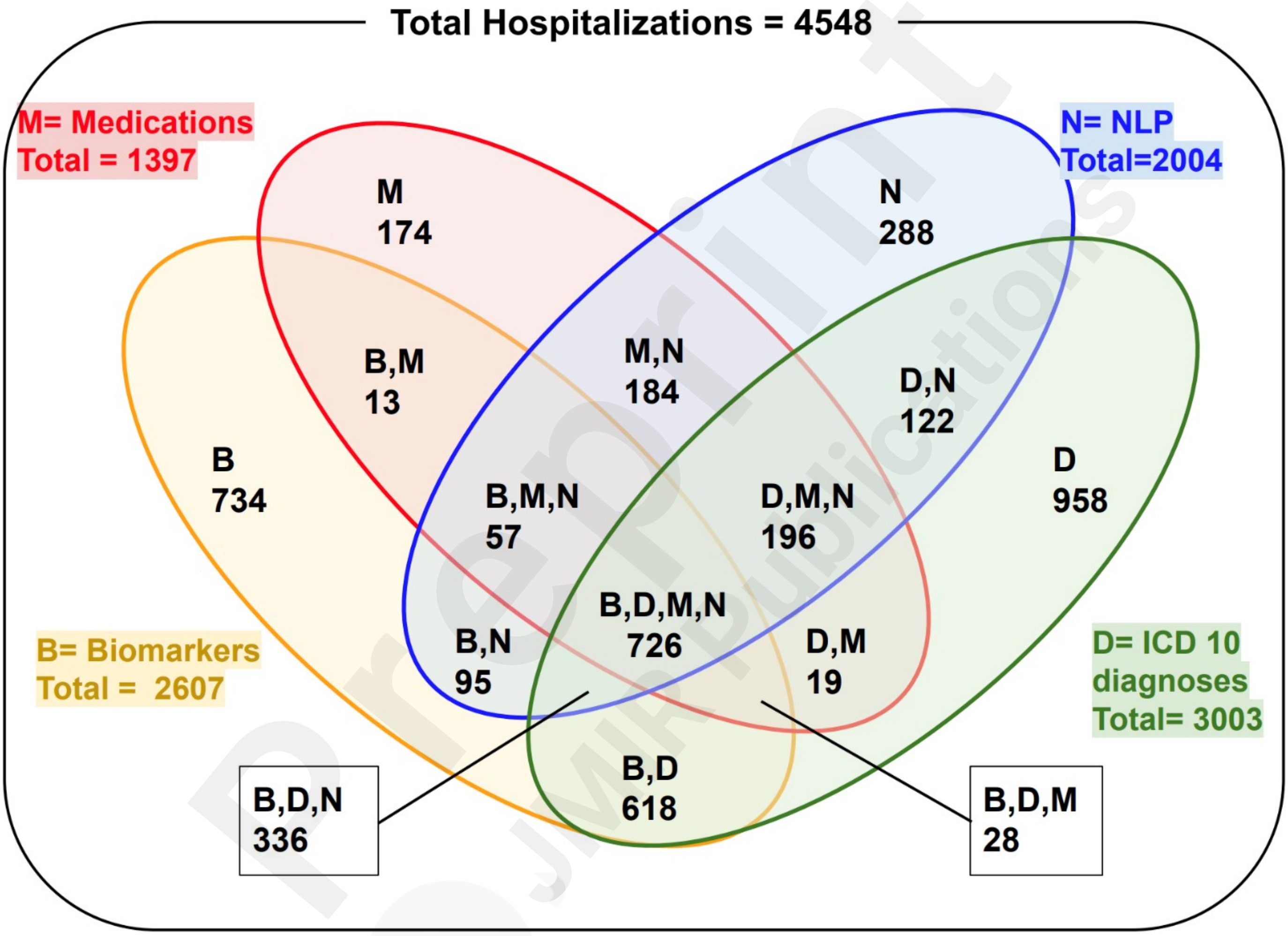
Table 2: Descriptive Analysis of PWUD Cohort and Factors Associated with Entering the Cohort As Highly-Documented (BDMN) or Minimally-Documented (NLP only)

		Criteria for Entering Cohort		Un-adjusted OR (95% CI)	Adjusted OR (95% CI)
	Entire Cohort	BDMN	N only		
	n=4548 (encounters)	N= 726 (encounters)	N= 288 (encounters)		
Age, Mean (SD)	47.9 (13.8)	43.3 (10.7)	45.6 (14.5)		
Sex (%) Male Female	 2,837 (62.4) 1,711 (37.6)	 457 (62.9) 269 (37.1)	 155 (53.8) 133 (46.2)		
Race/Ethnicity (%) Racially/Ethnically Minoritized* White/Non-Hispanic	 1,583 (34.8) 2,965 (65.2)	 176 (24.2) 550 (75.8)	 114 (60.4) 174 (39.6)	 1.00 (Ref) 2.04 (1.53, 2.73)	 1.00 (Ref) 2.07 (1.54, 2.79)*
Length of Hospitalization, Mean (SD)	38.7 (26.3)	41.5 (25.7)	34.7 (26.4)		
SVI (Quartile) 3rd-4th 1st-2nd Missing	 2,462 (54.1) 2,070 (45.51) 16 (0.4)	 461 (63.5) 262 (36.1) 3 (0.4)	 163 (56.6) 124 (43.1) 1 (0.4)	 1.00 (Ref) 1.34 (1.01, 1.77)	 1.00 (Ref) 1.41 (1.06, 1.88)**
Urine Toxicology (%) Opiate Fentanyl Oxycodone Methadone Cocaine Amphetamine	 658 (14.5) 1,313 (24.9) 369 (8.1) 272 (5.9) 622 (13.7) 323 (7.1)	 136 (18.7) 430 (59.2) 66 (9.1) 224 (30.9) 258 (35.5) 153 (21.1)	 0 0 0 0 0 0		
Primary Language (%) English Spanish	 4,296 (94.5) 123 (2.7)	 703 (96.8) 23 (3.2)	 270 (93.8) 11 (3.8)		
The group of Racially/Ethnically Minoritized Populations: Black: 773 (17%); Hispanic: 469 (10.3%) Asian: 122 (2.7%): Asian Indian: 24 (0.5%); Hawaiian (0.02%): Other 22 (0.5%): Unknown 172 (3.8%)					

*Multivariable model adjusted for age, sex, and SVI

**Multivariable model adjusted for age, sex, and race

Figure 1: Venn Diagram illustrating the number of hospitalizations in each cohort



Discussion:

Most studies about healthcare for hospitalized PWUD rely on using ICD10 codes, however we found limiting to only ICD-10 codes, we would have identified only 43% of PWUD in the cohort. NLP has the potential to uncover cases of SUD that have gone undocumented and give a more accurate view of the population impacted. NLP had greater PPV than diagnostic codes, but still, the PPV was low. We found that PWUD from racially and ethnically minoritized communities and people who are poor were more likely to be represented in the minimally documented cohort than the maximally documented cohort.

Our research augments this body of work by integrating NLP with diverse identification methods, including urine toxicology and medication records, while simultaneously addressing observed demographic disparities in documentation.[28] Largely a result of stigma and racism, people who suffer from drug use disorder still do not have universal access to evidence-based treatment. Black individuals with SUD tend to enter treatment with a more severe prognosis compared to their white counterparts, due in part to economic barriers in accessing treatment earlier.[29] Black, Latino, and Native American individuals also face additional challenges in accessing SUD treatment due to geographic barriers, healthcare access, and potential community characteristics or rapport with clinicians.[30]

Identification of PWUD who access medical care is important for several reasons. Best practice guidelines for hospitalized PWUD include management of substance use disorder, pain, and acute infection, testing and management for HIV and HCV, (3) vaccinations for hepatitis or other relevant infections, and (4) prevention of HIV with medications. [31,32] In the current study, we applied NLP retrospectively. Building on work done that identified low HIV testing rates, we plan to use NLP to augment PWUD cohort creation in a study examining patterns of HIV testing.[33,34] NLP could indeed become a valuable tool for identifying PWUD prior to discharge and facilitating intervention during hospitalization if electronic medical records could use NLP to trigger clinical decision support tools that trigger clinicians to consider SUD treatment, prescribe overdose prevention medications at discharge, order labs to prepare for pre-exposure prophylaxis (PrEP), or offer vaccines.

As we consider this study in the larger context of improving health equity, we believe that the next step would be refining the NLP system by adding more keywords, including and excluding certain conditions and medication, and conducting analyses on false positives and false negative cases. This study should be replicated in other medical centers across the U.S.; its wider application across different hospitals, encapsulating diverse populations and regions, will be instrumental. The study also has multifaceted applications, spanning epidemiological tracking, optimizing hospital resource utilization, and influencing the design of specific interventional studies. The current data could serve as a launchpad for integrated care for PWUD with less prejudice and inequity.

This study is not without its limitations. The NLP system, despite its effectiveness, occasionally misidentifies certain keywords. The constant calibration of the algorithm and frequent addition of keywords is needed to optimize and sustain accuracy. There are potential flaws in our characterization of domains. For example, the use of "amphetamine" as an indicator of SUD has caused inadvertent inclusion of patients prescribed amphetamines for conditions like ADHD. Similarly, methadone prescriptions at the end of a hospital stay might suggest pain management for conditions such as sickle cell disease, rather than its more commonly known use in addiction treatment. Achieving a balance between NLP's inclusivity and exclusivity presents a significant challenge for this project. Future steps should include evaluating the NLP system's sensitivity and specificity and iterating on the model to enhance these metrics. This will involve refining the keyword list for SUD, enhancing the NLP algorithm to better account for common confounding variables, and fine-tuning the inclusion and exclusion criteria to ensure the system's accuracy and reliability.

Despite these limitations, we feel that this research helps frame the future of systems to measure healthcare delivery to PWID. Hospitalization represents a crucial moment when non-judgmental trauma-informed, culturally competent care can be offered to PWUD. By harnessing NLP, we can transcend human limitations of error and bias, striving for more comprehensive and equitable care. In this era of digital innovation, algorithms can effectively link providers with patients. The time for this transformative change is now.

REFERENCES

1. Chiosi JJ, Mueller PP, Chhatwal J, Ciaranello AL. A multimorbidity model for estimating health outcomes from the syndemic of injection drug use and associated infections in the United States. BMC Health Serv Res 2023;23:760.
2. Mattson CL, Tanz LJ, Quinn K, Kariisa M, Patel P, Davis NL. Trends and Geographic Patterns in Drug and Synthetic Opioid Overdose Deaths - United States, 2013-2019. MMWR Morb Mortal Wkly Rep 2021;70:202-7.
3. Sun J, Mehta SH, Astemborski J, et al. Mortality among people who inject drugs: a prospective cohort followed over three decades in Baltimore, MD, USA. Addiction 2022;117:646-55.
4. Hollander MAG, Chang CH, Douaihy AB, Hulsey E, Donohue JM. Racial inequity in medication treatment for opioid use disorder: Exploring potential facilitators and barriers to use. Drug Alcohol Depend 2021;227:108927.
5. Hamdan S, Smyth E, Murphy ME, et al. Racial and Ethnic Disparities in HIV Testing in People Who Use Drugs Admitted to a Tertiary Care Hospital. AIDS Patient Care STDS 2022;36:425-30.
6. Leo Knudsen Westgard JPR, Finlay Pilcher, Ayesha A. Appa, Jaimie P. Meyer, Uriel R. Felsen, Emily D. Grussing, Taisuke Sato, Laura R. Marks, Kinna Thakarar, Brian T. Montague, Ank Nijhawan, William S Bradford, Alysse G. Wurcel and the PWUD Care Workgroup (PCW). Abstract: National HIV and HCV Testing Rates for Hospitalized People Who Use Drugs are Heterogeneous and Suboptimal Across 11 U.S. Hospitals. Conference on Retroviruses and Opportunistic Infections (CROI). Denver, Colorado2023.
7. Ti L, Ti L. Leaving the Hospital Against Medical Advice Among People Who Use Illicit Drugs: A Systematic Review. American Journal of Public Health 2015;105:e53-e9.
8. Hazen A, Pizzicato L, Hom J, Johnson C, Viner KM. Association between discharges against medical advice and readmission in patients treated for drug injection-related skin and soft tissue infections. Journal of Substance Abuse Treatment 2021;126:108465.
9. Eaton EF, Westfall AO, McClesky B, et al. In-Hospital Illicit Drug Use and Patient-Directed Discharge: Barriers to Care for Patients With Injection-Related Infections. Open Forum Infectious Diseases 2020;7.
10. Straub L, Gagne JJ, Maro JC, et al. Evaluation of Use of Technologies to Facilitate Medical Chart Review. Drug Saf 2019;42:1071-80.
11. Schaper E, Padwa H, Urada D, Shoptaw S. Substance use disorder patient privacy and comprehensive care in integrated health care settings. Psychol Serv 2016;13:105-9.
12. Saczynski JS, Andrade SE, Harrold LR, et al. A systematic review of validated methods for identifying heart failure using administrative data. Pharmacoepidemiology and drug safety 2012;21:129-40.
13. Segar MW, Keshvani N, Rao S, Fonarow GC, Das SR, Pandey A. Race, social determinants of health, and length of stay among hospitalized patients with heart failure: an analysis from the Get With The Guidelines-Heart Failure registry. Circulation: Heart Failure 2022;15:e009401.
14. Campanile Y, Silverman M. Sensitivity, specificity and predictive values of ICD-10 substance use codes in a cohort of substance use-related endocarditis patients. The American Journal of Drug and Alcohol Abuse 2022;48:538-47.
15. McGrew KM, Homco JB, Garwe T, et al. Validity of International Classification of Diseases codes in identifying illicit drug use target conditions using medical record data as a reference standard: A systematic review. Drug and Alcohol Dependence 2020;208:107825.
16. Ball LJ, Sherazi A, Laczko D, et al. Validation of an Algorithm to Identify Infective Endocarditis in People Who Inject Drugs. Med Care 2018;56:e70-e5.
17. Zalesak M, Francis K, Gedeon A, et al. Current and future disease progression of the chronic HCV population in the United States. PLoS One 2013;8:e63959.
18. Ehrenfeld JM, Gottlieb KG, Beach LB, Monahan SE, Fabbri D. Development of a Natural

Language Processing Algorithm to Identify and Evaluate Transgender Patients in Electronic Health Record Systems. Ethn Dis 2019;29:441-50.
19. Sims SA, Snow LA, Porucznik CA. Surveillance of methadone-related adverse drug events using multiple public health data sources. Journal of biomedical informatics 2007;40:382-9.
20. Ridgway JP, Uvin A, Schmitt J, et al. Natural language processing of clinical notes to identify mental illness and substance use among people living with HIV: retrospective cohort study. JMIR Medical Informatics 2021;9:e23456.
21. Goodman-Meza D, Tang A, Aryanfar B, et al. Natural Language Processing and Machine Learning to Identify People Who Inject Drugs in Electronic Health Records. Open Forum Infect Dis 2022;9:ofac471.
22. Bartholomew TS, Tookes HE, Spencer EC, Feaster DJ. Application of machine learning algorithms for localized syringe services program policy implementation - Florida, 2017. Ann Med 2022;54:2137-50.
23. Cresta Morgado P, Carusso M, Alonso Alemany L, Acion L. Practical foundations of machine learning for addiction research. Part I. Methods and techniques. Am J Drug Alcohol Abuse 2022;48:260-71.
24. Rivero-Juárez A, Guijo-Rubio D, Tellez F, et al. Using machine learning methods to determine a typology of patients with HIV-HCV infection to be treated with antivirals. PLoS One 2020;15:e0227188.
25. Reed JR, Jordan AE, Perlman DC, Smith DJ, Hagan H. The HCV care continuum among people who use drugs: protocol for a systematic review and meta-analysis. Syst Rev 2016;5:110.
26. Hoffman KA, Ponce Terashima J, McCarty D. Opioid use disorder and treatment: challenges and opportunities. BMC Health Serv Res 2019;19:884.
27. Registry AfTSaD. CDC/ATSDR Social Vulnerability Index. 2024.
28. Goodman-Meza D, Tang A, Aryanfar B, et al. Natural Language Processing and Machine Learning to Identify People Who Inject Drugs in Electronic Health Records. Open Forum Infectious Diseases 2022;9.
29. Matsuzaka S, Knapp M. Anti-racism and substance use treatment: Addiction does not discriminate, but do we? J Ethn Subst Abuse 2020;19:567-93.
30. Acevedo A, Panas L, Garnick D, et al. Disparities in the Treatment of Substance Use Disorders: Does Where You Live Matter? J Behav Health Serv Res 2018;45:533-49.
31. Calcaterra SL, Bottner R, Martin M, et al. Management of opioid use disorder, opioid withdrawal, and opioid overdose prevention in hospitalized adults: A systematic review of existing guidelines. J Hosp Med 2022;17:679-92.
32. Thakarar K, Weinstein ZM, Walley AY. Optimising health and safety of people who inject drugs during transition from acute to outpatient care: narrative review with clinical checklist. Postgraduate medical journal 2016;92:356-63.
33. E DG, Pickard B, Khalid A, et al. Implementation of a bundle to improve HIV testing during hospitalization for people who inject drugs. Implement Res Pract 2023;4:26334895231203410.
34. Zubiago J, Murphy M, Guardado R, Daudelin D, Patil D, Wurcel A. Increased HIV testing in people who use drugs hospitalized in the first wave of the COVID-19 pandemic. J Subst Abuse Treat 2021;124:108266.