Original Research

Substance Use & Addiction Journal
1–9

DOI: 10.1177/29767342241288112
journals.sagepub.com/home/saj
Sage

One Third of Alcohol Use Disorder Diagnoses are Missed by ICD Coding

Laura Mercurio, MD[1], Augusto Garcia, PhD[2], Stephanie Ruest, MD, MPH[1,3], Susan J. Duffy, MD, MPH[1], and Carsten Eickhoff, PhD[4,5]

Abstract

Background/Significance: Alcohol use carries significant morbidity and mortality, yet accurate identification of alcohol use disorder (AUD) remains a multi-layered problem for both researchers and clinicians.

Objective: To fine-tune a language model to AUD in the clinical narrative and to detect AUDs not accounted for by ICD-9 coding in the MIMIC-III database.

Materials and Methods: We applied clinicalBERT to unique patient discharge summaries. For classification, patients were divided into nonoverlapping groups stratified by the presence/absence of AUD ICD diagnosis for model training (80%), validation (10%), and testing (10%). For detection, the model was trained (80%) and validated (20%) on 1:1 positive/negative patients, then applied to remaining negative patient population. Physicians adjudicated 600 samples from the full model confidence spectrum to confirm AUD by Diagnostic and Statistical Manual of Mental Disorders-V criteria.

Results: The model exhibited the following characteristics (mean, standard deviation): precision (0.9, 0.02), recall (0.65, 0.03), F-1 (0.75, 0.02), area under the receiver operating curve (0.97, 0.01), and area under the precision-recall curve (0.86, 0.01). Adjudication produced an estimated 4% under-documentation rate for the total study population. As model confidence increased, AUD under-documentation rate rose to 30% of the number of patients identified as positive by ICD-9 coding.

Conclusion: Our model improves the identification of patients meeting AUD criteria, outperforming ICD codes in detecting cases of AUD. Detection discrepancy between ICD and free-text highlights clinician *under documentation*, not under recognition. Adjudication revealed model over-sensitivity to language around substance use, withdrawal, and chronic liver disease; future study requires application to a broader set of patient age and acuity. This model has the potential to improve rapid identification of patients with AUD and enhance treatment allocation.

Keywords

alcohol use disorder, natural language processing, electronic health records

Highlights

- Natural language processing improves identification of alcohol use disorder in the electronic health record
- A fine-tuned clinical BERT model identified > 1400 patients with alcohol use disorder not previously detected by structured diagnostic codes
- The discrepancy between diagnostic codes and free-text highlights clinician under-documentation, not under-recognition

Introduction

In the 2022 National Survey on Drug Use and Health, 1 in 5 respondents above 18 years of age reported binge drinking in the past month,[1] 6.3% reported heavy alcohol use in the past month,[2] and 1 in 10 met alcohol use disorder (AUD) criteria.[1] Researchers studying alcohol consumption also identified that individuals exceeding the gender-specific binge drinking thresholds were 70 times more likely to have an alcohol-related emergency department visit.[1,2] Meanwhile, epidemiologic studies demonstrate a sharp rise in almost all types of substance use since the onset of the COVID-19 pandemic,[3] with 1 national survey indicating the highest documented levels to date of alcohol (32%) and marijuana (43%) use among adults aged 19 to 30 years.[4] To date, AUD has been associated with disease in almost every organ system, ranging from depression and cognitive decline to liver failure, pancreatitis, gastrointestinal bleeding, impaired wound healing, ischemic heart disease, and arrhythmias.[5] Given the associated morbidity and mortality, both researchers and clinicians seek to better understand the landscape of AUDs.

Yet, accurate identification of AUD remains a multi-layered problem.[6-9] Prior studies examining medical record documentation reveal significantly lower rates of ICD-coded alcohol use disorder (6.3%),[9,10] compared with those of AUD solicited via structured research interviews (13.9%-22%).[11,12] Furthermore, patients are hesitant to disclose the truth of their use due to fear of stigma, while clinicians do not always solicit or document this information. When patients *do* disclose their alcohol use and clinicians record it, their disclosure often occurs in the clinical narrative without a corresponding diagnosis code in the international classification of diseases (ICD).[9,13-17] As a result, what we *do* know about patient alcohol use remains "buried" as unstructured, sparse data; investigations relying on discrete variables gathered from medical and diagnostic codes significantly underestimate alcohol consumption and its effects. In response, several large-scale natural language processing studies have shown promise in extracting substance use disorder (SUD) information from data sources including the Medical Information Mart for Intensive Care (MIMIC) database[13,15] and other large clinical corpora.[18-20]

In this study, we constructed a text classification model to accurately identify AUDs in the clinical narrative not accounted for by traditional diagnostic codes. We built a text-recognition model using a "bidirectional encoder representations from transformers (BERT)" text encoder, modeled with clinical notes,[21] to handle language complexity in the medical context[22] and applied it to 2 main scenarios: AUD classification and detection. In the first scenario, we evaluated the model's ability to accurately classify known patients with/without an AUD based on ICD diagnostic codes. In the detection experiment, we evaluated the model's ability to prospectively identify cases of AUD among previously uncharacterized patients from patient discharge summaries *without* an ICD diagnosis of AUD. We hypothesized that a text-recognition screener will accurately identify patients with diagnoses of known AUDs; we further posited that this model would identify additional patients with clinical characteristics of AUD but without related ICD diagnoses.

Methods

Data Source

Data for this study were extracted from the MIMIC-III relational database,[24] a publicly available dataset of 46 520 patients admitted to the critical care units of the Beth Israel Deaconess Medical Center from 2001 to 2012, in 58 976 unique hospital admissions. This database contains multiple types of documentation related to patient care such as vital signs, medications, laboratory testing, nursing, social work, and physician documentation; patient discharge summaries were extracted from the database for model training and evaluation.

Inclusion and Exclusion Criteria

All patients with available narrative text were included (Figure 1), with isolation of the single *most recent* hospital encounter discharge summary for each unique MRN (N = 42 929). We selected the most recent encounter to reflect the most up-to-date and comprehensive problem list. We henceforth refer to each encounter as a "Patient." Patients were classified as "true positives" for an AUD only if their medical record contained an ICD-9 diagnosis in the AUD category (Supplemental Appendix A). Discharge summaries were also preprocessed to combine any erroneously duplicated or divided text or addenda.

Definition of AUD

Classification of AUD in medical settings is typically determined by patient disclosure documented in the medical record, or a combination of positive screening responses and medical conditions. In this study, patients that receive at least 1 ICD-9 diagnosis in the AUD category (Supplemental Appendix A) qualified as having a "known" disorder; ICD-10 diagnostic codes were not yet implemented at the time of dataset publication. These patients are referred to as "positive." Those patients without an associated ICD-9 diagnosis are referred to as "negative." Patients identified as having potential AUD diagnoses by the text-recognition model were hand-reviewed by 3 physician adjudicators to determine whether the patient met AUD criteria as determined by the Diagnostic and Statistical Manual of Mental Disorders-V (DSM-V)[23] (Table 1).

Model Construction and Evaluation

Our text-recognition model was constructed based on clinicalBERT,[21] which has an input-context limitation of approximately 318 words (512 tokens) based on clinicalBERT's pretraining vocabulary and our dataset. Adapting

[1]Departments of Emergency Medicine & Pediatrics, Alpert Medical School of Brown University, Providence, RI, USA
[2]Division of Rheumatology, University of California San Francisco, San Francisco, CA, USA
[3]Injury Prevention Center of Rhode Island Hospital-Hasbro Children's Hospital, Providence, RI, USA
[4]Department of Computer Science, Brown University, Providence, RI, USA
[5]School of Medicine, University of Tubingen, Tubingen, Germany

Corresponding Author:
Laura Mercurio, Department of Emergency Medicine, Section of Pediatric Emergency Medicine, Alpert Medical School of Brown University, 55 Claverick Street, 2nd Floor, Providence, RI 02903, USA.
Email: laura.mercurio@brownphysicians.org

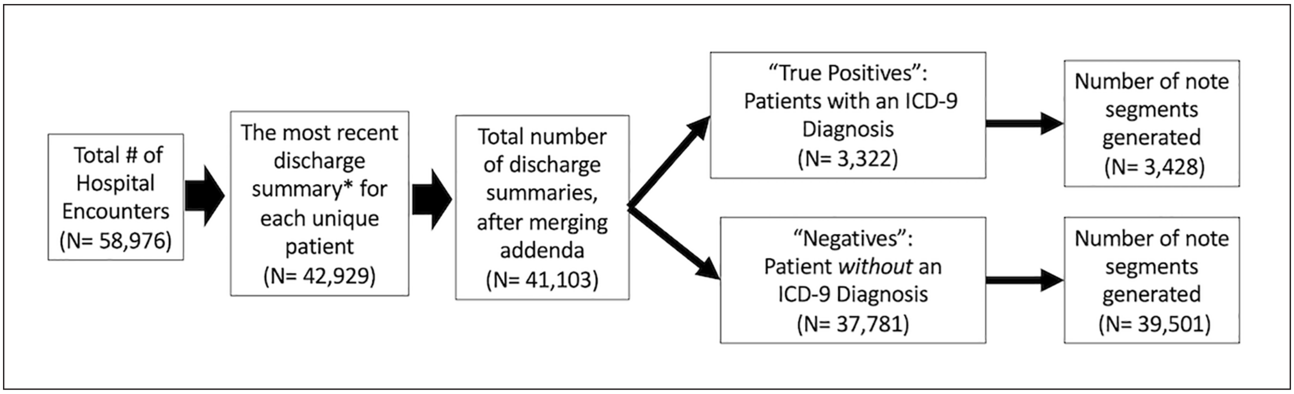

Figure 1. Patient inclusion flow diagram.

Table 1. Criteria for AUD, from the Diagnostic and Statistical Manual of Mental Disorders (DSM-V).[23]

The presence of at least 2 of these symptoms indicates an AUD	The severity of the AUD is defined as: *Mild*: The presence of 2 to 3 symptoms *Moderate*: The presence of 4 to 5 symptoms *Severe*: The presence of 6 or more symptoms

- Had times when you ended up drinking more, or longer, than you intended?
- More than once wanted to cut down or stop drinking, or tried to, but couldn't?
- Spent a lot of time drinking? Or being sick or getting over other aftereffects?
- Wanted a drink so badly you couldn't think of anything else?
- Found that drinking—or being sick from drinking—often interfered with taking care of your home or family? Or caused job troubles? Or school problems?
- Continued to drink even though it was causing trouble with your family or friends?
- Given up or cut back on activities that were important or interesting to you, or gave you pleasure, in order to drink?
- More than once gotten into situations while or after drinking that increased your chances of getting hurt (such as driving, swimming, using machinery, walking in a dangerous area, or having unsafe sex)?
- Continued to drink even though it was making you feel depressed or anxious or adding to another health problem? Or after having had a memory blackout?
- Had to drink much more than you once did to get the effect you want? Or found that your usual number of drinks had much less effect than before?
- Found that when the effects of alcohol were wearing off, you had withdrawal symptoms, such as trouble sleeping, shakiness, restlessness, nausea, sweating, a racing heart, or a seizure? Or sensed things that were not there?

Abbreviation: AUD, alcohol use disorder.

an approach originally published by Huang et al,[29] the model split each patient discharge summary into 318-word segments, with each segment being associated with the patient's global outcome label (Figure 2). As a result, the model backpropagated from the output label to each individual segment during the training phase. In the evaluation phase, the individual scores (the output logit for the positive class) of each segment were combined according to the following equation:

$$p(h=1 \mid (P_i,\ldots,P_{\mathrm{n}})=\frac{P_{\mathrm{n}_{\max}}+P_{\mathrm{n}_{\mathrm{mean}}}\dfrac{\mathrm{n}}{c}}{1+\dfrac{\mathrm{n}}{c}}$$

where n is the number of segments for that patient; $(P_i,\ldots,P_{\mathrm{n}})$ are the individual scores; $P_{\mathrm{n}_{\max}}$ and $P_{\mathrm{n}_{\mathrm{mean}}}$ are the maximum and mean individual scores, respectively; and c is a weighing coefficient chosen heuristically from the training and validation sets. For the final label assignment, a cutoff value of p was also heuristically calculated from the train and validation sets.

Experiment 1, Classification of AUD

Patients were divided into nonoverlapping groups, stratified by outcome class (presence/absence of AUD ICD diagnosis), for model training, validation, and testing: 80% in the training set, 10% in the validation set, and 10% in the test set (Figure 3). Patients were assigned to these subsets prior to the division of each discharge summary into 318-word segments; exact segment-counts for stages of each experiment are outlined in Figure 3. Final segment counts varied in each subset due to the unique length of each discharge summary. Training set class balancing reduced model performance, so it was not applied to the

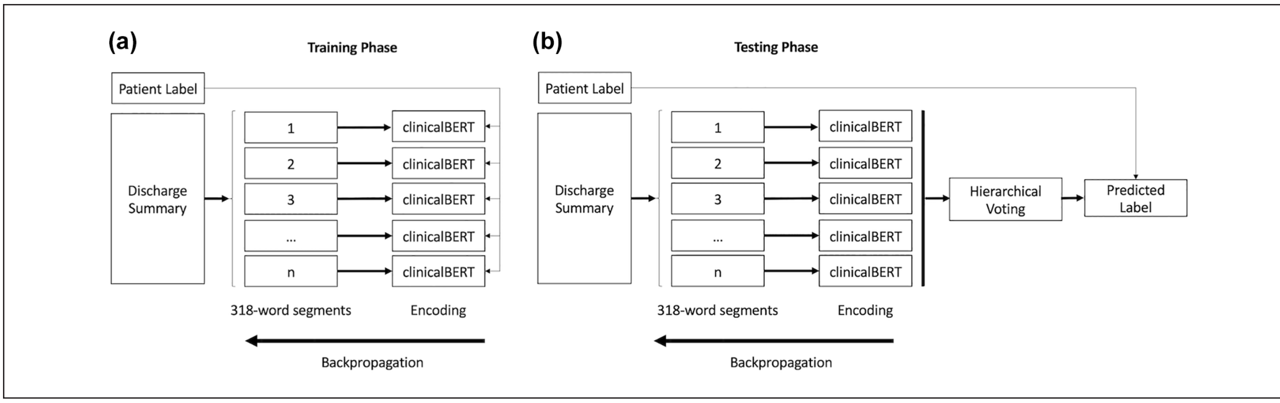

Figure 2. (a) clinicalBERT model construction, training phase. Each discharge summary is associated with an outcome label, then divided into 318-word segments, the approximate input limit for the clinicalBERT model. The outcome label for each patient is then backpropagated to every corresponding text segment. (b) clinicalBERT model construction, testing phase. After segmentation and encoding, segments are used to predict an outcome label, which is then compared with the originally assigned patient label.
Abbreviation: BERT, bidirectional encoder representations from transformers.

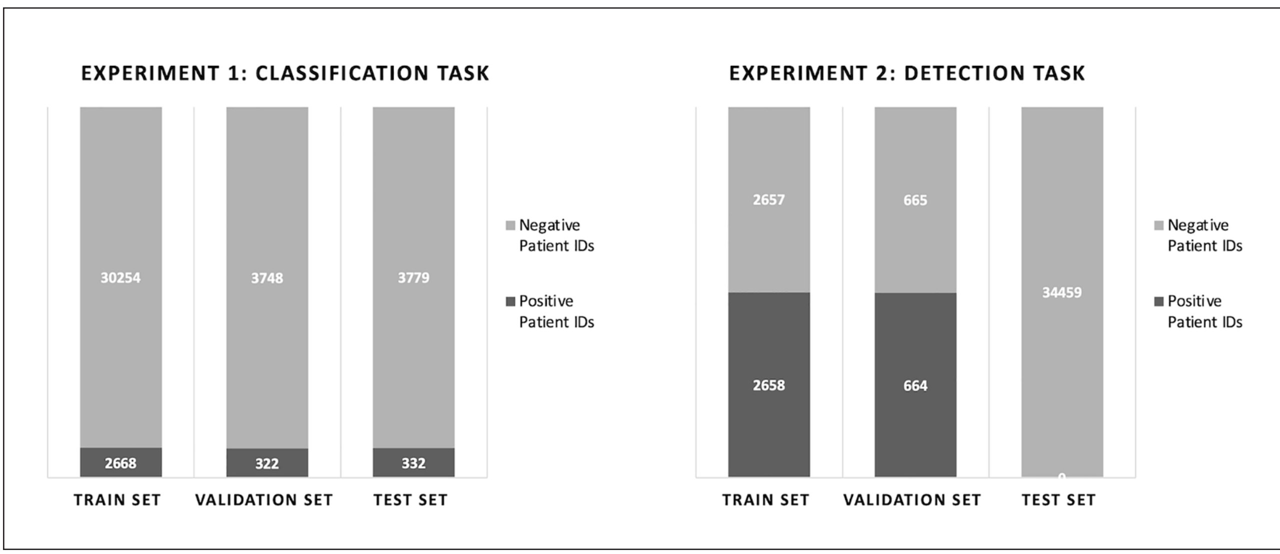

Figure 3. Distribution of patients in Experiments 1 and 2. Each "patient" represents the most recent discharge summary available for a unique person.

final model. We used a 5-fold cross-validation technique to measure confidence intervals of model performance in classifying AUD (Figure 3). For the final label assignment, a cutoff value of p was heuristically calculated from the train and validation sets (see section on "Model Construction & Evaluation").

We evaluated model performance using accuracy, precision, recall, F-1 score—the harmonic mean between precision and recall, area under the receiver operating curve (AUROC), and, given the significant class imbalance, the area under the precision-recall curve (AUPRC). Of note, our outcome label—ICD-9 coding—imperfectly reflects the actual documented presence or absence of AUD for multiple reasons. First, we can assume clinicians failed to evaluate some patients for AUD; in others, it may have been assessed but not documented. Finally, some patients may have narrative text supporting AUD but did not receive an ICD-9 code.

Experiment 2: Detection of AUD in Previously Unidentified Patients

All patients with an AUD ICD-9 diagnosis—"positives," were combined in a 1:1 proportion to randomly chosen patients without an ICD-9 diagnosis—"negatives." This 1:1 combination was then divided into training (80%) and validation (20%) subsets. The remainder of the study population did not have an associated ICD-9 AUD diagnosis; these patients comprised final test set, with a total of N = 34 459 discharge summaries. The discharge summaries

Table 2. Inter-Rater Reliability Sample: Top 100 Patients with AUD as Identified by the Text Classification Model.

Physician Reviewer	Patients meeting AUD criteria	Consensus agreement	Inter-rater reliability (95% CI)
Adjudicator 1	63%	Full Consensus = 38%	Adjudicators 1 and 2 Kappa = 0.40 (0.24, 0.59)
Adjudicator 2	55%	Majority Consensus = 63%	Adjudicators 1 and 3 Kappa = 0.39 (0.23, 0.54)
Adjudicator 3	47%		Adjudicators 2 and 3 Kappa = 0.69 (0.56, 0.81)

Abbreviation: AUD, alcohol use disorder.

in each subset were then split into the 318-word segments required by clinicalBERT; Figure 3 demonstrates the final distribution of patient discharge summaries for each task.

Using the test set, the model generated a confidence score reflecting the likelihood that each segment contained language supporting an AUD. Aggregating the scores of each segment, the model ultimately categorized each patient discharge summary as positive or negative for AUD.

We validated model performance by extracting a representative sample of patients across the *full* spectrum of model confidence (N = 600). Sample selection was performed by binning all test set discharge summaries into 99 "buckets" of 348 to 349 patients, each of which containing a proportional percentile of model confidence space (Supplemental Appendix B). Afterward, 6 summaries per bucket were randomly selected for adjudication. Blinded to the model's confidence and categorization, 3 physicians hand-annotated the patient discharge summaries to confirm the presence of an AUD according to the DSM-V criteria.[23] We then determined the adjudicator ratio—the proportion of patients determined to meet AUD criteria and multiplied this ratio by the sample size to determine the rate of previously unidentified patients with AUD.

Finally, we extracted a separate subset of patient discharge summaries to assess inter-rater reliability (IRR). This subset contained the top 100 patients most likely to meet AUD criteria as identified by the model in Experiment 2. These discharge summaries were hand-annotated by the same 3 physicians. We then calculated the percentage of full and majority agreement by adjudicators, as well as an unweighted the Cohen Kappa statistic to determine IRR (Table 2).

Results

A total of 3428 patients carried an ICD-9 diagnosis of AUD disorder in the MIMIC-III database. The most common ICD-9 code was 305.00, alcohol abuse. Patients with an AUD diagnosis were on average younger, more likely to be male, and publicly insured (Table 3); the median hospital length of stay was similar between groups.

In Experiment 1—classifying ICU patients with AUD diagnoses, the model was trained for 3 epochs with a learning rate of 0.00002. The following results are presented as (mean, standard deviation) based on 5-fold cross-validation of all 3 subsets: training, validation, and testing. The text-recognition model demonstrated high precision (0.9, 0.02) and moderate recall (0.65, 0.03) in classifying patients with AUD. It also exhibited a good F-1 measure (0.75, 0.02)—the harmonic mean between precision and recall. Finally, the model demonstrated a very high AUROC (0.97, 0.01) and a high AUPRC (0.86, 0.01).

In Experiment 2, the model identified several previously undocumented patients with AUD. The cumulative histogram (Figure 4) depicts the distribution of model confidence in detecting AUD in the "all negative" test set. Among previously undiagnosed patients, the model identified approximately 35% of text segments to contain AUD language with at least 70% likelihood; the 70% cutoff was determined by the calculation as described in the section, "Model Construction & Evaluation." Based on physician adjudication, the estimated under documentation rate was 4% of the total population; as model confidence increased, the estimated rate of AUD under documentation rose to 1.5×, or 30% of the number of positive patients by ICD coding; the combination of clinicalBERT and manual annotation revealed up to 1416 additional patients meeting AUD criteria not detectable by structured diagnostic codes alone (Figure 4).

Finally, in our IRR sample, 38/100 of these patients qualified for AUD with unanimous adjudication, and an additional 30 candidates qualified for AUD with majority adjudicator agreement. Adjudicator 1 demonstrated lower IRR scores than the other adjudicators, while Adjudicators 2 and 3 were more consistent in their outcome determination.

Those patients identified by ICD-9 with AUD were similar to those identified by our cBERT model (Table 4) in median age, assigned sex, select substance use (cannabis, cocaine/amphetamines), and concurrent mental health diagnoses except mood disorders ($P = .02$). There was a higher proportion of patients who spoke English, or Other languages, among those identified by ICD-9 as compared to cBERT. Meanwhile, the population with AUD identified by cBERT demonstrated a proportion of concurrent opioid/sedative use, "other" substance use, and mood disorder diagnosis.

Discussion

Given the significant morbidity and mortality associated with AUDs,[1,25,26] this study offers a novel method of screening medical records for AUD. Our pilot model significantly improves the identification of patients meeting

Table 3. Sociodemographic Information, by ICD-9 Diagnosis of an AUD.

Patient Characteristic	Patients with AUD diagnosis (N = 3322), N (%)	Patients without AUD diagnosis (N = 37 781), N (%)	X^2 (degrees of freedom), *P*-value
Median age, in years [IQR]	53 [44, 62]	69 [48, 78]	<.01
Assigned sex[a]			
Male	2538 (76)	20 644 (55)	X^2 (1) = 617, <.001
Female	784 (24)	17 137 (45)	
Language			
English	2143 (65)	18 912 (50)	X^2 (2) = 2100, <.001
Spanish	87 (2.6)	687 (1.8)	
Other	1092 (33)	18 182 (48)	
Insurance			
Private	1259 (38)	14 081 (37)	X^2 (4) = 1378, <.001
Medicare	991 (30)	19 375 (51)	
Medicaid	716 (22)	2945 (7.8)	
Government	229 (7)	1051 (2.7)	
Self-pay	127 (3.8)	357 (0.9)	
Median hospital length of stay, in days [IQR]	7 [4, 14]	7 [4, 12]	>.05
Substance use, by type			
Opioids/sedatives	238 (7.2)	490 (1.3)	X^2 (1) = 604, <.001
Cannabis	253 (7.6)	282 (0.7)	X^2 (1) = 1121, <.001
Cocaine/amphetamines	61 (1.8)	66 (0.2)	X^2 (1) = 273, <.001
Other	88 (2.6)	153 (0.4)	X^2 (1) = 264, <.001
Mental health diagnosis			
Mood disorder	709 (21)	3456 (9.1)	X^2 (1) = 504, <.001
Delirium/dementia	173 (5.2)	2676 (7.1)	X^2 (1) = 3.03, .08
Anxiety	166 (5.0)	1244 (3.3)	X^2 (1) = 27, <.001
Drug-induced mental disorder	105 (3.2)	341 (0.9)	X^2 (1) = 145, <.001
Trauma/stressor-related	69 (2.1)	241 (0.6)	X^2 (1) = 84, <.001
Attention deficit hyperactivity disorder	35 (1.1)	114 (0.3)	X^2 (1) = 48, <.001
Developmental disorder	29 (0.8)	323 (0.9)	X^2 (1) = 0.01, .91
Schizophrenia/psychosis	32 (1.0)	268 (0.7)	X^2 (1) = 2.71, .10

Abbreviations: AUD, alcohol use disorder; IQR, interquartile range.
[a]Assigned sex is referred to as "gender" in the original database.

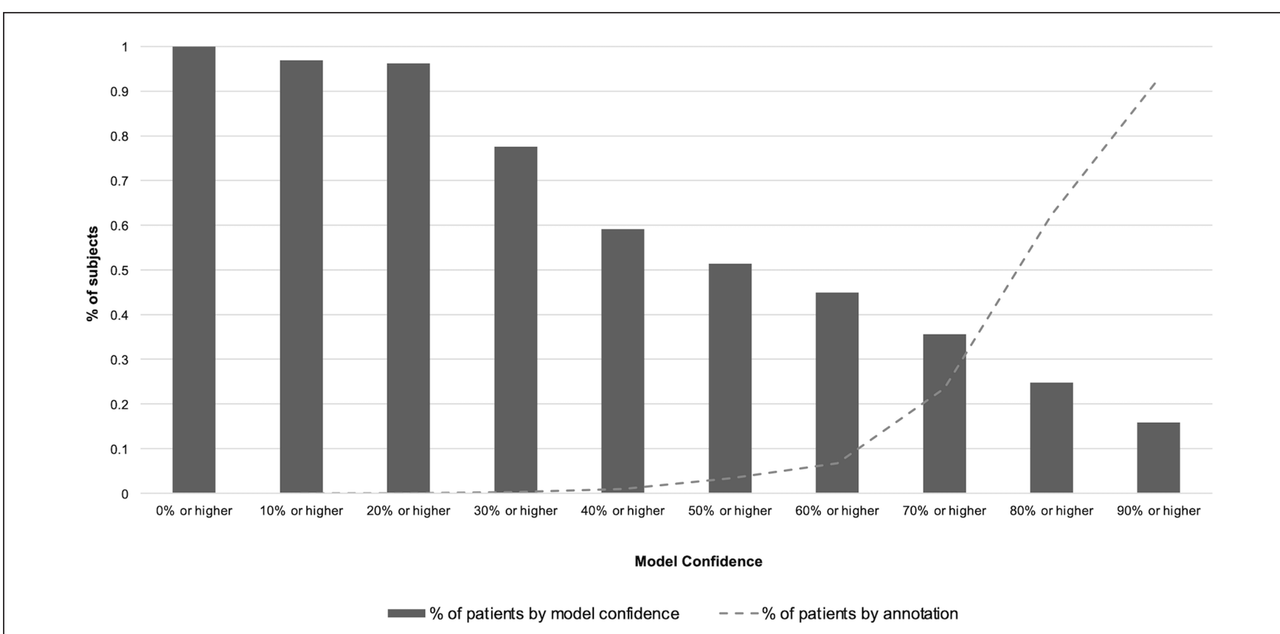

Figure 4. Histogram of model likelihood in identifying AUD in text segments of previously negative patients, and estimated rate of under-documentation by annotation. The higher the model confidence, the more likely the text segment contains language supporting an AUD diagnosis. For example, at 70% confidence, approximately 35% of text segments were identified to contain language supporting an AUD. Abbreviation: AUD, alcohol use disorder.

Table 4. Sociodemographic Information of Patients Identified with AUD, by ICD-9 Diagnosis Versus Model Identification (cBERT).

Population Characteristic	Patients with AUD by ICD-9 (N = 3322), N (%)	Patients with AUD by cBERT (N = 1693), N (%)	Patients without AUD by ICD-9 and cBERT (N = 36 088), N (%)	AUD by ICD9 vs AUD by cBERT X^2 (degrees of freedom*) *P*-value
Median age, in years [IQR]	53 [44, 62]	52.9 [37,63]	70 [49, 78]	>.05
Assigned sex				
Male	2538 (76)	1304 (77)	19 340 (54)	X^2 = 0.243 (1), >0.05
Female	784 (24)	389 (23)	16 748 (46)	
Language				
English	2143 (65)	925 (55)	17 987 (50)	X^2 = 266 (2), <.001
Spanish	87 (2.6)	45 (2.7)	642 (1.7)	
Other	1092 (33)	77 (4.5)	17 459 (48)	
Insurance				
Private	1259 (38)	680 (40)	13 401 (37)	X^2 = 10 (4), .04
Medicare	991 (30)	530 (31)	18 845 (52)	
Medicaid	716 (22)	1306 (18)	2639 (7.3)	
Government	229 (7)	106 (6)	917 (2.5)	
Self-pay	127 (3.8)	71 (4.2)	286 (0.7)	
Median hospital length of stay, in days [IQR]	7 [4, 14]	6 [3, 11]	7 [4, 12]	>.05
Substance use, by type				
Opioids/sedatives	238 (7.2)	217 (12.8)	273 (0.8)	**2 > 1, <.001**
Cannabis	253 (7.6)	141 (8.3)	141 (0.1)	>.05
Cocaine/amphetamines	61 (1.8)	21 (12)	45(0.2)	>.05
Other	88 (2.6)	69 (4.1)	83 (0.2)	**2 > 1, .001**
Mental health diagnosis				
Mood disorder	709 (21)	312 (18)	3456 (9.6)	**1 > 2, .02**
Delirium/dementia	173 (5.2)	82 (4.8)	2676 (7.4)	>.05
Anxiety	166 (5.0)	59 (3.5)	1244 (3.4)	>.05
Drug-induced mental disorder	105 (3.2)	72 (4.3)	341 (0.9)	>.05
Trauma/stressor-related	69 (2.1)	40 (2.4)	241 (0.7)	>.05
ADHD	35 (1.1)	24 (1.4)	114 (0.3)	>.05
Developmental disorder	29 (0.8)	13 (0.8)	323 (0.9)	>.05
Schizophrenia/psychosis	32 (1.0)	21 (1.2)	268 (0.7)	>.05

Contrasts: 1 = Patients with AUD by ICD-9; 2 = Patients with AUD by cBERT; disorder.
Abbreviations: IQR, interquartile range; ADHD, attention deficit hyperactivity; AUD, alcohol use disorder.
*If unlisted, degrees of freedom = 1.

AUD criteria without a structured diagnosis, as highlighted by the 1400+ patients identified in Experiment 2. While the medical community has engaged in several efforts to improve outpatient screening,[3,11,12] the 4% overall discrepancy between structured diagnoses and AUD in the medical narrative highlights clinician *under documentation*, as opposed to under recognition. In particular, physician annotation revealed several cases of documentation alluding to a new complication or chronic illness exacerbation related to AUD *without* implementation of a formal AUD diagnosis. At a system level, this could reflect provider hesitancy to levy an AUD diagnosis due to stigma, or limited clinician knowledge of AUD criteria. In response, a text-classification tool could facilitate improved documentation around AUD and better capture the epidemiology of AUD for both research and clinical use.

When compared to ICD-9 diagnoses, AUD patients identified by our model exhibited similar general demographics, with some differences in reported language proportions (Table 4). Yet, patients identified by our model had a higher proportion concurrent mood disorders, opioid/sedative use, and "other" drug use. This pattern raises a couple of potential explanations: (1) Physicians may document alcohol use, but are more likely to implement ICD-9 codes for other substances, or (2) our model may be conflating AUD with other substance use.

While our model performed well in its ability to correctly classify patients with AUD, adjudication results highlight that the model remains overly sensitive to certain language, including any substance use, drug withdrawal—not specific to alcohol, as well as chronic alcohol-related medical conditions, such as "cirrhosis," "liver failure," "lethargy," and "CIWA" (Clinical Institute Withdrawal Assessment). This sensitivity may also explain the above differences in substance use diagnoses in the model-identified population (Table 4). The model's over-emphasis on these terms could be explained by our study population, which represents older, hospitalized patients requiring an intensive level of care. The model also had difficulty distinguishing between a history of alcohol abuse and a current use disorder. This limitation potentially reflects a combination of outdated problem lists, under documentation of AUD resolution, and an unclear temporal relationship between a patient's AUD and their current illness.

The low IRR adjudication scores reflect difficulty in meeting AUD diagnostic criteria based on a single discharge summary (Table 2). The physician adjudicators included pediatricians with subspecialty training in emergency medicine and represented early, mid, and late-career experience. While all were familiar with the DSM-IV criteria, certain inferences had to be made about temporality and severity of use as it was documented in the history of present illness or hospital course. Review of false positives in the IRR sample (N = 7) revealed model conflation of phrases *including* the term alcohol that do not actually represent AUD. Examples included the following: "seen by social work for alcohol screen," "with alcohol intoxication," and "a toxic alcohol screen was checked [listing several volatile alcohols]," and "father with etoh cirrhosis."

We recognize that false identification could cause significant distress for both patient and providers. As such, future study involves model application to a broader set of patients, both in age and acuity; a more diverse training environment should enhance the language identified and utilized by the model and therefore improve our model's accuracy and precision.

As a retrospective evaluation of AUD, this study has several limitations. Geographically, the MIMIC-III database encompasses patients from a specific region of the United States (Boston, MA) and likely does not account for variations in substance use in other parts of the United States. As a study limited to ICU patients, the available text represents only those with the most extreme physical manifestations of disease; documentation and description of patients with AUD may differ among healthier and less symptomatic individuals. Finally, patient data collection extends only through 2012. Alcohol use patterns have significantly shifted in the context of the opioid epidemic as well as recreational marijuana legalization use since the publication of this database.[27] The medical community has also changed its understanding and documentation around social determinants of health,[28,29] both of which likely affect the prevalence and presentation of AUD. Model application on a more recent and inclusive dataset is required to examine the impact of sociodemographic data on AUD.

This study provides a pilot approach to improving identification of AUD in the medical record and lays the groundwork for a text-recognition model that can accurately and efficiently screen the medical record for *all* patients seeking medical care. A refined AUD-screening model would empower health officials to better allocate limited SUD screening and treatment resources to those at greatest risk.

Acknowledgments

Thank you to Dr. Janette Baird for her assistance in statistical analysis.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Dr Mercurio's involvement in this study was supported by NIH T32DA013911 from the National Institute on Drug Abuse. Dr. Ruest's involvement in this study was supported, in part, by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P20GM139664. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Compliance, Ethical Standards, and Ethical Approval

Institutional Review Board approval was not required.

ORCID iD

Laura Mercurio https://orcid.org/0000-0003-4346-8773

Supplemental Material

Supplemental material for this article is available online at the *SAJ* website http://journals.sagepub.com/doi/suppl/10.1177/29767342241288112.

References

1. National Institute on Drug Abuse and Alcoholism. Alcohol Use in the United States: Age Groups and Demographic Characteristics. *Alcohol Facts and Statistics*. 2022. Accessed December 14, 2023. https://www.niaaa.nih.gov/alcohols-effects-health/alcohol-topics/alcohol-facts-and-statistics/alcohol-use-united-states-age-groups-and-demographic-characteristics
2. Hingson RW, Zha W, White AM. Drinking beyond the binge threshold: predictors, consequences, and changes in the U.S. *Am J Prev Med*. 2017;52(6):717-727. doi:10.1016/j.amepre.2017.02.014
3. Roberts A, Rogers J, Mason R, et al. Alcohol and other substance use during the COVID-19 pandemic: a systematic review. *Drug Alcohol Depend*. 2021;229(Pt A):109150. doi:10.1016/j.drugalcdep.2021.109150
4. National Institute on Drug Abuse. Marijuana and hallucinogen use among young adults reached all time-high in 2021.

August 22, 2022. Accessed December 14, 2023. https://nida.nih.gov/news-events/news-releases/2022/08/marijuana-and-hallucinogen-use-among-young-adults-reached-all-time-high-in-2021
5. Berger D. Medical Complications: Common Alcohol-Related Concerns. *National Institute on Alcohol Abuse and Alcoholism: core resource on alcohol*. 2022. Accessed June 5, 2024. https://www.niaaa.nih.gov/health-professionals-communities/core-resource-on-alcohol/medical-complications-common-alcohol-related-concerns#pub-toc0
6. Nilsen P, Aalto M, Bendtsen P, Seppä K. Effectiveness of strategies to implement brief alcohol intervention in primary healthcare. A systematic review. *Scand J Prim Health Care*. 2006;24(1):5-15. doi:10.1080/02813430500475282
7. Edlund MJ, Unützer J, Wells KB. Clinician screening and treatment of alcohol, drug, and mental problems in primary care: results from healthcare for communities. *Med Care*. 2004;42(12):1158-1166. doi:10.1097/00005650-200412000-00002
8. Rehm J, Anderson P, Manthey J, et al. Alcohol use disorders in primary health care: what do we know and where do we go? *Alcohol Alcohol*. 2016;51(4):422-427. doi:10.1093/alcalc/agv127
9. Hallgren KA, Witwer E, West I, et al. Prevalence of documented alcohol and opioid use disorder diagnoses and treatments in a regional primary care practice-based research network. *J Subst Abuse Treat*. 2020;110:18-27. doi:10.1016/j.jsat.2019.11.008
10. Bahorik AL, Satre DD, Kline-Simon AH, Weisner CM, Campbell CI. Alcohol, cannabis, and opioid use disorders, and disease burden in an integrated health care system. *J Addict Med*. 2017;11(1):3-9. doi:10.1097/ADM.0000000000000260
11. John WS, Zhu H, Mannelli P, Schwartz RP, Subramaniam GA, Wu LT. Prevalence, patterns, and correlates of multiple substance use disorders among adult primary care patients. *Drug Alcohol Depend*. 2018;187:79-87. doi:10.1016/j.drugalcdep.2018.01.035
12. Wu LT, McNeely J, Subramaniam GA, et al. DSM-5 substance use disorders among adult primary care patients: results from a multisite study. *Drug Alcohol Depend*. 2017;179:42-46. doi:10.1016/j.drugalcdep.2017.05.048
13. Poulsen MN, Freda PJ, Troiani V, Davoudi A, Mowery DL. Classifying characteristics of opioid use disorder from hospital discharge summaries using natural language processing. *Front Public Health*. 2022;10:850619. doi:10.3389/fpubh.2022.850619
14. Tai B, McLellan AT. Integrating information on substance use disorders into electronic health record systems. *J Subst Abuse Treat*. 2012;43(1):12-19. doi:10.1016/j.jsat.2011.10.010
15. 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;264:1056-1060. doi:10.3233/SHTI190386
16. 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 Med Inform*. 2021;9(3):e23456. doi:10.2196/23456
17. Boscarino JA, Moorman AC, Rupp LB, et al. Comparison of ICD-9 codes for depression and alcohol misuse to survey instruments suggests these codes should be used with caution. *Dig Dis Sci*. 2017;62(10):2704-2712. doi:10.1007/s10620-017-4714-8
18. Carrell DS, Cronkite DJ, Shea M, et al. Clinical documentation of patient-reported medical cannabis use in primary care: toward scalable extraction using natural language processing methods. *Subst Abus*. 2022;43(1):917-924. doi:10.1080/08897077.2021.1986767
19. Wang Y, Chen ES, Pakhomov S, et al. Automated extraction of substance use information from clinical texts. *AMIA Annu Symp Proc*. 2015;2015:2121-2130. eCollection 2015.
20. Ni Y, Bachtel A, Nause K, Beal S. Automated detection of substance use information from electronic health records for a pediatric population. *J Am Med Inform Assoc*. 2021;28(10):2116-2127. doi:10.1093/jamia/ocab116
21. Alsentzer E, Murphy JR, Boag W, et al. Publicly available clinical BERT embeddings. *arXiv*. 2019. arXiv:1904.03323. Submitted on 6 Apr 2019 (v1), last revised 20 Jun 2019 (this version, v3). https://arxiv.org/abs/1904.03323
22. Huang K, Altosaar J, Ranganath R. Clinicalbert: modeling clinical notes and predicting hospital readmission. *arXiv*. arXiv:1904.05342. Accessed April 10, 2019. https://arxiv.org/abs/1904.05342
23. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association Publishing; 2022. doi:10.1176/appi.books.9780890425787
24. Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data*. 2016;3(1):160035. doi:10.1038/sdata.2016.35
25. Rehm J, Gmel G, Sempos CT, Trevisan M. Alcohol-related morbidity and mortality. *Alcohol Res Health*. 2003;27(1):39-51.
26. Carvalho AF, Heilig M, Perez A, Probst C, Rehm J. Alcohol use disorders. *Lancet*. 2019;394(10200):781-792. doi:10.1016/S0140-6736(19)31775-1
27. Witkiewitz K, Vowles KE. Alcohol and opioid use, co-use, and chronic pain in the context of the opioid epidemic: a critical review. *Alcohol Clin Exp Res*. 2018;42(3):478-488. doi:10.1111/acer.13594
28. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep*. 2014;129(Suppl 2):19-31. doi:10.1177/00333549141291S206
29. McCormack LA, Madlock-Brown C. Social determinant of health documentation trends and their association with emergency department admissions. *AMIA Annu Symp Proc*. 2020; 2020:823-832.