NATIONAL CENTER FOR HEALTH STATISTICS
White Paper

August 12, 2026

Using Clinical Notes and Medical Codes to Identify Stimulant Use in Hospital Data

Doreen M. Gidali, M.D./M.P.H., Nikki B. Adams, Ph.D., Minchan Shi, M.S., Dwayne L. Mitchell, Dr.PH.

U.S. CENTERS FOR DISEASE CONTROL AND PREVENTION

www.cdc.gov/nchs/products

Abstract

Objectives: This report documents the development of the Stimulant Algorithm by the National Center for Health Statistics (NCHS). The algorithm is designed to identify non-therapeutic stimulant use—including prescription misuse and illicit use—as well as instances of stimulant use co-occurring with opioid use (co-use), therapeutic stimulant use, and the presence of drug screening during a hospital encounter at any point in the patient's clinical history.

Methods: The Stimulant Algorithm was designed as a two-component algorithm. One component analyzes the clinical notes using natural language processing (NLP), including machine learning, while the other searches for medical codes related to stimulant or opioid use. The algorithm was designed using data from the 2020 National Hospital Care Survey (NHCS), which collects inpatient and emergency department (ED) encounter (or visit) data from a nationally representative sample of hospitals in the United States. Clinical reviewers annotated a subset of 2020 NHCS encounters to identify patients using stimulants, opioids, or both based on information found in the encounters. The annotated data were then used to develop and test the Stimulant Algorithm.

Results: The Stimulant Algorithm was evaluated using the Matthews Correlation Coefficient (MCC) and F1-score. It achieved an MCC of 0.95 and F1-score of 0.98 for identifying non-therapeutic stimulant use, MCC of 0.90 and F1-score of 0.92 for detecting non-therapeutic opioid use, and MCC of 0.94 and F1-score of 0.95 for recognizing the non-therapeutic co-occurrence of stimulant and opioid use (co-use) in hospital encounters.

Conclusion: The Stimulant Algorithm showed strong performance in identifying instances of non-therapeutic stimulant use, although prescription misuse—a subtype of non-therapeutic use—was more difficult to detect in the clinical notes. Because clinical data do not consistently distinguish between current and historical substance use, the algorithm captures any documented use within a patient's clinical history and may include past use when identifying cases.

Introduction

While death rates related to drug use in the United States have been decreasing, substance use remains a major public health concern. Overdoses are among the leading causes of injury-related deaths in the U.S. for ages 1-44 (1–3). Opioids have been the primary drugs involved. However, starting in 2011, stimulants (cocaine and the psychostimulants methamphetamine, methylphenidate and amphetamines) have played a growing role in what is now considered the fourth wave of the opioid epidemic—the opioid/stimulant wave—characterized by the use of either opioids, stimulants, or both together (4). The rate of overdose deaths involving psychostimulants was 14.3 times higher in 2021 than it was in 2011 (3,5). The rate of overdose deaths involving both stimulants and opioids has increased since 2017, underscoring the growing challenge of polysubstance use and the need for better surveillance and research tools (5).

Public health is increasingly leveraging ‘big data’ sources for more timely and detailed insights. In the clinical domain, medical claims data and electronic health records (EHR) are types of big data that support disease surveillance and research on health and other outcomes (6). Claims data, collected during patient visits primarily for billing, provide standardized information about healthcare utilization (7). EHRs are digital records of a person’s clinical profile and typically include both current and past medical conditions, treatments (including medications and immunizations), tests and procedures (7). EHR data can be divided into structured and unstructured forms. Structured data includes standardized, coded information—such as medical codes for diagnoses, medications or procedures—that can be easily organized, searched, and analyzed. In contrast, unstructured data consists of free-text entries written by healthcare providers, such as progress notes, surgical reports, or discharge summaries.

Substance use research using health data most commonly relies on structured EHR or claims to identify relevant patient records. These sources mostly contain medical codes, which can offer a broad overview but may not fully capture a patient’s clinical picture. Detailed aspects of care, including substance use history, social context, or behavioral indicators, are often further documented in free-text clinical notes. As a result, analyses based solely on coded data may lead to underreporting or misclassification of substance use-related encounters (8,9). Unstructured data, though more complex to work with, contains clinical detail and context, that can supplement coded data. Advanced methods such as natural language processing (NLP) and machine learning (ML) are required to extract information from text-based health records to provide a more comprehensive understanding of substance use and its associated health outcomes.

For example, in 2018, with support from the Office of the Secretary's Patient-Centered Outcomes Research Trust Fund (PCORTF), the National Center for Health Statistics (NCHS) developed an 'Opioid algorithm' using natural language processing to identify opioid-related hospital visits through clinical notes (10). When applied to 2016 data from the National Hospital Care Survey (NHCS), it identified 277,958 encounters not captured in structured data (the medical codes) demonstrating the value of unstructured text in tracking substance use (10).

Following this initial effort, NCHS received additional OS-PCORTF funding in 2019 to expand this work, validate the Opioid Algorithm, and develop a second algorithm that could identify co-occurring substance use disorders (SUD) and mental health issues (MHI) among patients with opioid-related encounters (11). As part of the validation process, clinicians independently reviewed and confirmed opioid involvement in a sample of EHR data with clinical notes detected by the Opioid Algorithm (12). These efforts showed that when used together, claims and both structured and unstructured EHR data are a rich source for studying substance use and related health outcomes.

Considering the evolving opioid-stimulant overdose crisis, and the continued pivot in public health to leverage 'big data', in 2023, OS-PCORTF sponsored NCHS to build on the progress of the two algorithms and develop a third one: the 'Stimulant Algorithm'. This one was intended to identify therapeutic stimulant use, non-therapeutic stimulant use—that is, illicit stimulant use (such as cocaine and methamphetamine) or misuse of prescription stimulants (like methylphenidate and amphetamine)—non-therapeutic co-use of stimulants and opioids, and drug screening. This report outlines the methodology used to develop the Stimulant Algorithm.

Background

The methodology for developing the Stimulant Algorithm was primarily adapted from the approaches used in creating the 'Opioid Algorithm' and the 'Co-occurring Disorders Algorithm (10-12). Insights and lessons learned from these prior developments provided a foundation for the Stimulant Algorithm's development work. The key difference was the shift in focus from opioids to stimulants. In addition to identifying stimulant use in hospital encounters, this algorithm would also be designed to distinguish between the types of stimulants used—whether prescription or illicit, and the manner of use, therapeutic or non-therapeutic (non-medical).

As with the previous efforts, a Technical Expert Panel played a critical role in shaping the development of the Stimulant Algorithm. The panel, composed of subject matter experts from the Centers for Disease Control and Prevention (CDC), the U.S. Food and Drug Administration (FDA), the Substance

Abuse and Mental Health Services Administration (SAMHSA), the National Institute on Drug Abuse (NIDA), and the National Center for Advancing Translational Sciences (NCATS), provided expert input on a range of methodological considerations, including case definitions, substance classification, and data science techniques.

All three algorithms were developed using data from the National Hospital Care Survey (NHCS)—2016 NHCS for the Opioid and Co-occurring Disorders Algorithms, and 2020 NHCS for the Stimulant Algorithm.

Methods

The Stimulant Algorithm consists of two components. A medical code component that uses structured (medical codes) data and a clinical note component that uses unstructured (clinical notes) data to identify stimulant-related outcomes of interest. The data source, case definitions, and methods for developing each component, and approaches used to evaluate them are described in the sections below.

Data Source

The NHCS is an establishment survey designed to produce national estimates of the characteristics of inpatient hospitalizations and emergency department (ED) visits, including, but not limited to, inpatient length of stay, diagnoses, and surgical and nonsurgical procedures (13). NHCS includes noninstitutional, non-federal hospitals in the 50 U.S. states and the District of Columbia, with six or more staffed inpatient beds.

Participating hospitals submit data to NHCS in one of two formats: Uniform Billing–04 (UB-04) administrative claims or electronic health records (EHR). In addition to direct submissions, NHCS also receives EHR data from the American College of Emergency Physicians (ACEP) for sampled EDs that did not respond directly to NHCS but participate in ACEP's data registry (14).

Among the data NHCS receives, only EHR submissions include clinical notes, but submitting notes is optional. As a result, only some hospitals that submit EHR data include free-text clinical notes. Conversely, data from all participating hospitals, regardless of the format in which the data is submitted, contains medical codes.

The 2020 NHCS sample included 608 hospitals across the United States. Of these, 205 submitted data (a response rate of 33.7%), resulting in a total of 10,010,479 encounters (i.e., inpatient hospitalizations or emergency department visits) (15). Among them, 137 hospitals submitted UB-04 administrative claims data, and 68 submitted EHR data. Of the 68 EHR hospitals, 25 included free-text clinical notes.

Of these, five had significant formatting issues or errors in transmission that rendered their notes unusable. The remaining 20 hospitals submitted notes of sufficient quality and detail to be considered meaningful, accounting for 888,556 encounters (8.9% of the entire dataset).

The Stimulant Algorithm was developed using a subset of these data, including encounters with meaningful clinical notes. Additional details on the data used for algorithm development and testing are provided in the section describing the algorithm methodology for the coded data and the clinical notes.

Case Definitions

Case definitions are an important aspect of algorithm development because they specify what the algorithm aims to detect. NCHS worked with the Technical Expert Panel to identify the data elements necessary to classify hospital encounters as positive for the following outcomes (defined below): therapeutic use, non-therapeutic use (including illicit use, prescription misuse, unspecified use, or co-use of stimulants and opioids), and drug screening. The methods for defining cases and the resulting definitions are explained below.

Considerations and General Definitions

Temporality: The nature of electronic health record (EHR) data makes it difficult to consistently determine whether a diagnosis in a patient's EHR is current (active) or historic (resolved) (16). Some health records clarify this distinction, but many do not, and the medical code fields are not always updated to reflect temporality (17). For example, a diagnosis code may remain marked as active long after a condition has resolved or may lack any indication of its status. To avoid misclassification, the project case definitions reflect substance use at any point in the patient's clinical history, as reflected in the 2020 NHCS data.

Manner or type of use: The terms used to describe substance use are often used interchangeably and can have different meanings in different settings. For clarity, the following section explains NCHS's definition of the following types of use specifically for this project. Moving forward, the term 'use' without modification, globally refers to any of the following types of use:

1. **Therapeutic use** refers to the use of a prescribed stimulant or opioid that is documented as medically indicated and taken according to clinical guidance (for example, opioids for pain management, or stimulants for attention deficit hyperactivity disorder).
2. **Non-therapeutic use**: Use of a substance for non-medical reasons (or recreational) and includes prescription misuse or illicit use as described below.

- **Prescription misuse** refers to:
  - Use of prescription stimulant (e.g., Adderall) or opioid (e.g., oxycodone) in a manner not directed by a doctor, such as diverted prescriptions (using someone else's prescription), use in greater amounts, more often, or longer than prescribed (18).
- **Illicit use** refers to**:**
  - Use of Schedule I stimulants or opioids —drugs with a high potential for abuse and no accepted medical use (for example, ecstasy, heroin) or
  - Use of illegally or illicitly manufactured Schedule II stimulants or opioids — substances with approved medical use but also a high potential for abuse (for example, methamphetamine or fentanyl) (19).
- **A note on prescription misuse and illicit use:** While methamphetamine, cocaine, and fentanyl all have clinical (therapeutic) use, they are rarely prescribed on an outpatient basis. Cocaine is typically administered by healthcare providers during medical procedures (20). Only one methamphetamine-containing medication, Desoxyn, is available by prescription, but its use is extremely limited—fewer than 1,800 prescriptions were dispensed in the U.S. in 2024, with no instances found in the NHCS 2020 data (21). Fentanyl, when prescribed for outpatient treatment, is usually dispensed as transdermal patches and restricted to a small group of patients under strict monitoring protocols (22,23). Consequently, for this project, any indication of non-therapeutic use of methamphetamine, fentanyl, or cocaine was classified as illicit use, unless explicitly stated otherwise in the clinical record.

  While misuse of prescribed drugs like Percocet or Adderall is illegal, here they were classified as prescription misuse to distinguish from use of illegally manufactured substances, which are key drivers of the overdose crisis (4,24,25).
- **Unspecified use** refers to non-therapeutic use of stimulants or opioids without evidence of a specific manner of use, such as 'illicit use' or 'prescription misuse.'
- **Co-use** refers to the non-therapeutic use of stimulants and opioids at the same time. For this project, it serves as a proxy for simultaneous use.

3. **Any use** incorporates all the types above, including instances where it is unclear whether the manner of use is therapeutic or non-therapeutic.

Defining Project Cases

Case definitions for the algorithms were developed through a three-phased approach in collaboration with the technical expert panel.

Phase 1: An exploratory analysis of 2020 NHCS data was conducted to verify whether the necessary information to meet the project objectives was available. Coded fields were examined for various classes of stimulant medications and related disorders. Clinicians on the team performed a cursory annotation of a small sample of stimulant or opioid-related records to identify potential terms for case detection and assess feasibility. The exploratory findings guided drug classifications, inclusion and exclusion criteria for specific medications, and patterns in physician documentation of substance use. The stimulant medication categories and the rationale for inclusion and exclusion are shown in **Tables A** and **B** respectively**.**

Table A. List of psychostimulant categories and the rationale for inclusion in designing the Stimulant Algorithm

Category	Drug/Class	Examples (Brand/Common Names)	Rationale for Inclusion
Prescription stimulants	Cocaine hydrochloride	Goprelto®, Numbrino®	Clinical use in select surgical specialties Rarely prescribed for outpatient use
	Amphetamine derivatives (amphetamine, dextroamphetamine, lisdexamfetamine)	Adderall®, Dynavel®, Zenzedi®, Vyvanse®	High abuse potential (Schedule II) Commonly prescribed for ADHD Widely used across clinical settings
	Methylphenidate	Ritalin®, Concerta®, Focalin®	
Illicit stimulants	Cocaine	…	High abuse potential (Schedule II) Common illicit use
	Methamphetamine	…	High abuse potential (Schedule I) Illegally manufactured forms common
	3,4-methylenedioxymethamphetamine	MDMA, Ecstasy	Illicit stimulant (Schedule I) Associated with recreational use

®Prescription brand medication
NOTES: Stimulants are a class of drugs that increase the activity of the central nervous system. Schedule I are drugs with high abuse potential and no current accepted medical use in the US. Schedule II are drugs with high abuse potential with accepted medical use in the US. NHCS is the National Hospital Care Survey. ADHD is attention deficit and hyperactivity disorder.

Table B. List of psychostimulant categories and the rationale for exclusion in designing the Stimulant Algorithm

Category	Drug/Class	Examples (Brand/Common Names)	Rationale for Exclusion
Excluded from the Stimulant Algorithm			
Prescription stimulants	Methamphetamine hydrochloride	Desoxyn®	Negligible frequency in 2020 NHCS
	Methamnetamine	…	Rarely prescribed in the U.S Negligible frequency in NHCS
	Modafinil	Provigil®	Low abuse potential Low frequency in NHCS
	Cathinone derivatives	Bupropion (Wellbutrin®)	Low abuse potential
	Anorexiants/stimulants	Adipex-P®, Lomaira®	Negligible frequency in NHCS
Illicit stimulants	Cathinones	"Bath salts"	Negligible frequency in NHCS
Other stimulants	Ephedrine, pseudoephedrine	…	Negligible frequency in NHCS
	Caffeine	…	Difficult to distinguish healthy vs. unhealthy use

®Prescription brand medication
NOTES: Stimulants are a class of drugs that increase the activity of the central nervous system. Schedule I are drugs with high abuse potential and no current accepted medical use in the US. Schedule II are drugs with high abuse potential with accepted medical use in the US. NHCS is the National Hospital Care Survey. Negligible frequency refers to a frequency of less than 20. ADHD is attention deficit and hyperactivity disorder.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020

Phase 2: A literature review was conducted to gather information on current methods for defining substance use visits relevant clinical terms. A high-level review of medical codes was also performed to confirm that codes existed to measure the project objectives.

Phase 3: Phases 1 and 2 generated an initial list of medical codes (for medications, diagnoses, procedures, services, and laboratory tests) and search terms (keywords or supporting language in clinical notes) as potential indicators of stimulant or opioid use. The medical codes and search terms were refined before being used to define cases; they are discussed further in the algorithm methodology sections. The final case definitions are presented below.

Illicit Stimulant or opioid use

An encounter was classified as positive for illicit stimulant use or illicit opioid use if at least one of the following criteria was met in the encounter record(s).

1. The presence of a diagnostic, procedure, or service code for illicit stimulant or opioid use or related disorders in the coded data fields or
2. Explicit mention or evidence of illicit stimulant or opioid use in the clinical notes (text fields). Here, an example of an explicit mention would be when the healthcare provider notes the patient's substance use or substance use disorder.

Prescription Misuse of Stimulants or Opioids

A hospital encounter was classified as positive for misuse of a prescription stimulant or opioid if at least one of the following criteria was true in the encounter record(s).

1. Presence of a diagnostic code indicating prescription misuse. For example, a diagnosis code for poisoning by methylphenidate or Vicodin®.
2. Explicit mention or evidence of prescription misuse in the clinical notes (text fields). For example, a provider stating that a patient misuses Adderall®.

Non-therapeutic Stimulant and Opioid Co-use

The presence of both opioid and stimulant non-therapeutic use in the same encounter served as a proxy for the co-use of stimulants and opioids. To meet the classification of stimulant and opioid co-use, both of the following had to be true within the encounter record(s):

1. Meet the definition for either prescription stimulant misuse or illicit stimulant use; and
2. Meet the definition for either prescription opioid misuse or illicit opioid use as defined above.

Therapeutic Stimulant Use

An encounter was classified as positive for therapeutic stimulant use if at least one of the following criteria was met in the encounter record(s).

1. The presence of a medication, diagnosis, procedure, or service code for therapeutic stimulants in the coded data fields or
2. Explicit mention or evidence of therapeutic stimulant use in the clinical notes (text fields).

Drug screening

An encounter was classified as positive for having a drug screen performed if at least one of the following criteria was met in the encounter record(s).

1. The presence of a laboratory, procedure, or service code for drug screening in the coded data fields or
2. Explicit mention or evidence of drug screening in the clinical notes (text fields).

Medical Code Component

The Stimulant Algorithm includes a medical code component, based on structured medical codes (the code component), and another based on clinical notes. The methodology for the code component is outlined below; the clinical note component is discussed separately.

To develop the code component, the project team adapted the approach used in the Opioid and Co-occurring Disorders algorithms, with modifications to reflect the focus on stimulant use and manner of

use. This multi-phased process began with selecting the appropriate medical code systems (detailed in the **Appendix**), followed by identifying the most relevant diagnosis, procedure, service, laboratory, and medication codes to use as indicators for the outcomes of interest.

Selecting Medical Codes

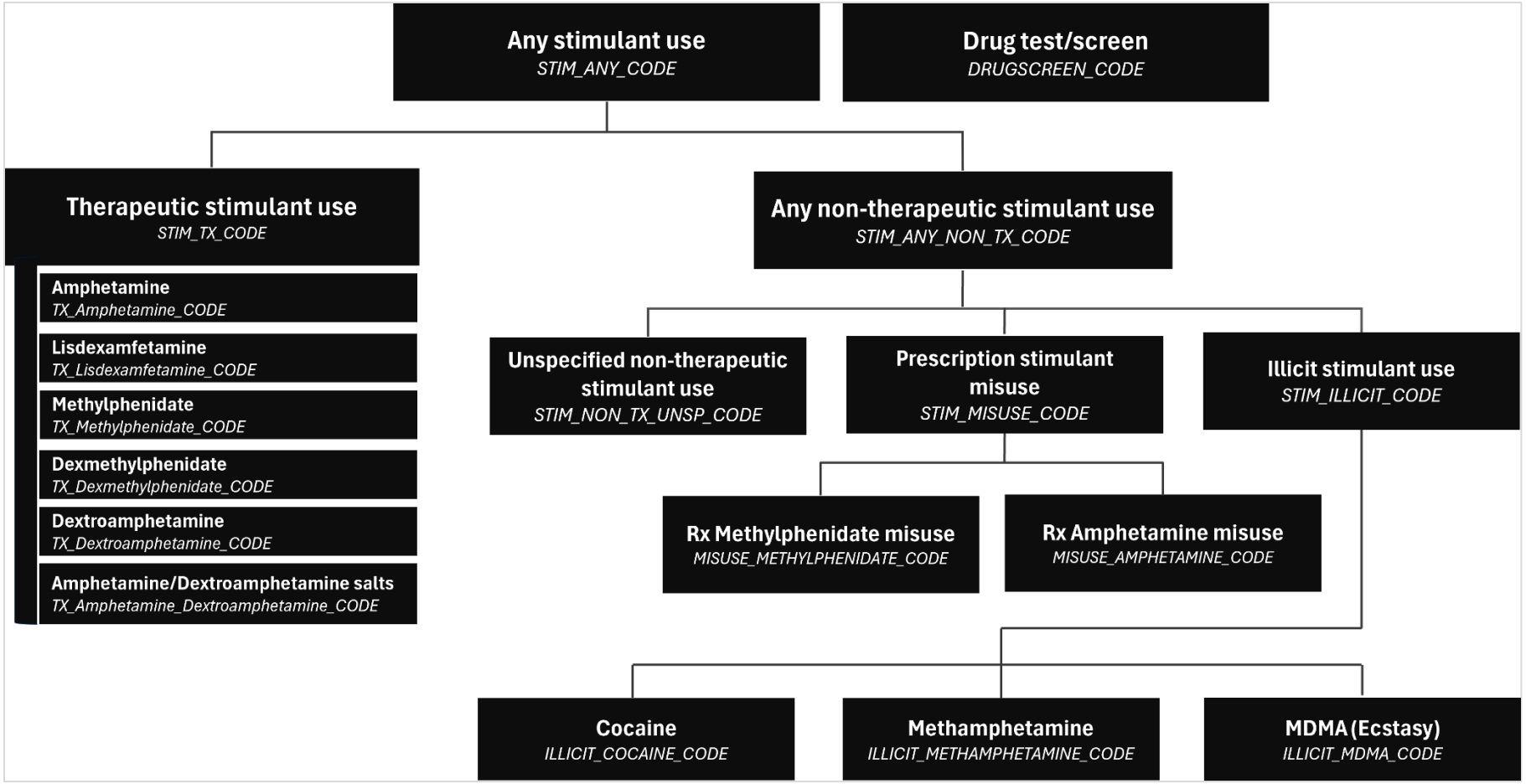
The two previous NCHS substance-use algorithms mainly focused on opioids and substance use disorders, with only limited inclusion of stimulant use (10,11). In comparison, the Stimulant Algorithm was designed to identify both the specific substance and the method of use, including distinguishing prescription misuse from illicit drug use, which required a separate set of medical codes.

Team clinicians conducted a detailed review of ICD-10-CM guidelines to identify codes that best reflected the project case definitions. While the opioid-related codes were primarily adapted from earlier algorithms, developing the stimulant code list required a targeted search. Clinicians compiled a list of stimulant codes and reviewed and refined the opioid codes, resulting in a preliminary list that met the project's specifications. The list was then reviewed by the team and vetted by the technical expert panel. Throughout the project, the list of codes was reviewed and revised to ensure alignment with case definitions.

Mapping Medical Codes to Algorithm Outcome Variables

Based on the project's case definitions for the outcomes of interest, the team developed a set of outcome variables —binary indicators used to determine whether a hospital record meets the criteria for therapeutic use, non-therapeutic use (including illicit use, prescription misuse, unspecified use, or non-therapeutic co-use), or drug screening. **Figure 1** provides a detailed breakdown of the outcome variables for the Stimulant Algorithm. Using a systematic approach led by the project's clinicians, each selected medical code was mapped to its corresponding outcome variable(s). This mapping served as the foundation for building the code-based algorithm.

Figure 1. Stimulant use output variables for the Code Component of the Stimulant Algorithm



NOTES: STIM is stimulant, NON_TX is non-therapeutic use (illicit or prescription misuse), TX is therapeutic (as prescribed) use, NON_TX_UNSP is non-therapeutic, unspecified use, Rx is prescription, MDMA is 3,4-methylenedioxymethamphetamine. The outcome variable for stimulant and opioid non-therapeutic co-use is not shown, as it is derived from using the non-therapeutic opioid and non-therapeutic stimulant use variables from both the code and clinical notes component. The complete codebook for the Stimulant Algorithm outcome variables, which includes the opioid outcome variables, can be found here: https://github.com/CDCgov/stimulant_opioid_algorithm_medical_codes_R/tree/main

Mapping diagnoses, procedures, services, and laboratory codes to the outcomes

Diagnosis codes were the primary data element used to identify non-therapeutic use—such as substance use, dependence, poisoning, and neonatal exposure codes—as well as therapeutic use, including underdosing and adverse reaction codes.

In addition to diagnosis codes, procedure and service codes were used to identify multiple outcomes of interest, depending on the code. For example, codes for provider-administered stimulant medications indicated therapeutic stimulant use, while codes for medication-assisted treatment for substance use disorders served as indicators of non-therapeutic use.

Drug screening was also an outcome of interest. Team clinicians classified toxicology tests (e.g., urine or blood) and verbal or written screening tools, such as questionnaires, as screening methods. The corresponding procedure, service, or laboratory codes were used to indicate drug screening. This indicator reflected only whether a drug screen was performed during the encounter, not the result of the screen.

Figure 2 provides a high-level overview of how the medical codes were mapped to each outcome of interest. The specific non-proprietary medical codes used in the algorithm, along with their corresponding stimulant use outcomes, can be found in the CDC GitHub repository for the medical code component of the Stimulant Algorithm. The repository, which includes a ReadMe file describing how medical codes are mapped to the outcomes is available at: https://github.com/CDCgov/stimulant_opioid_algorithm_medical_codes_R/tree/main

Mapping medication codes to the outcomes

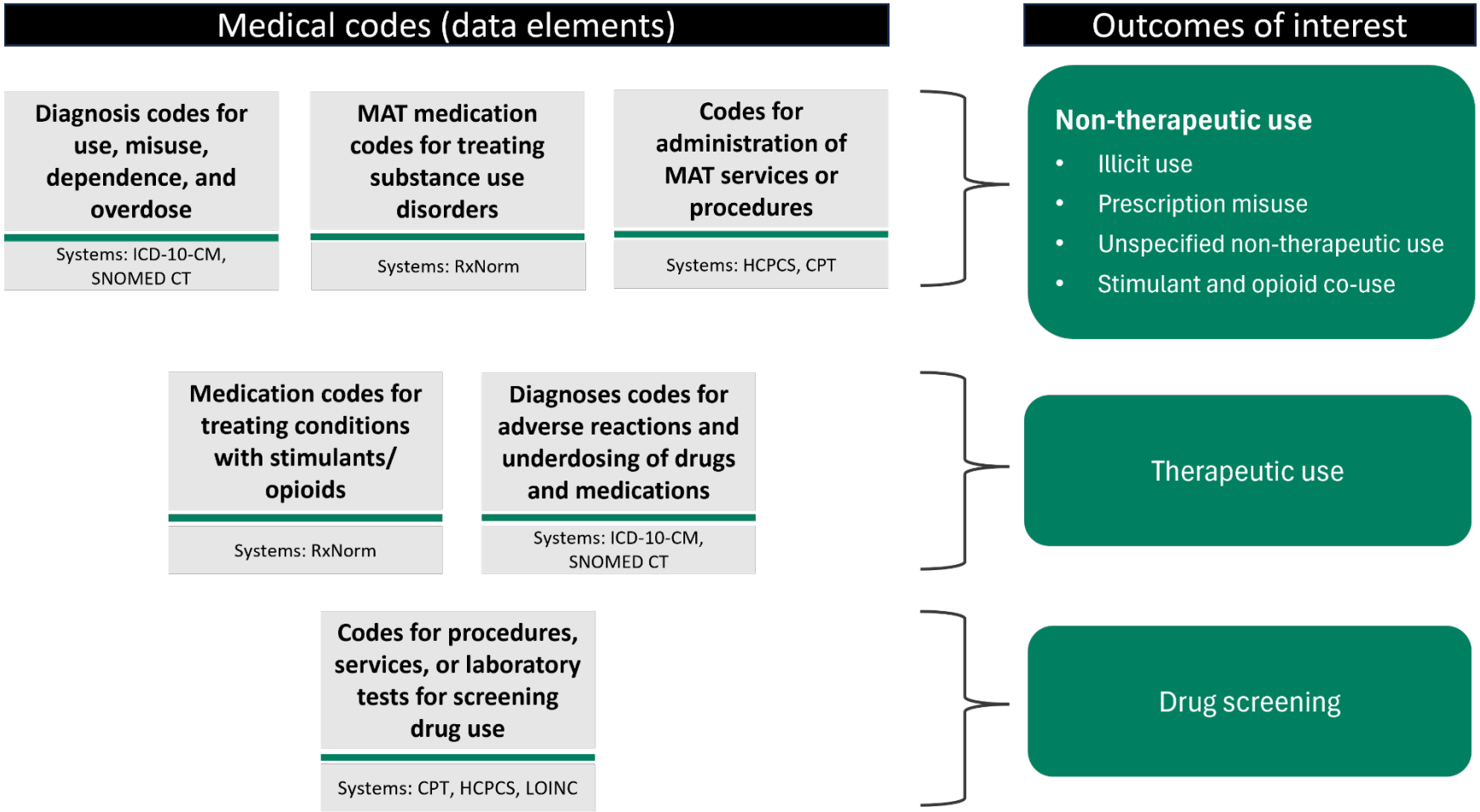
The Stimulant Algorithm development included identifying medications in the 2020 NHCS using a list of RxNorm codes for prescription stimulants and opioids. All stimulant medication codes flagged therapeutic use. While opioid medications where the main active ingredient was methadone or buprenorphine have previously been used for analgesia, their current primary indication is for the treatment of opioid use disorder or dependence (26,27). Therefore, for this project, methadone, buprenorphine, and their analogs served as indicators for non-therapeutic opioid use, which was important for flagging one of the outcomes, non-therapeutic co-use of stimulants and opioids.

The opioid antagonists naloxone (also known as Narcan®) and naltrexone (also known as Vivitrol® or Revia®) were excluded as indicators for the Stimulant Algorithm. Naltrexone, while used to treat opioid dependence, is also prescribed for alcohol use disorder (28). The indication could not be reliably determined from available data. To avoid potential misclassification, it was excluded.

Naloxone, used in both opioid overdose treatment and prevention, is frequently prescribed prophylactically to patients on opioids or other high-risk medications (29). However, a prescription alone does not confirm non-therapeutic use. The available codes did not allow the team to determine whether naloxone was administered during an overdose or prescribed as a preventive measure. As such, to avoid misclassification, naloxone was also excluded.

The medication codes used in the algorithm, along with their corresponding stimulant use outcomes, can be found in the CDC GitHub repository available at: https://github.com/CDCgov/stimulant_opioid_algorithm_medical_codes_R/tree/main.

Figure 2. Mapping medical codes to outcomes of interest for the Stimulant Algorithm



NOTES: MAT is Medication Assisted Therapy for substance use, ICD-10-CM is International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), SNOMED-CT is Systematized Nomenclature of Medicine--Clinical Terms, HCPCS is Healthcare Common Procedure Coding System, CPT is Common Procedure Terminology, LOINC is Logical Observation Identifiers Names and Codes, RxNorm is a standardized method of naming prescription and over the counter drugs, and part of the Unified Medical Language System. The **Appendix** describes the medical code systems in more detail.

Designing the Algorithm to Search Coded Fields

The code component was designed to analyze structured hospital data and generate an output file containing outcome variables corresponding to the outcomes of interest—therapeutic use, non-therapeutic use, and drug screening (30). As part of the code component, NCHS developed a dataset of selected medical codes and their associated code systems (**Appendix**), mapped to the outcome variable(s) each code indicates. The medical code algorithm scans coded fields in the medical record for the code system and the actual code found in its mapping dataset and then flags records as positive for the outcome(s) when a match is found.

As with any real-world data, some of the coded fields in the NHCS 2020 were missing codes. All coded fields in the data have a corresponding text field with a descriptive label (a character string that describes the medical code). To account for the absence of medical codes, a text-based search was added to the code component to search the descriptive code label. For example, if a record was missing a medication code, but the medication name (the descriptive label) was available, searching the label allowed this case to be flagged where relevant. This text-based search for key terms is conducted on the labeled fields for medications, procedures, services, and laboratory tests.

The code component was iteratively run on the NHCS 2020 coded data and their corresponding labels, with necessary revisions made until it functioned as intended. For a clearer understanding of what the algorithm was searching, a more detailed explanation of the NHCS data structure is provided in the **Appendix**.

Clinical note component: Natural Language Processing and Machine Learning

The clinical note component is designed to extract information from the clinical notes associated with a hospital encounter using natural language processing techniques. This process involved multiple steps, some of which occurred concurrently. First, a dataset was created for development and testing through a process called annotation. The annotated dataset was then used to develop both a rule-based NLP algorithm and a machine learning model. The following sections provide a detailed overview of each step.

Annotation

For classification algorithms, ground truth data—information verified by direct observation or measurement—is essential for guiding development and testing performance (31,32). This data, known as “gold standard data,” is collected through “annotation,” where subject matter experts, in this case,

the team clinicians, review the data and record relevant information. The annotation process involved team clinicians using a custom-built electronic form designed to help them identify information in a subset of the data that met project case definitions. The following questions were included in the form:

1. Was a drug screen performed?
2. What drugs were used and in what manner? Specifically, was there:
    - Illicit stimulant or opioid use?
    - Misuse of prescription stimulants or opioids?
    - Therapeutic use of stimulants?
3. For each type of drug use:
    - What was the specific substance seen? (“unspecified” was an allowed answer)
    - Where was this information found (for example, lab results, diagnoses, clinical note text)?

Annotation Sample

To create a gold standard dataset against which both the medical code component and clinical notes component could be evaluated, the annotation sample included only encounters with clinical notes since all hospital encounters have coded data. Data exploration revealed that a random sample would not yield enough positive cases for many outcomes of interest due to their low frequency. Therefore, key terms related to stimulant and opioid use were used to select a sample that captured relevant encounters for annotation.

To better identify non-therapeutic use, encounters were oversampled based on key terms and their context within clinical notes. For directly collected EHR (or Direct EHR notes), which include note type metadata, encounters were selected when key terms appeared in specific note types (e.g., “Adderall” in a “Social History” note). For ACEP notes, which lack note type metadata, encounters were selected based on the co-occurrence of specific phrases within the same note (e.g., “social hx” and “amphetamines”). Encounters with both stimulant and opioid key terms or combined phrases were also oversampled to capture co-use. Additionally, a subset of encounters without key terms was included as negative examples. The **Appendix** further explains the differences between Direct EHR and ACEP clinical notes.

The final 1,000 encounter sample was proportionally representative of the 2020 NHCS by source, with encounters from Direct EHR making up 69.8% of the annotation sample. Most encounters (90.0%) contained stimulant terms, 45.0% contained both stimulant and opioid terms, and 10.0% contained

neither stimulant nor opioid terms. **Table C** provides additional details on the characteristics of the annotation sample.

Table C. Composition of the annotation sample by data source

Sample characteristic	Direct EHR*	ACEP EHR†	Total
Stimulant term only, targeted note type	188	82	270
Stimulant term only, non-targeted note type	126	54	180
Stimulant and opioid terms, targeted note type	188	82	270
Stimulant and opioid terms, non-targeted note type	126	54	180
Neither stimulant nor opioid terms	70	30	100

*Direct EHR represents electronic health records submitted to NCHS directly from the hospital.
†ACEP EHR represents electronic health records with clinical notes submitted from the American College of Emergency Physicians.
NOTE: EHR is electronic health record.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020

Gold Standard Dataset

An NCHS NLP expert led team clinicians in annotating and creating a gold standard dataset for training, developing, and testing the Stimulant Algorithm. Annotators were trained on how to conduct annotation, and pre-tests were conducted to refine the case definitions and address any differences in interpretation between the clinicians. During pre-tests, clinicians independently annotated the same encounters, and inter-rater reliability—measuring agreement between two annotators —was calculated. After two rounds of pre-tests, the annotators agreed at least 90% of the time on each top-level question regarding the type and manner of stimulant or opioid use. Any discrepancies were discussed and resolved, with the annotation guide revised concurrently as clinicians developed it throughout the process.

Team clinicians were then assigned equal batches of different encounters to annotate separately, using the co-developed guide, resulting in a gold standard dataset divided into two parts. One part was used for development *(n*=804), and the other served as a test set (*n*=196) to evaluate the algorithm's performance against the clinician's assessment of the same data.

Building the Clinical note component

The Natural Language Processing (NLP) algorithm uses a combination of researcher-defined rules and machine learning techniques to identify relevant outcomes related to stimulant use. It analyzes clinical notes for specific drug terms and evaluates the context in which these terms are mentioned to determine whether the note indicates a positive outcome, as defined by the project case definitions.

“Context” refers to the setting (for example, note type or the location in the note), clinical circumstance, or surrounding text in which the term being searched is located. A non-therapeutic context was one in which the mentioned drug was most likely being used non-therapeutically (for non-medical reasons). Examples of non-therapeutic contexts in the note include a “social history” context, where substance use history is often documented, or a “problems” context, where diagnoses (including substance use disorders) are typically reported. In contrast, drug terms found in contexts such as "medications" or "allergies" would not fall under a non-therapeutic context.

The following sections explain the methods behind the rules-based approach and machine learning model used in developing the Clinical note component of the Stimulant Algorithm.

Logical Rules

Drug terms in the list of key terms discussed under Case Definitions were categorized by the substance name and type (prescription, illicit, unspecified). Then, depending on the context in which the drug term was found in the note, the combination was mapped to a specific outcome and the corresponding outcome variable. For example, the drug term “Vyvanse”, a stimulant, would be categorized as the generic prescription compound ‘Lisdexamfetamine.’ Based on these rules, if ‘Vyvanse’ appeared in a clinical note in a non-therapeutic context, this would be a positive indicator of the outcome of non-therapeutic stimulant use, specifically, prescription misuse of lisdexamfetamine.

Similarly, if the term 'Vyvanse' appeared in a context that was not a non-therapeutic context, but was relevant for identifying other outcomes, such as 'medications' or 'treatment', then this would be a positive indicator of therapeutic use.

Other contexts where drug mentions were considered outside of non-therapeutic use but were relevant for identifying other outcomes included 'laboratory or diagnostic orders' and 'results,' which would indicate drug screening. Additionally, contexts where drug terms were mentioned but not related to drug use were also crucial for reducing false positives by the algorithm. For instance, mentions of drugs in the context of “stimulant use risk score” were regarded as outside a non-therapeutic context or any other context relevant to the outcomes of interest.

Some phrases were identified as being strongly indicative of a specific outcome during exploration and were therefore mapped to an “override” category. This meant that these phrases flagged their associated outcome variable without needing to determine the non-therapeutic context. For instance, the phrase “opioid-seeking” was directly mapped to the opioid misuse variable without first determining the non-therapeutic context. **Table D** summarizes the logic rules and the corresponding outcomes; combinations that did not map to a particular outcome variable and are indicated with “N/A”.

Table D. Outcomes of interest determined by algorithm category and non-therapeutic context for the clinical notes analysis

Algorithm category	Non-therapeutic context*	Outcome
Stimulant		
Cocaine	Yes	Illicit cocaine use
Cocaine	No	N/A
Rx_Cocaine	Yes	Prescription stimulant misuse
Rx_Cocaine	No	Therapeutic stimulant use
Methamphetamine	Yes	Illicit methamphetamine use
Methamphetamine	No	N/A
MDMA	Yes	Illicit MDMA use
MDMA	No	N/A
Rx_Amphetamine	Yes	Prescription amphetamine misuse
Rx_Amphetamine	No	Therapeutic stimulant use
Lisdexamfetamine	Yes	Prescription lisdexamfetamine misuse
Lisdexamfetamine	No	Therapeutic stimulant use
Methylphenidate	Yes	Prescription methylphenidate misuse
Methylphenidate	No	Therapeutic stimulant use
Unspecified_Stimulant	Yes	Unspecified non-therapeutic stimulant use
Unspecified_Stimulant	No	Therapeutic stimulant use
Opioid		
Illicit_Opioid	Yes	Illicit opioid use
Illicit_Opioid	No	Any opioid use
Rx_Opioid	Yes	Prescription opioid misuse
Rx_Opioid	No	Any opioid use
Unspecified_Opioid	Yes	Unspecified non-therapeutic opioid use
Unspecified_Opioid	No	Any opioid use

*Context within which the drug or use was mentioned as either illicit drug use and/or misuse of prescription drugs.
NOTES: Rx is prescription, and MDMA is 3,4-Methylenedioxyamphetamine also known as ecstasy, N/A cases where a combination of category and context did not map to a particular outcome variable.

Applying the logical rules

The clinical note component methodology hinges on the determination of whether a term in the note is found within a non-therapeutic context. The presence of metadata optimizes this rule-based approach. In the case of direct EHR, where informative metadata was available, this rules-based approach proved sufficient; however, due to the lack of metadata in the ACEP data source, an alternative method was needed to assess the clinical context of ACEP notes. An illustration of the difference between Direct EHR and ACEP clinical notes in NHCS data, highlighting why a different approach was necessary to analyze the ACEP clinical notes is included in the **Appendix**.

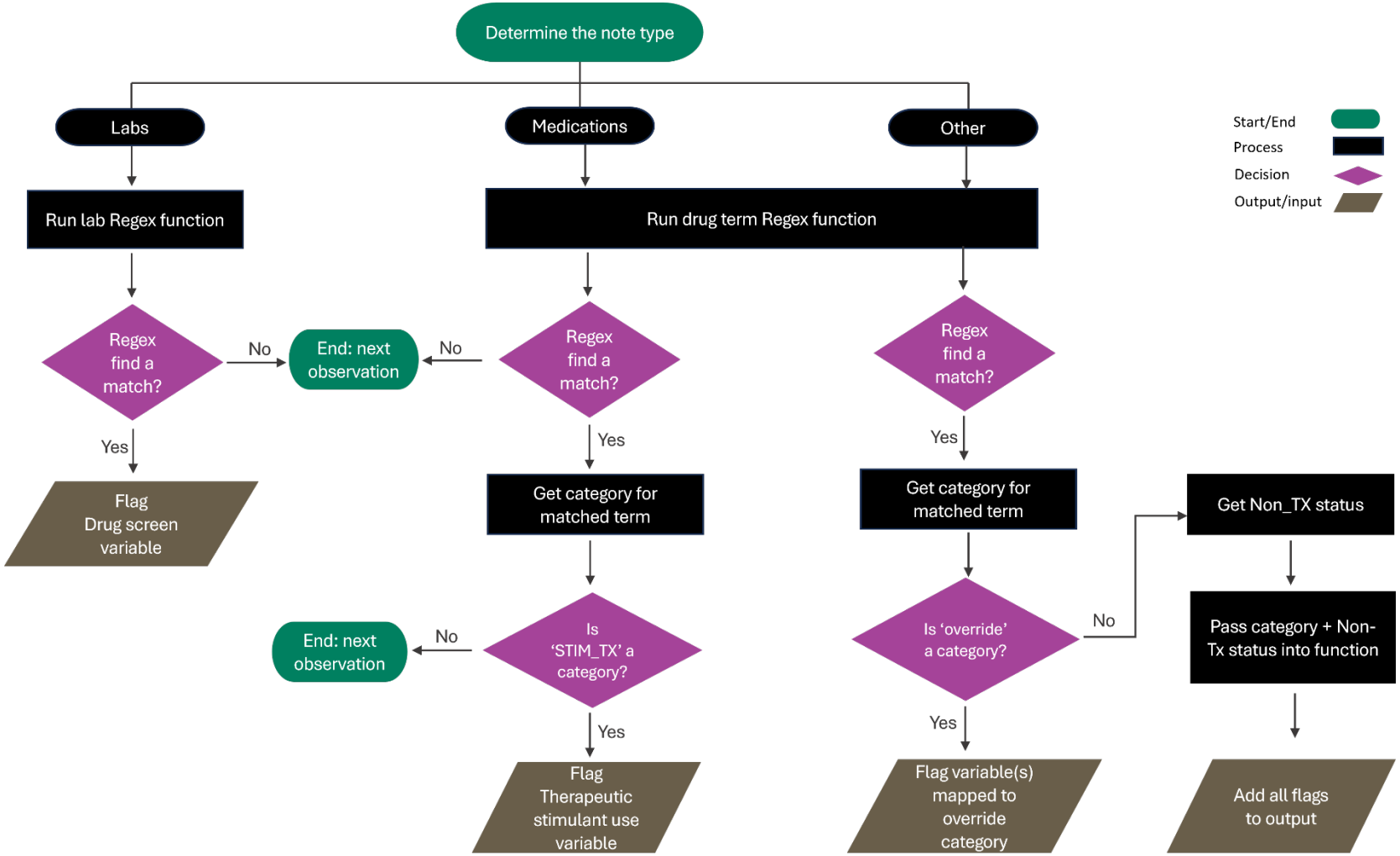
In NHCS data, a “record” in the clinical notes corresponds to one row in the database, typically representing a subsection of the clinical note. Each record can contain character strings ranging from a

single word to thousands of words. For Direct EHR clinical notes, a single record corresponds to a specific note type, such as “social history,” which indicates the context. In contrast, ACEP note records may include multiple note types or contexts within the same record; for instance, therapeutic (medications) and non-therapeutic (problem lists) contexts can coexist in one single note record. For this reason, a machine learning component was incorporated to identify whether a specific drug was mentioned in a non-therapeutic context for ACEP notes.

Logical Rules for Clinical Notes with Metadata

Because Direct EHR notes contained metadata in the form of note types, determining the clinical context of the note was straightforward, using the logic rules above. The combination of a drug term’s type (illicit, prescription, or unspecified) and its context (non-therapeutic or otherwise) would result in the appropriate flag. These steps are outlined in **Figure 3.**

Figure 3. Natural language processing flow chart for analyzing clinical notes from the directly collected electronic health records.



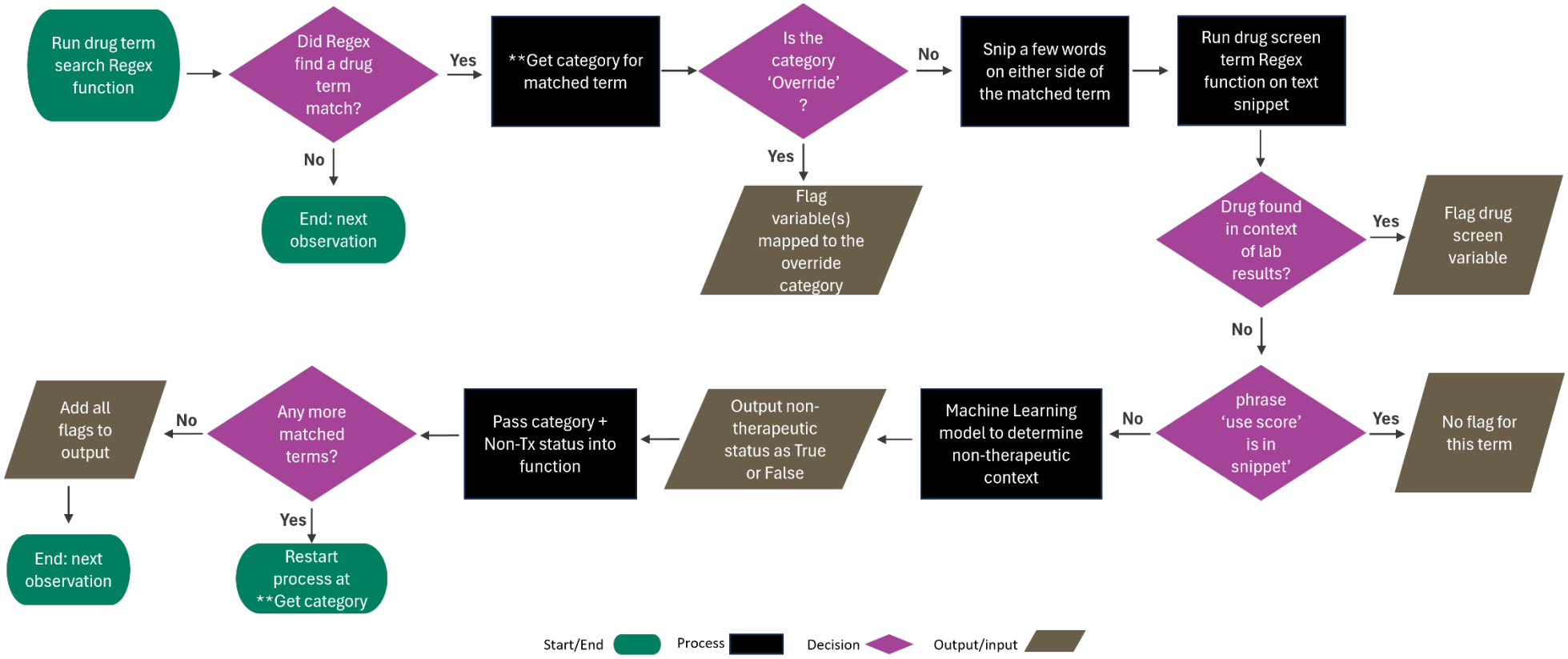
NOTES: Regex is regular expression, STIM is stimulant, TX is therapeutic, Non-_TX is non-therapeutic. The flow chart happens for each single match

Machine Learning Model for Clinical Notes without Metadata

Given the lack of metadata in ACEP notes, outcome variables could not be determined at the record level as they were for Direct EHR encounters; instead, determinations were made based on the surrounding text of each key term within a record. Drug terms were searched first. For each match, a snippet of text—words on either side of the term—was extracted. Within that snippet, a regular expression determined if the drug was mentioned in the context of a lab result. If so, the drug screening variable was flagged, and no other variable could be flagged within that snippet. If the context was not related to drug screening or drug risk use scores, the snippet was evaluated for non-therapeutic use using a logistic regression machine learning model from the Python scikit-learn package, which the team trained to detect contexts of non-therapeutic drug use (33).

To create the training data, a sample of ACEP clinical notes was searched for stimulant and opioid drug terms, and 370 snippets were extracted from the matches. A team clinician annotated these snippets to indicate whether the drug mentioned was used non-therapeutically. The snippets underwent basic textual preprocessing, including converting to lowercase, replacing continuous digits with the string “num,” and parsing sentences into words and their stems using the Natural Language Toolkit’s word_tokenize() method and PorterStemmer class (34). The normalized text was then vectorized using TF-IDF, with the vectorizer fitted over all available snippets—about 8,000, not just those labeled. **Figure 4** illustrates the processing steps of the machine learning model for analyzing ACEP clinical notes.

Figure 4. Machine learning process flow chart for analyzing clinical notes from the American College of Emergency Physicians electronic health record database



**The process from this point is carried out for each matched term, until all terms have been assigned flags.
NOTES: Regex is a regular expression function that takes in a list of terms and searches for them in the clinical note. STIM is stimulant, TX is therapeutic.

Mapping Clinical Notes to Outcome Variables

Once the non-therapeutic status of a drug term was established, either through metadata in the Direct EHR notes or the non-therapeutic classifier for ACEP notes, the presence or absence of a specific outcome in a record was determined based on the combination of the drug's category and its non-therapeutic context **(Table D).**

Some combinations of drug category and clinical context did not correspond to a specific outcome variable. For example, although methamphetamine and 3,4-methylenedioxymethamphetamine (MDMA) can be used medically, there was little to no evidence in the clinical notes supporting therapeutic use. Similarly, nearly all mentions of cocaine use were non-therapeutic.

Figure 5 illustrates the outcome variables from the clinical note component. Some of these variables, shown as parent and child variables in the figure, were partially or entirely derived from other variables. The outcome variables for prescription stimulant misuse, unspecified non-therapeutic stimulant use, and illicit MDMA use, along with their derivatives, are marked as unreliable. Further details about this are provided in the upcoming section on algorithm performance.

Validating and refining the NLP algorithm against gold standard data

Algorithm development and annotation were conducted simultaneously, allowing for interim evaluation and iterative refinement of the algorithm. Code was written to extract ground truth data for each encounter from the annotated dataset and to compare the annotator's determinations against the algorithm's output.

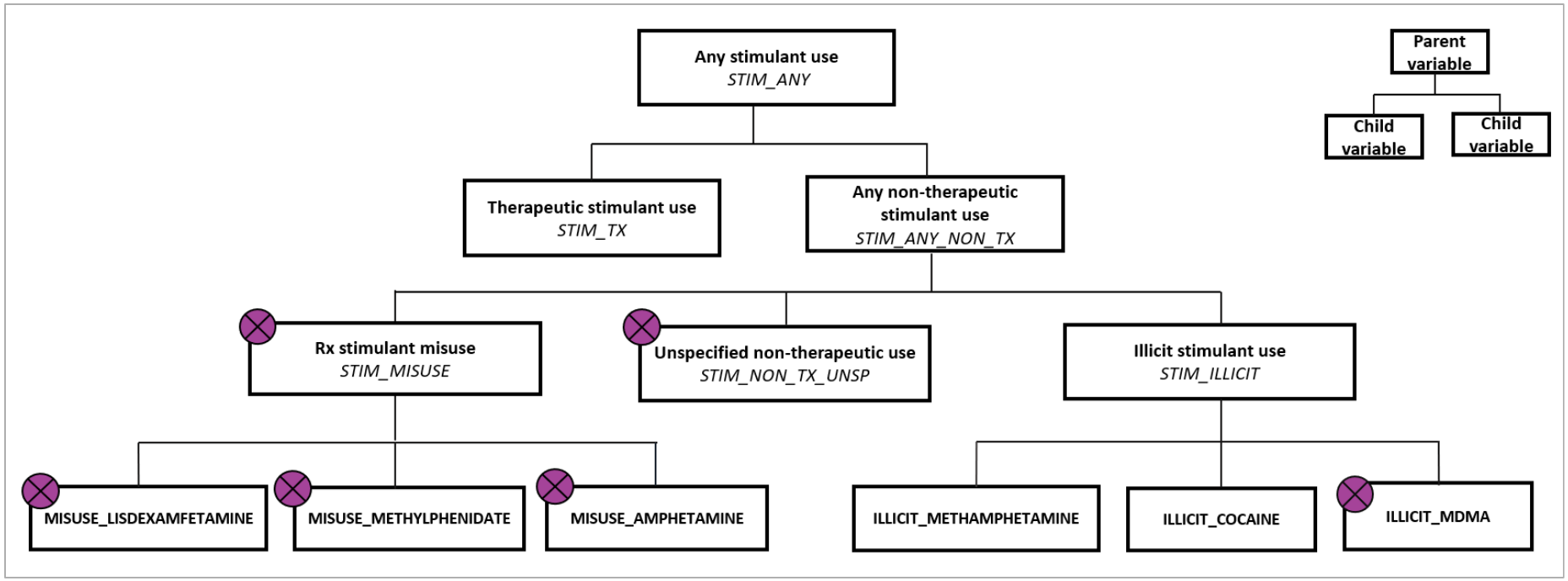
Some of the proposed outcome variables had so few actual positives (that is, true positives plus false negatives) even in the development set that developing rules to find them was difficult; there were too few cases from which to generalize. For variables with many actual positive cases (all variables had many actual negatives), the algorithm could be adjusted, re-run, and performance metrics could be calculated on a large enough number of instances to be meaningful. In contrast, if a variable only had a handful of actual positive cases, even if the algorithm could be adjusted to yield perfect performance for that variable (100% raw accuracy), there was doubt about how generalizable the algorithm would be in identifying that outcome variable.

For this reason, the Technical Expert Panel recommended setting a minimum threshold for actual positives in the test set, below which the outcome variable's accuracy would be unreliable. Only categories with at least 24 actual positives in the test set were kept to ensure accurate performance metrics. When a variable's actual positives fell below this threshold, it was excluded as a separate

variable. However, if this variable was used to derive another variable, its positive flags would be consolidated into the parent variable where applicable.

For example, if the illicit MDMA variable had too few actual positive cases to be considered reliable, it was still used to develop the illicit stimulant use variable; the positive cases from the illicit MDMA would contribute to it even if the illicit MDMA variable was eventually discarded, as long as the illicit stimulant use variable itself met the threshold**. Figure 5** illustrates how the output variables from the clinical note component are grouped and highlights those that were ultimately removed for not meeting the threshold of 24 positive cases in the test set. The difficulty in finding the outcomes whose variables were discarded is elaborated on in the section on limitations.

Figure 5. Stimulant use outcome variables for the clinical note component of the Stimulant Algorithm



⊗ Indicates variables dropped from the final research file because they did not meet the threshold of 24 actual positive cases in the test set.
NOTES: STIM is stimulant, NON_TX is non-therapeutic use (illicit or prescription misuse), TX is therapeutic (as prescribed) use, NON_TX_UNSP is non-therapeutic, unspecified use, Rx is prescription, MDMA is 3,4-methylenedioxymethamphetamine. The outcome variable for stimulant and opioid non-therapeutic co-use is not shown, as it is derived from using the non-therapeutic opioid and non-therapeutic stimulant use variables from both the code and clinical note component. The complete codebook for the Stimulant Algorithm outcome variables, which includes the opioid outcome variables, can be found here: https://github.com/CDCgov/stimulant_opioid_algorithm_clinical_notes

Testing Algorithm Performance

Performance assessment was conducted for each component separately as well as for the overall algorithm (component results combined). When calculating the performance for each component, the ground truth for the clinical note component was based on information found by annotators in the clinical notes, while the ground truth for the medical code component was based on the identified medical codes. Using illicit cocaine as an example, if the annotator only found cocaine use as a code in the diagnoses table for that encounter, the ground truth for the clinical note component's illicit cocaine use variable is “False,” while it is “True” for the medical code component.

For the combined algorithm, the ground truth is "True" regardless of where the information was identified by the annotators. The outcome of stimulant and opioid co-use was determined by the full algorithm and not the component level. The co-use variable was flagged if both non-therapeutic stimulant use and non-therapeutic opioid use were found, regardless of whether either was identified by the NLP or code component. As such, the algorithm’s performance for the *co-use* variable was only calculated for the full algorithm.

As outlined in the methodology above, a logistic regression machine learning model was integrated into the NLP algorithm to identify non-therapeutic contexts of relevant drug mentions within ACEP clinical notes that lack metadata. This sub-component of the NLP algorithm was itself independently evaluated. It had a dataset of 370 labeled examples and was divided into a training set (75%) and a test set (25%). Performance of this sub-component is also reported in the Results section.

Performance Metrics

The following metrics were used to assess the algorithm’s performance: recall, precision, the F1-score, and Matthew’s correlation coefficient (MCC). “Recall” measures how well a model can find all the positive cases in a data set. It is calculated as the ratio of true positives (TP) predictions—correctly identified relevant instances—to the sum of true positives and false negatives (FN)—cases incorrectly classified as non-cases. “Precision” measures how well a model can correctly identify positive cases out of the cases it predicted as positive. It is calculated as the ratio of true positive predictions to the sum of true positives and false positives (FP)—negative cases incorrectly classified as positive.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

The F1-score, the harmonic mean of recall and precision, is one of the most common performance metrics in binary classification. MCC (identical to Pearson's Phi coefficient) accounts for true and false negatives and true and false positives to assess model performance. It measures the model's binary classification ability and is superior in measuring imbalanced data sets (35).

$$F1\ score = \frac{TP}{TP + 1/2(FP + FN)}$$

$$MCC = \frac{(TN * TP) - (FP * FN)}{\sqrt{(TN + FN)(FP + TP)(TN + FP)(FN + TP)}}$$

Results

The Stimulant Algorithm (both components) assigns flags for the outcome variables at the record level. These record-level flags are grouped to the encounter level for analysis (an encounter can have multiple records) rather than to a person or patient level. Each encounter represents a single emergency department visit or hospitalization; therefore, multiple encounters for the same patient are treated as separate observations. The results presented below are from analysis performed at the encounter level for the test portion of the annotated dataset.

Results focus on performance measures for the high-level outcomes of interest, evaluated separately for the medical code component, the clinical note component, and the combined algorithm. Because not all outcomes are identified by each component, **Table E** summarizes which outcomes are captured by each component and by the combined algorithm and provides context for the results that follow.

Table E. Outcomes identified by each component of the Stimulant Algorithm

Outcome	Code	Clinical Note	Full Algorithm
Any stimulant use	✓	✓	✓
Therapeutic stimulant use	✓	✓	✓
Any non-therapeutic stimulant use	✓	✓	✓
Any illicit stimulant use	✓	✓	✓
Illicit methamphetamine use	✓	✓	✓
Illicit MDMA use	✓	✗	✓
Illicit cocaine use	✓	✓	✓
Any non-therapeutic opioid use	✓	✓	✓
Stimulant–opioid co-use	✓	✓	✓
Drug screening	✓	✓	✓
Prescription misuse	✓	✗	✓

NOTES: ✓ is yes; the outcome is captured. ✗ *is no; the outcome is not captured.* MDMA is 3,4-methylenedioxymethamphetamine, also commonly known as ecstasy or molly

Performance of the Clinical Note Component

Algorithm performance, measured by *MCC*, was higher for stimulant non-therapeutic use (0.95) than for opioid non-therapeutic use (0.90) **(Table 1)**. Recall was higher than the precision for both stimulant and opioid non-therapeutic use. Recall was 1.0 for stimulant non-therapeutic use and 0.98 for opioid non-therapeutic use. Precision was 0.95 and 0.87 for non-therapeutic use of stimulants and opioids, respectively. This illustrates that the NLP algorithm found all or nearly all the true non-therapeutic use encounters (high recall) in the test set but incorrectly identified some encounters as non-therapeutic **(Table 2).**

Table 3 shows performance metrics for the logistic regression machine learning model. The precision and recall values for this model were 0.86 and 0.96, respectively, and the *F1-score* was 0.91 on a balanced test set comprising 92 texts**.**

Performance of the Medical Code Component

Detailed results of the medical code component performance are shown in **Table 1**. As with the clinical note component, the medical code component performed better in identifying non-therapeutic use of stimulants (*F1*-score=0.95, *MCC*=0.94) than it did for the same measure for opioids (*F1-score*=0.92, *MCC*=0.91).

Performance of the Full Stimulant Algorithm

The overall performance for the full algorithm (that is, combined medical code and clinical note components) was relatively high, although no standardized definition of high performance exists in the literature. Broadly, performance for stimulant-related variables was higher than that of opioid-related variables. Note that clinical note variables concerning prescription misuse and unspecified non-therapeutic use were dropped because they did not meet the threshold of 24 for actual positives.

The full algorithm performed well for the top-level variables of non-therapeutic stimulant use (*F1-score*=0.98, *MCC*=0.95), and non-therapeutic opioid use (*F1-score:* 0.92, *MCC*: 0.90) **(Table 1)**. For recall, all encounters flagged for non-therapeutic use by the medical code component were also flagged by the clinical note component, but the clinical note component found true positives that the medical code component did not find. In terms of precision, the medical code component is generally more precise. The medical code component did not produce any false positives that were not also produced by the clinical note component. There were, however, false positives from the clinical note component that the medical code component properly identified **(Table 2**). The performance for the co-use variable was only calculated for the full algorithm, with an *F1-score* of 0.95 and an *MCC* of 0.94.

The Full algorithm's performance was better for Direct EHR than for ACEP data. For non-therapeutic stimulant use, the *F1-score* and *MCC* on the annotated test set for Direct EHR were 0.99 and 0.99, and 0.94 and 0.87 for ACEP **(Table 4).**

The counts of true and false positives, and of true and false negatives used to calculate the performance measures in **Table 1** can be found in **Table 2**, where the detailed counts of algorithm predictions and annotated ground truth are shown.

Limitations and Analytic Considerations

Although the performance metrics substantiate the successful development of the Stimulant Algorithm, notable limitations and analytic considerations for users applying the algorithm to their data are discussed below.

Limitations of the Medical Code Component

Diagnosis codes submitted to the National Hospital Care Survey (NHCS) in various code systems are translated into the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) for internal use, including development of this algorithm **(Appendix).** A limitation of this standardization is the potential loss of specificity during translation. For example, for the 2020 NHCS,

specificity regarding methamphetamine use was lost when Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) codes were mapped to ICD-10-CM.

ICD-10-CM diagnosis codes are revised annually, with new codes introduced as needed. Methamphetamine use codes were added to ICD-10-CM in 2022. Therefore, in 2020—the data year for the algorithm development data—specific SNOMED CT codes for methamphetamine use existed, but there was no ICD-10-CM equivalent, resulting in a loss of specificity during translation, which is further explained in the **Appendix.**

Fortunately, this issue was identified during annotation, leading to the incorporation of SNOMED CT codes for Methamphetamine use, and all active ICD-10-CM opioid and stimulant use diagnosis codes as of 2024 into the algorithm. Previous ICD-10-CM codes were also included to ensure algorithm effectiveness when applied to data predating 2020. However, without a comprehensive analysis, it is impossible to determine whether other similar, but less frequent issues occurred and went undetected.

The code component is therefore limited in that it can only detect ICD-10-CM stimulant or opioid diagnosis codes, and SNOMED CT diagnosis codes specific to methamphetamine use. Despite this limitation, users can specify additional code systems and expand beyond those originally incorporated.

The code component of the algorithm is publicly available here: https://github.com/CDCgov/stimulant_opioid_algorithm_medical_codes_R

Limitations of the Clinical Note Component

The NLP algorithm is limited in detecting prescription misuse. In the annotation sample of 1,000 encounters, only 17 encounters were labeled for evidence of prescription misuse in the clinical notes (five for stimulant misuse and 12 for opioid misuse). Of these, 13 were in the development set—too few to generalize how prescription misuse is presented in clinical notes. Only four were in the test set, which was not enough for reliable performance metrics, given the threshold of 24 actual positives previously discussed.

For some of the few cases labeled for prescription misuse, team clinicians had to piece together disjointed sections of the clinical notes to make that determination. This is difficult to capture with rule-based regular expressions, and even some techniques with advanced understanding of language often have character limits that prevent analyzing the full span of text needed to determine prescription misuse in these instances. Furthermore, even if the text was short enough for classification, there were too few examples of prescription misuse in the annotation sample to train a classifier properly.

Perhaps if the annotated sample had more actual positives, more cases of prescription misuse might have been captured through NLP; thus, oversampling prescription misuse for annotation could have provided more cases to train the models. Unfortunately, this was not done due to resource limitations and the project timeline.

Consequently, the NLP algorithm does not contribute to the algorithm's output variables for prescription misuse; instead, such cases identified by the Stimulant Algorithm are captured solely by the medical code component.

Limitations of the full Stimulant Algorithm

The algorithm has limitations in specifying the manner of use for fentanyl, buprenorphine, and methadone, based on assumptions made for the case definitions (see Case Definition section for more details).

Fentanyl use identified in clinical notes or diagnosis codes for poisoning, use, or dependency was assumed to be illegally manufactured fentanyl (IMF) and classified as illicit opioid use, unless the record clearly indicated therapeutic administration (e.g., medication records, procedure codes or diagnosis codes for therapeutic use). In cases without clear evidence of therapeutic use, fentanyl was assumed to represent non-prescribed (illicit) use, consistent with evidence that non-therapeutic fentanyl use typically involves IMF (23). Consequently, some cases of prescribed fentanyl misuse may be misclassified as illicit opioid use if the data does not clearly distinguish the context.

A second assumption is that buprenorphine and methadone, unless otherwise specified, are used for medication-assisted therapy (MAT) for opioid use disorder. While both drugs can be prescribed for pain and have potential for misuse, MAT is their most common indication (26,27). Therefore, the algorithm treats methadone and buprenorphine as proxies for non-therapeutic opioid use unless the record explicitly states otherwise. This approach may under-detect prescription misuse involving these medications.

Despite these limitations in specificity, the algorithm reliably detects non-therapeutic use of these opioids, which is essential for identifying the stimulant-opioid co-use outcome.

Analytical Considerations

When analyzing or interpreting the Stimulant Algorithm's output, it's important to note that the algorithm cannot distinguish between historic and current substance use due to inherent limitations in clinical data (16,17). In both coded data and clinical notes from the 2020 NHCS development set, it was difficult to

consistently determine whether a patient's substance use was resolved or active. As a result, unless the distinction is clear in the data, the algorithm captures any documented stimulant or opioid use—past or present—in a patient's clinical history for the data year(s) analyzed.

This is a considerable limitation. Although the algorithm was developed using 2020 data, it searches clinical records that may include historical information; therefore, without a distinction in the data, results should not be used if the timing of use is critical. For example, a patient with a diagnosis code for methamphetamine use from ten years ago will still be flagged for methamphetamine use in the year of interest. Similarly, a patient with historical stimulant use in the clinical notes without distinction, and a current diagnosis code for heroin use, would be detected for co-use, even if the stimulant use is not current.

Without consistent, accurate timing information in clinical records, this is the closest estimate for identifying use in both coded and clinical notes data. However, if the data being analyzed do distinguish timing, users of the Stimulant Algorithm (36-39) can subset their data accordingly before applying the algorithms**.** Additionally, if only administrative claims data are being analyzed, the output would reflect current or active substance use for the calendar year, as these diagnoses are typically current for billing purposes.

Discussion

With support from OS-PCORTF, the Stimulant Algorithm was developed to identify stimulant use in electronic health records (EHR), including administrative claims data. The algorithm detects: 1) therapeutic stimulant use, 2) non-therapeutic stimulant use—including illicit use and prescription misuse, 3) non-therapeutic co-use of stimulants and opioids, and 4) whether a drug screen was performed. The full algorithm consists of two components: a medical code component that searches coded data (including claims), and a clinical note component that uses a Machine Learning (ML) model with Natural Language Processing (NLP) to analyze clinical notes.

Evaluation against annotated gold standard data demonstrated robust performance. The algorithm achieved high accuracy in detecting non-therapeutic stimulant use and stimulant-opioid co-use (*MCC* 0.95 and 0.94, respectively), and good accuracy for detecting non-therapeutic opioid use (*MCC* 0.90). This slightly lower performance for opioid detection was expected, given the algorithm's primary focus on stimulants.

A key objective of this project was to identify prescription stimulant misuse and distinguish it from illicit use, as well as to gain insight into which prescription stimulants are most frequently misused. Only two

diagnosis codes for prescription stimulant misuse are available (methylphenidate and amphetamine), so it was anticipated that clinical notes would provide additional detail regarding the types of prescription stimulants being misused. However, several challenges emerged in identifying prescription misuse within clinical notes.

As discussed in the limitations, the primary issue was the small number of actual cases in the development and test data, compounded by fragmented information in clinical notes that was difficult to capture with available resources. Additionally, provider documentation often used vague and non-standardized language when describing misuse.

Several important insights emerged from these limitations. First, greater standardization of reporting language for prescription misuse in clinical notes would improve detection. Second, as the range of prescription stimulants expands, more specific codes would enhance providers' ability to document misuse, facilitating research and reporting. Finally, distinguishing between a patient misusing their own prescription and one using another person's medication (also prescription misuse) proved challenging; therefore, more specific information in the clinical notes would be valuable for analyzing usage patterns.

Despite the limitation in identifying prescription misuse in clinical notes, the Stimulant Algorithm performed well in identifying the main outcomes of interest: prescription misuse (solely through the medical code component), therapeutic stimulant use, and the co-use of stimulants and opioids, and drug screening (by the full algorithm). Although not documented in this report, the algorithm was successfully applied to EHR data outside the NHCS to assess its generalizability to other clinical data sources. Findings from these assessments, to be reported later, will further inform researchers on how the Stimulant Algorithm may be improved and applied more broadly in healthcare research.

The Stimulant Algorithm was also applied to the complete 2020 National Hospital Care Survey (NHCS) data and released as a supplemental file to the nationally representative dataset, available in the NCHS Research Data Center (https://www.cdc.gov/nchs/data/nhcs/Stimulant-Use-Dataset-RDC-Data-Documentation_5-27-26.pdf). When using the 2020 NHCS, the Stimulant Algorithm results, combined with the ability to link hospital encounter data to external datasets such as the National Death Index and U.S. Housing Urban and Development administrative data, provide researchers with a rich resource to study patient-centered health outcomes associated with stimulant use.

The full algorithm programming code (clinical notes and code components), along with documentation—including instructions, code lists, and search terms—is publicly available on the CDC GitHub page (https://github.com/CDCgov) for use by researchers and other end users (36-39).

References

1. Drug overdose deaths – Health, United States. National Center for Health Statistics; 2023. Available from: https://www.cdc.gov/nchs/hus/topics/drug-overdose-deaths.htm

2. WISQARS-Web-based injury Statistics Query and Reporting System. Animated leading causes of death. National Center for Injury Prevention and Control; 2024. Available from: https://wisqars.cdc.gov/animated-leading-causes/

3. Garnett M, Miniño A. Drug overdose deaths in the United States, 2023-2024. NCHS Data Brief. 2026 Jan;(549):1-13. Hyattsville, MD: National Center for Health Statistics. DOI: https://dx.doi.org/10.15620/cdc/174639

4. Ciccarone D. The rise of illicit fentanyls, stimulants and the fourth wave of the opioid overdose crisis. Current Opinion in Psychiatry [Internet]. 2021 May 10;34(4):344–50. DOI: https://doi.org/10.1097/yco.0000000000000717

5. Spencer MR, Miniño AM, Garnett MF. Co-involvement of opioids in drug overdose deaths involving cocaine and psychostimulants, 2011–2021. NCHS Data Brief, no. 474. Hyattsville, MD: National Center for Health Statistics; 2023. DOI: https://dx.doi.org/10.15620/cdc:129733

6. Dattangire, R., & Biradar, D. (2024). Leveraging Big Data for Disease Surveillance and Public Health Interventions. International Journal of Global Innovations and Solutions (IJGIS). DOI: https://doi.org/10.21428/e90189c8.8f592e36

7. National Institutes of Health. National Library of Medicine. US. Department of Health and Human Services. Finding and Using Health Statistics. Available from: https://www.nlm.nih.gov/oet/ed/stats/03-200.html

8. Wei WQ, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. J Am Med Inform Assoc 23(e1):e20–7. 2016.

9. Smoller JW. The use of electronic health records for psychiatric phenotyping and genomics. Am J Med Genet B: Neuropsychiatric Genetics 177(7):601–12. 2018.

10. White DG, Adams NB, Brown AM, O'Jiaku-Okorie A, Badwe R, Shaikh S, Adegboye A. Enhancing identification of opioid-involved health outcomes using National Hospital Care Survey data. National Center for Health Statistics. Vital Health Stat 2(188). 2021.

11. Brown AM, White DG, Adams NB, O'Jiaku-Okorie A, Badwe R, Shaikh S, Adegboye A. Identifying co-occurring disorders among patients with an opioid-involved hospital encounter using National Hospital Care Survey data. National Center for Health Statistics. Vital Health Stat 2(193). 2022.

12. Brown AM, White DG, Adams NB, Badwe R, Shaikh S, Guluma L. Validation of the enhanced opioid identification and co-occurring disorders algorithms. National Center for Health Statistics. Vital Health Stat 2(205). 2024.

13. Centers for Disease Control and Prevention. National Center for Health Statistics. National Hospital Care Survey. 2024. Available from: https://www.cdc.gov/nchs/nhcs/index.html

14. American College of Emergency Physicians. Clinical Emergency Data Registry (CEDR). Available from: https://www.acep.org/cedr/

15. Technical Documentation: National Center for Health Statistics. Division of Health Care Statistics. 2020 National Hospital Care Survey (NHCS) Public Use Data Files Documentation. Hyattsville, Maryland. July 2024. Available from: https://www.cdc.gov/nchs/data/nhcs/2020-NHCS-PUF-Tech-Doc-508.pdf

16. Schulz, W.L., Young, H.P., Coppi, A. *et al.* Temporal relationship of computed and structured diagnoses in electronic health record data. *BMC Med Inform Decis Mak* 21, 61 (2021). https://doi.org/10.1186/s12911-021-01416-x

17. Holmes JH, Beinlich J, Boland MR, Bowles KH, Chen Y, Cook TS, et al. Why Is the Electronic Health Record So Challenging for Research and Clinical Care? Methods of Information in Medicine. 2021 May;60(01/02):032–48. DOI: https://doi.org/10.1055/s-0041-1731784

18. Substance Abuse and Mental Health Services Administration. Results from the 2022 National Survey on Drug Use and Health: A companion infographic [Report]. Rockville, MD: Substance

Abuse and Mental Health Services Administration; 2022. Available from: https://www.samhsa.gov/data/sites/default/files/reports/rpt42730/2022-nsduh-infographic-report.pdf

19. U.S. Department of Justice. U.S. Drug Enforcement Administration. Drug Scheduling. 2018. Available from: https://www.dea.gov/drug-information/drug-scheduling

20. U.S. Food and Drug Administration. Center for Drug Evaluation and Research. Application Number: 209963 Clinical pharmacology and biopharmaceutical review(s): NDA 209963. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/209963Orig1s000ClinPharmR.pdf

21. U.S. Department of Justice. U.S. Drug Enforcement Administration. Methamphetamine. Available from: https://deadiversion.usdoj.gov/drug_chem_info/meth.pdf

22. American College of Medical Toxicology. ACMT Position Statement: Safety Issues Regarding Prescription Fentanyl Products. Journal of Medical Toxicology. 2016;12(2):211-212. DOI: https://doi.org/10.1007/s13181-016-0535-y

23. Centers for Disease Control and Prevention. Fentanyl fact sheet: general information. Available from: https://www.cdc.gov/stop-overdose/media/pdfs/2024/05/CDC_Fentanyl-Fact-Sheet_General_508.pdf

24. U.S. Department of Justice. The Methamphetamine Threat: A Report from the National Drug Intelligence Center. Available from: https://www.justice.gov/archive/ndic/pubs5/5140/5140p.pdf

25. Overdose Epidemic Report 2023. Available from: https://end-overdose-epidemic.org/wp-content/uploads/2023/11/23-894446-Advocacy-2023-overdose-report_FINAL.pdf

26. Drug Enforcement Administration. Diversion Control Division. Drug & Chemical Evaluation Section. Buprenorphine. 2022. Available from: https://www.deadiversion.usdoj.gov/drug_chem_info/buprenorphine.pdf

27. Drug Enforcement Administration. Diversion Control Division. Drug & Chemical Evaluation Section. Methadone. 2019. Available from: https://deadiversion.usdoj.gov/drug_chem_info/methadone/methadone.pdf

28. Singh D, Saadabadi A. Naltrexone. Treasure Island (FL): StatPearls Publishing; 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK534811/

29. U.S. Food and Drug Administration. New recommendations for naloxone [press release]. July 23, 2020. Available from: https://www.fda.gov/drugs/drug-safety-and-availability/new-recommendations-naloxone

30. Data Documentation on Hospitalizations and Drugs. 2020 National Hospital Care Survey. National Center for Health Statistics Research Data Center. Centers for Disease Control and Prevention; Hyattsville, MD: National Center for Health Statistics; October 2024. Available from: https://www.cdc.gov/rdc/data/nhcs/Stimulant-File-RDC-Data-Documentation.pdf

31. Roberts, Angus, Robert Gaizauskas, Mark Hepple, George Demetriou, Yikun Guo, Ian Roberts, and Andrea Setzer. Building a semantically annotated corpus of clinical texts. Journal of biomedical informatics 42(5): 950-966. 2009.

32. Xia, Fei, and Meliha Yetisgen-Yildiz. Clinical corpus annotation: challenges and strategies. In: Proceedings of the third workshop on building and evaluating resources for biomedical text mining (BioTxtM'2012) in conjunction with the international conference on language resources and evaluation (LREC). Istanbul, Turkey. 2012.

33. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res 12:2825-2830. 2011.

34. Loper E, Bird S. NLTK: the Natural Language Toolkit. Association for Computational Linguistics-02: Workshop on effective tools and methodologies for teaching natural language processing and computational linguistics. July 2002.

35. Chicco, Davide, and Giuseppe Jurman. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics 21:1-13. 2020.

36. Centers for Disease Control and Prevention. Stimulant opioid algorithm medical codes. GitHub. Available from: https://github.com/CDCgov/stimulant_opioid_algorithm_medical_codes_R

37. Centers for Disease Control and Prevention. Stimulant opioid algorithm medical codes (SAS). GitHub. Available from: https://github.com/CDCgov/stimulant_opioid_algorithm_medical_codes_SAS

38. Centers for Disease Control and Prevention. Stimulant opioid algorithm clinical notes. GitHub. Available from: https://github.com/CDCgov/stimulant_opioid_algorithm_clinical_notes

39. National Center for Health Statistics. Non-therapeutic drug use. Hugging Face. Available from: https://huggingface.co/NCHS/Non_Therapeutic_Drug_Use

Detailed Tables

Table 1. Performance metrics for the Stimulant Algorithm evaluated against the annotated dataset, by component, for the high-level outcomes of interest

Performance Metric by Component	Non-therapeutic stimulant use	Non-therapeutic opioid use	Stimulant-opioid co-use*
Clinical note component			
Recall†	1.00	0.98	…
Precision§	0.95	0.87	…
F1¶	0.97	0.92	…
Matthews Correlation Coefficient	0.94	0.90	…
Medical code component			
Recall†	0.97	0.85	…
Precision§	0.93	1.00	…
F1¶	0.95	0.92	…
Matthews Correlation Coefficient	0.94	0.91	…
Full algorithm			
Recall†	1.00	0.98	0.98
Precision§	0.95	0.87	0.93
F1¶	0.98	0.92	0.95
Matthews Correlation Coefficient	0.95	0.90	0.94

... Category is not applicable.
*Co-use is a proxy for the co-occurrence of non-therapeutic stimulant and opioid use in an encounter.
†Percentage of correctly identified positives out of all true positives, also known as sensitivity.
§Percentage of identified positives that are true positives.
¶Harmonic mean of recall and precision, a common measure of algorithm performance.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020.

Table 2. Agreement counts between the Stimulant Algorithm and the annotated dataset, by component, for the high-level outcomes of interest

Characteristic	Non-therapeutic stimulant use		Non-therapeutic opioid use		Stimulant-opioid co-use*	
	Annotator positive	Annotator negative	Annotator positive	Annotator negative	Annotator positive	Annotator negative
Clinical note component						
Algorithm positive	105	5	47	9	…	…
Algorithm negative	–	86	1	139	…	…
Medical code component						
Algorithm positive	28	2	14	3	…	…
Algorithm negative	1	165	6	173	…	…
Full algorithm						
Algorithm positive	105	5	47	9	42	3
Algorithm negative	–	86	1	139	1	150

– Quantity zero
…Category not applicable
*Co-use is a proxy for the co-occurrence of non-therapeutic stimulant and opioid use in an encounter and was only identified by the full algorithm, not by component.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020

Table 3. Performance metrics for the non-therapeutic context machine learning model

Performance Metric	Precision*	Recall†	F1 score§
Score	0.86	0.96	0.91

*Percentage of identified positives that are true positives.
†Percentage of correctly identified positives out of all true positives, also known as sensitivity
§Harmonic mean of recall and precision, a common measure of algorithm performance.
NOTES: There were 46 cases of actual positives for non-therapeutic context, and 46 cases of actual negatives, for a balanced test of 92 text snippets.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020

Table 4. Full algorithm performance metrics by data source

Characteristic	Non-therapeutic stimulant use		Non-therapeutic opioid use	
	Direct EHR*	ACEP†	Direct EHR	ACEP
Recall§	1.00	1.00	0.97	1.00
Precision¶	0.99	0.88	0.97	0.65
F1§§	0.99	0.94	0.97	0.79
Matthews Correlation Coefficient	0.99	0.87	0.96	0.75

*Direct EHR represents electronic health records submitted to NCHS directly from the hospital.
†ACEP EHR represents electronic health records with clinical notes submitted from the American College of Emergency Physicians.
§Percentage of correctly identified positives out of all true positives, also known as sensitivity.
¶Percentage of identified positives that are true positives.
§§Harmonic mean of recall and precision, a common measure of algorithm performance.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020

Table 5. Performance metrics for all reliable outcome variables from the clinical note component

Variable[1]	Outcome*	Performance Metric	
		F1†	MCC¶
STIM_ANY	Any stimulant use	0.96	0.88
STIM_ANY_NON_TX	Any non-therapeutic stimulant use	0.97	0.94
STIM_TX	Therapeutic stimulant use	0.87	0.85
STIM_ILLICIT	Illicit stimulant use	0.98	0.96
ILLICIT_METHAMPHETAMINE	Illicit methamphetamine use	0.99	0.98
ILLICIT_COCAINE	Illicit cocaine use	0.97	0.97
OPIOID_ANY_NON_TX	Any non-therapeutic opioid use	0.92	0.9
OPIOID_ILLICIT	Illicit opioid use	0.94	0.92

*Outcome variables from the clinical note component.
†Harmonic mean of recall and precision, a common measure of algorithm performance.
§Matthew's correlation coefficient
NOTES: Clinical note component refers to the natural language processing and machine learning model. Variables shown are those that had enough actual positives to be considered reliable for identifying the outcomes of interest.

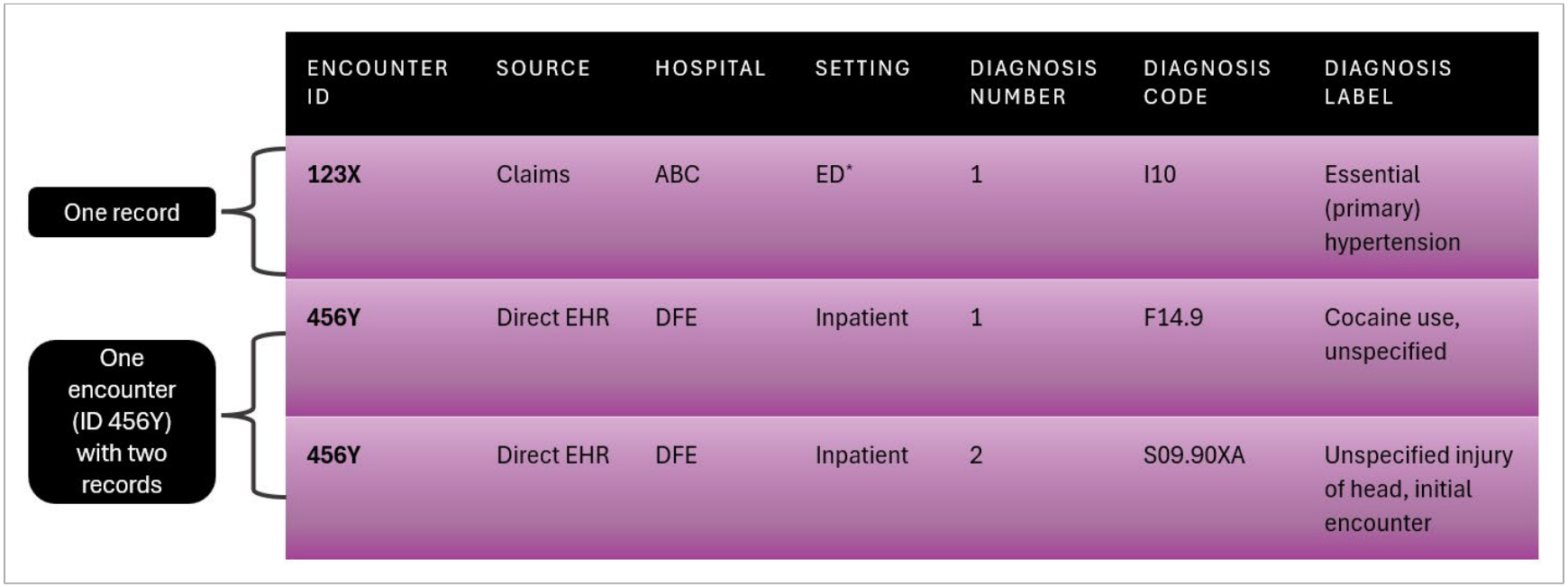
Appendix

National Hospital Care Survey Data Structure

Processed NHCS data are stored in a relational database. An encounter, assigned a unique identifier (ID), is the data unit of collection and represents all the information reported in the EHR during a patient's visit to the emergency department or an inpatient hospital stay. Note that a patient seen in the ED and later admitted into the hospital will have two encounters. Collected data are organized into multiple tables, each representing information related to a particular aspect of care provided. NHCS data have tables on diagnoses, medications, laboratory tests, clinical notes, and procedures. Each row in these tables is a 'record' representing any information related to the table (aspect of care provided) submitted for an encounter. Each record can be linked to the encounter it belongs to by the encounter ID, allowing for complex queries across these multiple tables. An encounter can have multiple records in one table or multiple tables.

For illustration, consider that patient AB was admitted to the hospital for two days. This patient's hospital stay is called an 'encounter' and is assigned ID '456Y'. While admitted, the patient receives two diagnoses: 'unspecified cocaine use' and an 'unspecified injury to the head'. AB also undergoes two lab tests, is given three types of medications, and undergoes a procedure for the head injury. The healthcare team documents the care given in five clinical notes. Encounter 456Y will have two records in the diagnoses and laboratory tables, three records in the medications table, five records in the clinical notes table, and so on. **Figure I** illustrates the NHCS data structure with an example of a diagnosis table that includes patient AB's diagnosis records during encounter 456Y.

Figure I. Simplified illustration of a diagnosis table with encounters and records in the National Hospital Care Survey



ENCOUNTER ID	SOURCE	HOSPITAL	SETTING	DIAGNOSIS NUMBER	DIAGNOSIS CODE	DIAGNOSIS LABEL
123X	Claims	ABC	ED*	1	I10	Essential (primary) hypertension
456Y	Direct EHR	DFE	Inpatient	1	F14.9	Cocaine use, unspecified
456Y	Direct EHR	DFE	Inpatient	2	S09.90XA	Unspecified injury of head, initial encounter

*Emergency Department
NOTES: An encounter is an emergency department visit, or an inpatient hospital stay. Actions taken or data collected while delivering care during the encounter are documented as records and organized in tables representing the broader aspects of care, such as diagnosis, procedures, lab tests…etc. The image illustrates the diagnosis table, with records belonging to different encounters. An encounter can have multiple records within one table or in multiple tables. For example, the encounters represented here can also have multiple records in the procedure table or the medications table.
SOURCE: National Hospital Care Survey

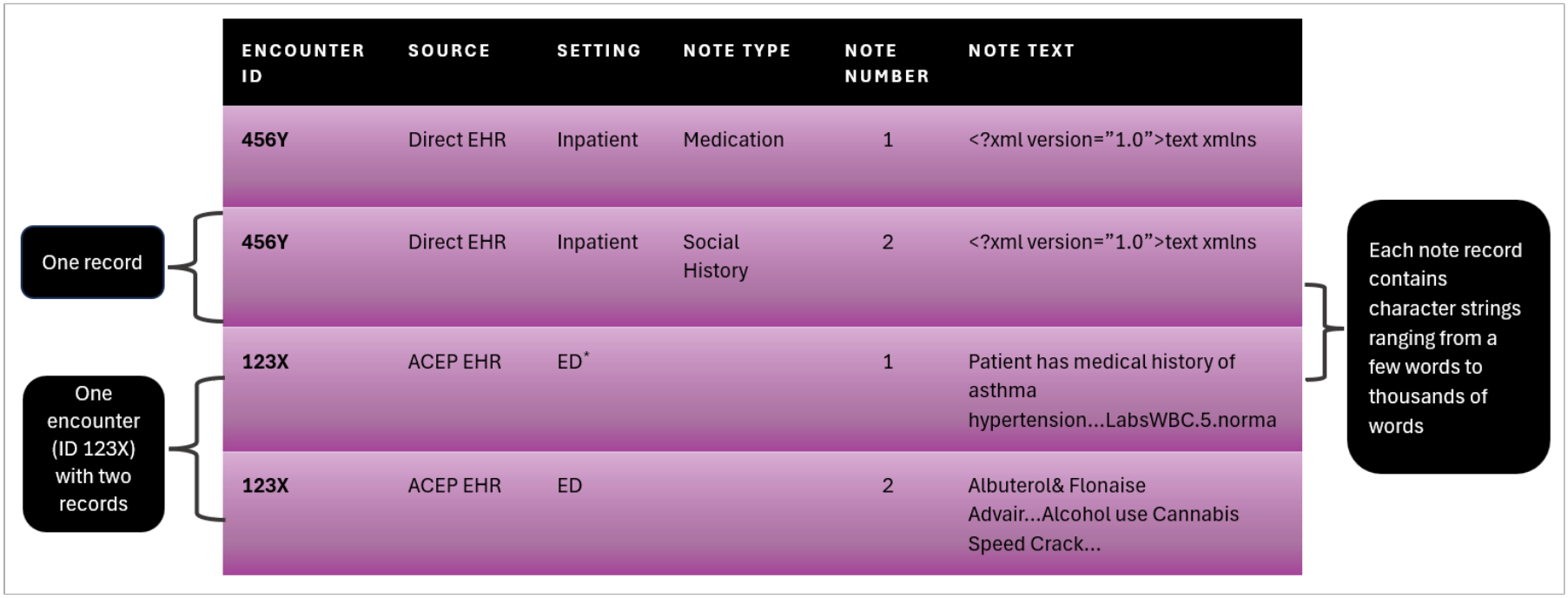
Data Structure: Clinical Notes Table

Electronic Health Record (EHR) data submitted to NHCS may include clinical notes, depending on whether the hospital chooses to provide them. NHCS receives EHR data from two sources: the American College of Emergency Physicians' clinical registry (ACEP) and direct submissions from participating hospitals (Direct EHR). A subset of the submitted clinical notes was used to develop and test the algorithm.

Direct EHR clinical notes were received in an Extensible Markup Language (XML) format designed for web browser viewing—that is, the data included HTML-style tags. These notes were accompanied by metadata (or note types) that identified the type of content in each record. For example, note types included "Medications," "Filed Vital Signs," "Inpatient Encounters," "Social History," and "Reason for Visit."

In contrast, ACEP clinical notes were received in plain-text format, without metadata describing the content type. A single record could contain multiple content types—for instance, information about medications in the first half and a list of lab results in the second—with no standardized formatting or structural cues to indicate the transitions. **Figure II** illustrates the formatting differences in Direct EHR and ACEP notes.

Figure II. Simplified illustration of the clinical notes table in the National Hospital Care Survey with note records and respective sources



*Emergency Department.
NOTES: An encounter is an emergency department visit or an inpatient hospital stay. Actions taken or data collected while delivering care during the encounter are documented as records and organized in tables representing the broader aspects of care, such as diagnosis, procedures, lab tests…etc. The image illustrates the clinical notes table, with records belonging to different encounters, from the two sources of clinical notes. ACEP clinical notes do not have a note type.
SOURCE: National Hospital Care Survey

Medical Codes in the 2020 National Hospital Care Survey

Medical or health care coding is a universal system of classifying health care services (i.e., diagnoses, procedures, lab tests) into specific alphanumeric codes for each service. This is achieved through various code systems, each particular to the care provided**.** Hospitals participating in the NHCS submit data in the code systems relevant to their specific hospital record structure, practice, and billing guidelines. NCHS retains these original data but performs post-processing to standardize certain coded data into select code systems, improving the internal (NCHS) use of NHCS data. For the 2020 NHCS, most diagnostic information was submitted in the *International Classification of Diseases, Tenth Revision, Clinical Modification* (ICD–10–CM), or the Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) and mapped into ICD-10-CM where possible for NCHS use. Medication information was submitted in multiple code systems and standardized into RxNorm nomenclature. Laboratory test data were submitted as customized codes or as Logical Observation Identifiers Names and Codes (LOINC). Procedure and service codes were submitted in the *International Classification of Disease, Tenth Revision, Procedure Coding System* (ICD-10-PCS), the Healthcare Common Procedure Coding System (HCPCS), or the Current Procedural Terminology (CPT) code system. Laboratory tests, procedures, and service codes were retained as reported.

The Stimulant Algorithm identifies codes from ICD-10-CM, RxNorm, LOINC, CPT, and HCPCS, with select SNOMED CT codes used specifically for Methamphetamine use, which did not have a direct ICD-10-CM correlate in 2020. Diagnosis codes are central to the algorithm, primarily including select ICD-10-CM codes for 'Mental and behavioral disorders due to psychoactive substance use F10–F19' and 'Poisoning by, adverse effect of, and underdosing of drugs, medicaments, and biological substances T36–T50'. These codes capture both non-therapeutic use (such as substance use, dependence, poisoning, and neonatal exposure) and therapeutic use (including underdosing and adverse reactions). Procedure and service codes are also used to identify outcomes, such as provider-administered stimulant medications for therapeutic use and medication-assisted treatment for non-therapeutic use (see **Figure 2** in the main text).

A complete list of medical codes for services, procedures, diagnoses, laboratory tests, and medications**,** along with their corresponding stimulant use outcomes is available in the CDC GitHub repository for the medical code component of the Stimulant Algorithm. The repository, which includes a ReadMe file describing how medical codes are mapped to the indicators, is available at: https://github.com/CDCgov/stimulant_opioid_algorithm_medical_codes_R/tree/main.

Acknowledgements

This work was funded by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF). The OS-PCORTF funding was made available to the Centers for Disease Control and Prevention by the Office of the Assistant Secretary for Planning and Evaluation (ASPE) through Intra-Departmental Delegation of Authority (IDDA) *OS-ASPE-PCORTF-CDC-2025-006.*

Special thanks to Priyam Patel, M.S.P.H., Rachel Fan, B.A., and the NCHS Division of Research Methodology for their assistance in data science methodology. As well as Rihem Badwe, PharmD. for her pharmaceutical expertise, Donielle G. White, M.P.H. and Lello Guluma, M.P.H. for their help in design and methodology, and the Technical Expert Panel members for their guidance throughout the project.

Suggested citation

Gidali DM, Adams NB, Shi M, Mitchell DL. Using clinical notes and medical codes to identify stimulant use in hospital data. Hyattsville, MD: National Center for Health Statistics; August 2026. DOI: https://dx.doi.org/10.15620/cdc/252467.

Copyright information

National Center for Health Statistics

For email updates on NCHS publication releases, subscribe online:

www.cdc.gov/nchs/updates/.

For questions or general information about NCHS:

Tel: 1–800–CDC–INFO (1–800–232–4636) | TTY: 1–888–232–6348

Keywords: Natural Language Processing (NLP) • Machine learning (ML) • Non-therapeutic opioid use • National Hospital Care Survey (NHCS)