National Center for Health Statistics Research Data Center

Stimulant Use Dataset Documentation

Drug Data from the National Hospital Care Survey, 2020-2021

Division of Health Care Statistics
National Center for Health Statistics
Updated May 2026

Contact Information

Data users can find the latest information about NHCS on our website, at: https://www.cdc.gov/nchs/nhcs/index.html. If data users have queries about the public use data files, they may send their question through email to nhcs@cdc.gov, or call us at 301-458-4321. A response to data user inquiries is generally provided in 1-2 business days.

The NHCS also has a listserv, where updates and information about the most recent NHCS news and data releases are sent out. To join the listserv, there are two options:

1. Send an email message to listserv@cdc.gov. Leave the subject line blank. In the body of the message, type or copy and paste: **subscribe NHCS-DATA your name** where "your name" is your first and last name.
2. Fill out and submit a subscription request form through the NCHS listservs webpage (https://www.cdc.gov/nchs/products/nchs_listservs.htm). You should receive an email notification of your subscription.

Contents

Section 1: Document Overview

This document describes the National Hospital Care Survey (NHCS) Stimulant Use Dataset, which was generated by applying the National Center for Health Statistics' (NCHS) Stimulant Algorithm to the 2020 - 2021 NHCS data. The Stimulant Algorithm was developed with funding from the Assistant Secretary for Planning and Evaluation, Office of the Secretary – Patient Centered Outcomes Research Trust Fund. The resulting dataset provides contextual information on patients' hospital encounters with documented evidence of stimulant use, and to a limited extent, opioid use in health records. The Stimulant Use Dataset is available through the NCHS Research Data Center (RDC) (1). The complete NHCS database is also available through the RDC and can be merged with the Stimulant Use Dataset for detailed analysis for the 2020 – 2021 data years. To learn more about the RDC see: https://www.cdc.gov/rdc/index.html

This document provides:

- An overview of the National Hospital Care Survey
- A high-level explanation of the Stimulant Algorithm and its methodology
- An introduction to the Stimulant Use Dataset and its key components
- Resources and links for accessing additional information

Section 2: Data Source

The Stimulant Algorithm was developed using data from the 2020 National Hospital Care Survey (NHCS), which provides nationally representative information on hospital utilization and patient care in non-institutional, non-federal hospital departments in the United States. NHCS collects inpatient and emergency department (ED) administrative claims or electronic health record (EHR) data from participating hospitals. These data include information documented during a hospital encounter, such as diagnoses, medications, discharge status, and procedures or services provided.

The 2020 NHCS data are the first to be nationally representative, allowing for production of national estimates. General information about the survey is available on the survey's webpage: https://www.cdc.gov/nchs/nhcs/index.html (2). Detailed variable descriptions and coding information for the full NHCS dataset are available in the survey's data dictionary: https://www.cdc.gov/rdc/data/b1/NHCS-RDC-Data-Dictionary.pdf (3).

Section 3: Overview of Methodology

The Stimulant Algorithm identifies clinical visits (encounters) with documented evidence of therapeutic stimulant use, non-therapeutic stimulant use, non-therapeutic co-use of stimulants and opioids, and whether a drug screen was performed during the visit.

The algorithm has two components that search the record for information related to stimulant or opioid use: a medical code component that searches for related medical codes; and a clinical note component that analyzes unstructured clinical notes, using Natural Language Processing and Machine Learning techniques.

Section 3.1: Case Definitions

This section describes the conceptual framework used in developing the Stimulant Algorithm, including the case definitions for the stimulant use indicators (i.e., outcomes of interest), essential for accurate analysis and interpretation of the Stimulant Use Dataset.

3.1.1 Manner or Type of Use

Terminology related to substance use varies across clinical settings. For the Stimulant Algorithm (and consequently the Stimulant Use Dataset), the following definitions apply. Unless otherwise specified, the term **"use"** refers to any of the categories below.

- **Therapeutic use** — stimulant or opioid consumption that is medically indicated and taken according to clinical guidance.
- **Non-therapeutic** use — consumption of a stimulant or opioid for non-medical reasons and includes:
    - **Prescription misuse** — taking a prescribed stimulant or opioid in a manner not directed by a clinician (medical professional).
    - **Illicit use** — taking illicitly manufactured substances (that is substances without accepted medical use such as methamphetamine), or use of substances typically obtained outside of legitimate clinical prescribing.

      Documented use of methamphetamine, cocaine, or fentanyl is classified as illicit use unless the medical record contains explicit evidence to the contrary.
- **Unspecified non-therapeutic use** — documented non-therapeutic stimulant or opioid use where the manner of use is not specified in the medical record.
- **Co-use** — documented non-therapeutic use of both stimulants and opioids; serves as a proxy for documented overlap of non-therapeutic stimulant and opioid use in the record.
- **Any use** — encompasses all categories above, including instances where it is unclear in the record whether use was therapeutic or non-therapeutic.

3.1.2 Temporality of Use

Electronic health record (EHR) data do not consistently distinguish between current (active) and past (resolved) substance use (4,5). In both medical code and clinical notes data from the NHCS development

set used to train the Stimulant Algorithm, timing and status of documented substance use could not be determined reliably, as diagnosis codes and other clinical documentation may persist in the patient record without clear temporal context. As a result, the Stimulant Algorithm identifies documented stimulant and opioid use, as defined above, at any point in a patient's clinical history, as reflected in the NHCS data for a given year, regardless of whether use occurred during the specific encounter. This has important implications for the analytic interpretation of these indicators, as described in Section 4.3.

Section 3.2: Medical Code Component

The NHCS collects medical codes for diagnoses, medications, laboratory tests, services, and procedures. The medical code component of the Stimulant Algorithm searches the NHCS database (or comparable data) using a defined set of medical codes to identify patient encounters that meet the primary case definitions outlined above.

In the 2020 NHCS, diagnostic information received was primarily coded in the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), the Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT), or vendor or site-specific codes. All data submitted in SNOMED-CT or vendor/site-specific codes were translated into ICD-10-CM codes where possible. Medication information was submitted using multiple coding systems and standardized to RxNorm nomenclature. Laboratory test data were submitted as either custom-codes (based on the hospital, health record vendor, or health insurance systems) or as Logical Observation Identifiers Names and Codes (LOINC). Procedure and service data were submitted using the International Classification of Diseases, Tenth Revision, Procedure Coding System (ICD-10-PCS), the Healthcare Common Procedure Coding System (HCPCS), or using Current Procedural Terminology (CPT).

The Stimulant Algorithm identifies cases using the following coding systems: ICD-10-CM for diagnoses, RxNorm for medications, LOINC for laboratory tests, and HCPCS or CPT for procedures and services. **The complete list of medical codes, along with their corresponding stimulant use indicators, is available in the** CDC GitHub repository for the medical code component of the Stimulant Algorithm (6,7). 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.

Section 3.3: Clinical Note Component

The clinical note component of the Stimulant Algorithm uses machine learning and natural language processing (NLP) methods to analyze free-text clinical notes for evidence of stimulant or opioid use. This component was developed and tested on a subset of clinical notes from the 2020 NHCS data.

The clinical note component uses predefined key search terms, together with the clinical context in which those terms appear, to identify encounters meeting the case definitions. **A complete list of search terms, their associated clinical context, and corresponding stimulant use indicators is available in the** CDC GitHub repository for the clinical note component of the Stimulant Algorithm (8). The repository,

which includes a **ReadMe file** describing how search terms are mapped to the indicators, is available at: https://github.com/CDCgov/stimulant_opioid_algorithm_clinical_notes/tree/main.

Clinical notes are submitted by only a subset of hospitals participating in the NHCS, primarily those that submit EHR data. As a result, the clinical note component can be applied only to encounters that include clinical notes. In addition, those notes must contain enough detail to be considered analyzable. Consequently, when the Stimulant Algorithm is run on NHCS data, the clinical note component processes only a fraction of all NHCS hospital encounters.

Table 1 summarizes, for each year of data included in the Stimulant Use Dataset, the total number of NHCS hospital encounters; the subset derived from EHR data; and the number of EHR encounters that contained one or more analyzable clinical note records. The final column shows the percentage of all NHCS encounters to which the clinical note component was applied.

Table 1. Summary of NHCS Encounters and Clinical Note Component Processing, 2020–2021

Survey Year	Number of Encounters			Percent Processed by the Clinical Note Component
	Total NHCS	EHR Data	Eligible for Clinical Note Processing	
2020	10,010,479	2,512,902	901,940	9.00%
2021	12,803,010	2,268,665	1,259,314	9.80%

NOTES: EHR data represents number of NHCS encounters submitted in the form of electronic health records. Analyzable clinical notes are those with sufficient and meaningful information.

Section 4: The Stimulant Use Dataset

The Stimulant Use Dataset is an encounter-level analytic dataset derived from the National Hospital Care Survey (NHCS) for each included data year. It contains only emergency department and inpatient encounters identified by the Stimulant Algorithm as meeting one or more case definitions for stimulant use and, to a limited extent, opioid use. **Encounters not identified** by the algorithm are **excluded.** The dataset is limited to encounter identifiers and indicator variables, and does not include additional clinical, demographic, or other encounter-level variables.

The dataset includes variables generated from both the medical code component and the clinical note component of the Stimulant Algorithm. These variables indicate whether an encounter met the algorithm's case definitions for therapeutic stimulant use, non-therapeutic stimulant use, non-therapeutic stimulant and opioid co-use, and related indicators, based on information from medical codes and, where applicable, clinical notes.

The Stimulant Use Dataset is primarily intended to support analyses of hospital encounters associated with stimulant use, though some opioid information is included. To analyze all hospital encounters from an NHCS data year—including those not identified by the Stimulant Algorithm—users may link this dataset to the full NHCS dataset using the linkage variables included within the Stimulant Use Dataset.

Additional details on these linkage variables, and the algorithm-specific variables are provided in upcoming sections. Users interested in studying opioid-related encounters can explore the Enhanced Opioid Identification Dataset, also available through the RDC. An overview of the dataset can be found here: https://www.cdc.gov/nchs/media/pdfs/2024/08/2021-Opioid-RDC-508.pdf (9).

Section 4.1: Stimulant Use Dataset Variables

The following section describes the variables in the Stimulant Use Dataset, including their definitions, format, values, and any notes relevant for interpretation. For clarity, a 'variable' refers to a single data element in the dataset, while an 'indicator' refers to a measure of stimulant or opioid use identified by the algorithm. This distinction is important because indicators may be derived from multiple variables and do not always correspond to a single variable.

The variables in the dataset can be grouped into several categories based on their purpose. These include (1) **component-specific variables,** which represent whether a given indicator is present in the clinical notes or in the medical codes; (2) **summary variables**, which reflect whether an indicator is present, regardless of whether it was identified in the clinical notes or medical codes; (3) **linkage variables** that can be used to merge this dataset with the complete NHCS dataset; and (4) an **auxiliary variable** to support further analysis. The sections below describe each group in more detail, beginning with the component-specific variables which make up much of the dataset.

4.1.1: Component-specific variables

Component-specific variables include a suffix indicating which algorithm component identified the indicator or alternatively, where evidence of that indicator was detected. Variables appended with "_NLP" represent positive cases identified by the clinical note component while variables appended with "_CODE" represent positive cases identified by the medical code component (e.g., **VARNAME_NLP** and **VARNAME_CODE** respectively).

Each indicator represents the outcome of interest (e.g., illicit stimulant use), while the corresponding component-specific variables reflect whether that indicator was identified by a specific component of the algorithm. All indicators have a corresponding "_CODE" variable (from the medical codes), and only some indicators also have a corresponding "_NLP" variable (from the clinical notes). In other words, not all indicators have an associated "_NLP" variable. This is because the clinical note component identifies only a subset of indicators due to challenges in training the models, and some component-specific variables were excluded due to their unreliable performance in testing. More details are available in **Appendix I**.

To determine whether an encounter is positive for a given indicator, users must consider both component-specific variables. A value of 1 in either the "_CODE" or "_NLP" variable indicates that the encounter was flagged by at least one algorithm component, while a value of 1 in both variables indicates that it was flagged by both components – that is, it was found in both the code data and the clinical notes.

For example, to determine whether an encounter is positive for the illicit stimulant use indicator, users should first examine both component-specific variables: that is **STIM_ILLICIT_NLP** and **STIM_ILLICIT_CODE**. An encounter is considered positive if either variable has a value of '1' indicating that it was identified by at least one algorithm component. If both variables have a value of '1' (**STIM_ILLICIT_NLP** = 1 and **STIM_ILLICIT_CODE** = 1), then the encounter was identified for illicit stimulant use by both components with evidence in both the clinical notes and the medical codes.

Component-specific variables allow users to distinguish between indicators identified from medical codes and those identified from clinical notes, supporting analyses that assess the added value of clinical note data relative to coded data sources.

4.1.2: Summary Variables

There are four **summary variables** that represent the primary stimulant and opioid use indicators. These variables are derived from component-specific variables from one or both algorithm components and are not appended with _NLP or _CODE. They represent the presence of an indicator in the record, regardless of whether it was identified in the clinical notes or medical codes. The summary variables are:

- **DRUGSCREEN** (whether a drug screen was performed),
- **OPIOID_ANY_NON_TX** (any non-therapeutic opioid use),
- **STIM_ANY_NON_TX** (any non-therapeutic stimulant use), and
- **STIM_OPIOID_CO_USE** (non-therapeutic stimulant and opioid co-use).

DRUGSCREEN

Can be identified by either algorithm component; however, the variable is defined such that it does not distinguish whether identification occurred in the clinical notes or medical codes.

OPIOID_ANY_NON_TX and **STIM_ANY_NON_TX**

These variables are derived from a logical combination of their "_NLP" and "_CODE" counterparts. A value of 1 indicates that the indicator was identified by either component (or both). For example:

- **OPIOID_ANY_NON_TX** = 1 **IF**:
  - **OPIOID_ANY_NON_TX_NLP** = 1, **OR**
  - **OPIOID_ANY_NON_TX_CODE** = 1

STIM_OPIOID_CO_USE

This variable is derived from the variables **OPIOID_ANY_NON_TX** and **STIM_ANY_NON_TX**. A value of 1 means that both indicators, that is any non-therapeutic opioid use and any non-therapeutic stimulant use, are present. For example:

- **STIM_OPIOID_CO_USE** = 1 **IF**:
  - **OPIOID_ANY_NON_TX** = 1, **AND**
  - **STIM_ANY_NON_TX** = 1

4.1.3: Linkage Variables

Linkage variables enable users to link the Stimulant Use Dataset to the complete NHCS dataset, for the corresponding data year, allowing users to incorporate additional encounter- or hospital-level information from NHCS. These linkage variables are not algorithm-derived, rather they come directly from the underlying NHCS encounter-level data. The variables are **ENCOUNTER_ID** (unique identifier for each hospital encounter), **HOSPID** (unique ID assigned to the hospital providing data), **SETTING** (emergency or inpatient), **FILESOURCE** (data source type), **SUR_YR** (the survey year), and **DATE_BEG** / **DATE_END** (start or end of the encounter). These variables may also be used to subset or stratify analyses by characteristics such as hospital setting, the type of health record, or year.

4.1.4: Auxiliary Variables

The dataset includes the **auxiliary variable** **NLP_ELIGIBLE**, which indicates whether a given encounter has clinical notes available for text-based analysis (i.e., can be analyzed by the clinical component, which uses natural language processing (NLP). Because analyzable clinical notes are available for only a subset of encounters, clinical note component variables — variables ending in "_NLP" — are only meaningfully interpreted when **NLP_ELIGIBLE**=1.

For encounters where **NLP_ELIGIBLE** = 0, a value of '0' for an _NLP variable (e.g., **ILLICIT_COCAINE_NLP** = 0) does not necessarily mean that the clinical note component searched the encounter and found no evidence of cocaine use. Rather, it indicates that no clinical notes were available for analysis. As such, a value of '0' assigned to a clinical note component variable is non-informative for encounters without clinical notes (where **NLP_ELIGIBLE** = 0), because it reflects missing input, not the absence of an indicator. This differs from the code component, where medical codes are present for virtually all encounters.

Data users should restrict analyses involving clinical note component variables (VARNAME_NLP) to encounters where NLP_ELIGIBLE = 1. For example, if a user wants to explore the proportion of encounters with evidence of illicit cocaine use found in the clinical notes, then the numerator should be the number of encounters where **ILLICIT_COCAINE_NLP** = 1, and the denominator should be limited to the number of encounters where **NLP_ELIGIBLE** = 1, not the total number of encounters in the dataset.

Section 4.2: Summary Statistics for Select Indicators

This section presents tables of descriptive statistics summarizing hospital encounters identified by the Stimulant Algorithm for the stimulant and opioid use indicators.

Categories shown in the tables are **not mutually exclusive**; a single hospital encounter may be flagged for more than one indicator. For example, an encounter identified as having both non-therapeutic stimulant use and non-therapeutic opioid use would be counted in each corresponding category. Therefore, counts across categories **should not be summed up** to obtain total numbers of encounters.

A patient encounter could be identified for stimulant or opioid use by either the Stimulant Algorithm's medical code component or its clinical note component. However, for some drug use categories, the clinical note component could not reliably identify cases; therefore, only the medical code component was used for those indicators. The total number of encounters for each indicator reflects the combined set of unique encounters identified by either component (medical codes and/or NLP), where applicable, and by medical codes alone when the clinical note component variable was not available. Encounters identified by both components are counted once.

To support users in understanding how these indicators are constructed and how the components contribute to case identification, a detailed breakdown of encounters by identification method is provided in **Appendix II.** This breakdown is shown for a single data year (2020) for demonstration purposes but reflects the same methodology applied across all data years.

4.2.1: Summary of Primary Indicators, Across Settings

Tables 2 and **3** display the number and percentage of NHCS emergency department and inpatient encounters identified by the Stimulant Algorithm for each stimulant- and opioid-related indicator, by data year. Counts and percentages are shown separately for emergency department and inpatient settings. Percentages are calculated using the total number of NHCS encounters within each hospital setting as the denominator (shown as the last row in the tables).

Table 2. Number and percentage of emergency department encounters identified for the primary indicators: National Hospital Care Survey, 2020 – 2021

Indicator	2020		2021	
	Number	Percent	Number	Percent
Any stimulant use	133,865	1.68	170,348	1.70
Therapeutic stimulant use	27,579	0.35	34,942	0.35
Non-therapeutic stimulant use	108,849	1.37	140,051	1.40
Non-therapeutic opioid use	139,170	1.75	183,914	1.83
Non-therapeutic stimulant & opioid co-use	30,377	0.38	39,995	0.40
Total encounters	7,960,956	100	10,025,499	100

NOTES: Groups are not mutually exclusive; a single encounter may involve use of multiple substances. "Total encounters" represents the total number of NHCS encounters in each hospital setting and serves as the denominator for percentage calculations.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020 - 2021

Table 3. Number and percentage of inpatient encounters identified for the primary indicators: National Hospital Care Survey, 2020 – 2021

Indicator	2020		2021	
	Number	Percent	Number	Percent
Any stimulant use	58,984	2.12	83,014	2.25
Therapeutic stimulant use	2,791	0.10	4,972	0.13
Non-therapeutic stimulant use	56,277	2.02	78,170	2.12
Non-therapeutic opioid use	66,300	2.38	87,567	2.37
Non-therapeutic stimulant & opioid co-use	15,618	0.56	21,232	0.58
Total encounters	2,781,542	100	3,689,208	100

NOTES: Groups are not mutually exclusive; a single encounter may involve use of multiple substances. "Total encounters" represents the total number of NHCS encounters in each hospital setting and serves as the denominator for percentage calculations.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020 - 2021

4.2.2: Stimulant Use Indicators

Tables 4 and **5** report the number of inpatient and emergency department encounters identified by the Stimulant Algorithm, in patient records with documented stimulant use, including non-therapeutic stimulant use (illicit use or prescription misuse) and therapeutic stimulant use.

Table 4. Number of emergency department encounters with documented evidence of stimulant use in the record: National Hospital Care Survey, 2020 – 2021

Indicator	Total Encounters	
	2020	2021
Any stimulant use	133,865	170,348
Non-therapeutic stimulant use		
Any non-therapeutic stimulant use	108,849	140,051
Illicit Stimulant use		
Any Illicit stimulant	77,847	99,268
Cocaine	60,106	80,465
MDMA*	159	174
Methamphetamine	21,220	22,467
Prescription stimulant misuse		
Any prescription stimulant	1,526	1,859
Prescription Amphetamine	1,394	1,697
Prescription Methylphenidate	135	161
Therapeutic stimulant use		
Any therapeutic stimulant	27,579	34,942
Amphetamine	455	587
Amphetamine/Dextroamphetamine	7,113	7,090

Dexmethylphenidate	435	452
Dextroamphetamine	304	312
Lisdexamfetamine	2,285	2,524
Methylphenidate	3,225	3,152

* MDMA is 3,4-methylenedioxymethamphetamine, also commonly known as ecstasy or molly
NOTE: Groups are not mutually exclusive; a single encounter may involve use of multiple substances.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020 - 2021

Table 5. Number of inpatient encounters with documented evidence of stimulant use in the record: National Hospital Care Survey, 2020 - 2021

Indicator	Total Encounters	
	2020	2021
Any stimulant use	58,984	83,014
Non-therapeutic stimulant use		
Any non-therapeutic stimulant use	56,277	78,170
Illicit Stimulant use		
Any Illicit stimulant	37,711	51,383
Cocaine	35,697	47,418
MDMA*	62	78
Methamphetamine	2,199	4,214
Prescription stimulant misuse		
Any prescription stimulant	1,339	1,821
Prescription Amphetamine	1,235	1,714
Prescription Methylphenidate	106	110
Therapeutic stimulant use		
Any therapeutic stimulant	2,791	4,972
Amphetamine	250	295
Amphetamine/Dextroamphetamine	1,317	2,303
Dexmethylphenidate	30	112
Dextroamphetamine	90	122
Lisdexamfetamine	310	745
Methylphenidate	680	1,150

* MDMA is 3,4-methylenedioxymethamphetamine, also commonly known as ecstasy or molly
NOTE: Groups are not mutually exclusive; a single encounter may involve use of multiple substances.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020 – 2021

4.2.3: Opioid Use Indicators

The Stimulant Algorithm also identifies encounters with non-therapeutic stimulant-opioid co-use by searching patient records for documented non-therapeutic opioid use. **Tables 6** and **7** report the

number of emergency department and inpatient encounters identified by the Stimulant Algorithm for non-therapeutic opioid use, including illicit opioid use, prescription opioid misuse, and unspecified non-therapeutic opioid use (i.e., when the medical record does not differentiate between illicit use and prescription misuse).

Table 6. Number of emergency department encounters with documented evidence of non-therapeutic opioid use in the record: National Hospital Care Survey, 2020 - 2021

Indicator	Total Encounters	
	2020	2021
Any non-therapeutic opioid use	139,170	183,914
Illicit opioid use	22,295	34,364
Prescription opioid misuse	4,781	6,873
Unsp non-therapeutic opioid use	89,856	103,697

NOTES: Groups are not mutually exclusive; a single encounter may involve use of multiple substances. Unsp is unspecified.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020 - 2021

Table 7. Number of inpatient encounters with documented evidence of non-therapeutic opioid use in the record: National Hospital Care Survey, 2020-2021

Indicator	Total Encounters	
	2020	2021
Any non-therapeutic opioid use	66,300	87,567
Illicit opioid use	3,080	6,706
Prescription opioid misuse	1,121	1,618
Unsp non-therapeutic opioid use	61,460	77,703

NOTES: Groups are not mutually exclusive; a single encounter may involve use of multiple substances. Unsp is unspecified.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020 - 2021

Section 4.3: Analytical Considerations

When using the Stimulant Use Dataset, several important analytic limitations and implications should be considered. Users should be aware of several important factors related to the interpretation of algorithm-identified cases and underlying limitations in medical code data and clinical notes.

Temporality cannot be distinguished

EHR data does not consistently differentiate between *current* and *past* substance use (4,5). Diagnosis codes and clinical documentation may persist in a patient's record without clear temporal context. As a result, the Stimulant Algorithm identifies **any documented stimulant or opioid use—past or present—**

within a given data year, not necessarily use that occurred during the specific encounter. Users analyzing a single year of data should interpret results cautiously, as identified cases may reflect prior substance use documented in the record rather than use specific to that year. When possible, examining multiple years of data or applying additional analytic restrictions, in the full NHCS dataset, may help users better approximate the timing of substance use.

Medication data are sparsely available in NHCS data

Hospitals may submit either administrative claims data or EHR data to NHCS, but not both. Administrative claims data do not contain medication information; therefore, medication data are available only from hospitals submitting EHR data. Furthermore, not all hospitals submitting EHR data include medication fields. For example, in the 2020 NHCS data used to develop the Stimulant Algorithm, medication data were available for 1,555,470 of the 2,512,902 EHR encounters submitted. Consequently, medication-based indicators, particularly therapeutic stimulant use, are likely underreported. This limitation reflects the structure of NHCS; moreover, therapeutic stimulant use is more commonly documented in ambulatory care settings (e.g., outpatient clinics), which NHCS does not include. Consequently, counts of therapeutic stimulant use should be interpreted with caution, as they likely underestimate the true prevalence of use.

Limited diagnosis codes for prescription misuse

Stimulant misuse codes are sparse, and prescription misuse is often not differentiated in medical codes and are challenging to identify in clinical notes. Counts of prescription stimulant or opioid misuse may be underreported, particularly when misuse is documented only in clinical notes or not documented at all.

Analytic assumptions applied to certain opioid medications

Buprenorphine, methadone, and their analogs were classified as indicators of non-therapeutic opioid use unless documentation clearly indicated therapeutic use. Counts of therapeutic opioid use or prescription opioid misuse involving these substances may be underrepresented, and some counts may reflect misclassification.

Co-use is derived indirectly rather than directly detected

The algorithm does not detect stimulant–opioid co-use directly; instead, co-use is assigned when the encounter is positive for any non-therapeutic stimulant use and any non-therapeutic opioid use. The co-use indicator reflects logical overlap, not a separate independent identification process. As a result, co-use counts may be sensitive to how each component (medical code or clinical note) captures the individual categories.

Section 5: Stimulant Use Dataset Data Dictionary

Variables are grouped as linkage, summary, component-specific, and auxiliary, as described in Section 4.1.

RDC Variable Name	Description	Format	Value	Notes
ENCOUNTER_ID	**Linkage variable.** Encounter identifier	Character		
SETTING	**Linkage variable.** Describes the hospital setting.	Character	"ED" = emergency department "IP" = inpatient department "ED-to-IP" = emergency department visits admitted to the inpatient department	To account for all inpatient or emergency department encounters you must include the encounters with a SETTING value of "ED-to-IP"
FILESOURCE	**Linkage variable.** Describes the source of the data provided by the hospital.	Character	"Claims"* "EHR"† "Vizient"§ "ACEP" ¶	ACEP only includes data from the ED setting. Detailed explanation of each file source is included in the data dictionary footnotes
HOSPID	**Linkage variable.** Arbitrary hospital identifier assigned to each hospital.	Character		
SUR_YR	**Linkage variable.** A 4-digit designation for the year of NHCS data collection	Character	2020, 2021	
DATE_BEG	**Linkage variable.** Start date of encounter	Date		
DATE_END	**Linkage variable.** End date of encounter	Date		
OPIOID_ANY_NON_TX	**Summary variable.** Any non-therapeutic opioid use, that is use of illicit opioids or misuse of prescription opioids.	Numeric	0="False" 1="True"	1="True" when OPIOID_ANY_NON_TX_NLP = 1 **or** OPIOID_ANY_NON_TX_CODE = 1

RDC Variable Name	Description	Format	Value	Notes
STIM_ANY_NON_TX	**Summary variable**. Any non-therapeutic stimulant use, that is use of illicit stimulants or misuse of prescription stimulants.	Numeric	0="False" 1="True"	1="True" when STIM_ANY_NON_TX_NLP = 1 **or** STIM_ANY_NON_TX_CODE = 1
STIM_OPIOID_CO_USE	**Summary variable.** Stimulant and opioid co-use, proxy for the co-occurrence of non-therapeutic stimulant and opioid use in the encounter.	Numeric	0="False" 1="True"	1="True" when OPIOID_ANY_NON_TX =1 **AND** STIM_ANY_NON_TX =1
DRUGSCREEN	**Summary variable.** Whether a drug screen was performed in the form of laboratory testing, verbal or questionnaire screening.	Numeric	0="False" 1="True"	
NLP_ELIGIBLE	**Auxiliary variable.** Indicates whether an encounter has readable clinical notes available for text-based (NLP) analysis.	Numeric	0="False" 1="True"	Analytical support to subset data and define denominators for analysis of clinical note component variables.
STIM_ANY_NON_TX_NLP	**Component-specific variable:** Any non-therapeutic stimulant use, that is use of illicit stimulants or misuse of prescription stimulants, that was identified by NLP analysis of clinical notes.	Numeric	0="False" 1="True"	
STIM_ILLICIT_NLP	**Component-specific variable:** Use of illicit stimulants, such as cocaine or illicitly manufactured methamphetamine, identified by NLP analysis of clinical notes.	Numeric	0="False" 1="True"	
ILLICIT_COCAINE_NLP	**Component-specific variable:** Use of illicit cocaine identified by NLP analysis of clinical notes.	Numeric	0="False" 1="True"	
ILLICIT_METHAMPHETAMINE_NLP	**Component-specific variable:** Use of illicit methamphetamine identified by NLP analysis of clinical notes.	Numeric	0="False" 1="True"	

RDC Variable Name	Description	Format	Value	Notes
STIM_ANY_NLP	**Component-specific variable:** Any stimulant use, non-therapeutic or therapeutic, identified by NLP analysis of clinical notes.	Numeric	0="False" 1="True"	
OPIOID_ILLICIT_NLP	**Component-specific variable:** Use of illicit opioids such as illicitly manufactured fentanyl, and heroin, identified by the NLP analysis of the clinical notes.	Numeric	0="False" 1="True"	
OPIOID_ANY_NON_TX_NLP	**Component-specific variable:** Any non-therapeutic opioid use, that is use of illicit opioids or misuse of prescription opioids that was identified by NLP analysis of clinical notes.	Numeric	0="False" 1="True"	
STIM_ANY_NON_TX_CODE	**Component-specific variable:** Any non-therapeutic stimulant use, that is, use of illicit stimulant or misuse of prescription stimulant, identified by the medical code search.	Numeric	0="False" 1="True"	
STIM_ILLICIT_CODE	**Component-specific variable:** Use of illicit stimulants, such as cocaine or illicitly manufactured methamphetamine, identified by the medical code search.	Numeric	0="False" 1="True"	
ILLICIT_COCAINE_CODE	**Component-specific variable:** Use of illicit cocaine identified by the medical code search.	Numeric	0="False" 1="True"	
ILLICIT_MDMA_CODE	**Component-specific variable:** Use of MDMA identified by the medical code search.	Numeric	0="False" 1="True"	
ILLICIT_METHAMPHETAMINE_CODE	**Component-specific variable:** Use of illicit methamphetamine identified by the medical code search.	Numeric	0="False" 1="True"	

RDC Variable Name	Description	Format	Value	Notes
STIM_MISUSE_CODE	**Component-specific variable:** Misuse of prescription stimulants identified by the medical code search.	Numeric	0="False" 1="True"	
MISUSE_AMPHETAMINE_CODE	**Component-specific variable:** Misuse of prescription stimulants containing Amphetamine identified by the medical code search.	Numeric	0="False" 1="True"	
MISUSE_METHYLPHENIDATE_CODE	**Component-specific variable:** Misuse of prescription stimulants containing Methylphenidate identified by the medical code search.	Numeric	0="False" 1="True"	
STIM_NON_TX_UNSP_CODE	**Component-specific variable:** Unspecified non-therapeutic stimulant use based on the presence of a stimulant class of medication or drug without further classification of the subtype, in a non-therapeutic context, identified by the medical code search.	Numeric	0="False" 1="True"	
STIM_TX_CODE	**Component-specific variable:** Presence of one or more medication codes for prescription stimulants prescribed by a clinical provider in a patient encounter, identified by the medical code search.	Numeric	0="False" 1="True"	
TX_AMPHET_DEXTROAMPHET_CODE	**Component-specific variable:** Therapeutic use of stimulant medications containing Amphetamine and Dextroamphetamine, identified by the medical code search.	Numeric	0="False" 1="True"	

RDC Variable Name	Description	Format	Value	Notes
TX_AMPHETAMINE_CODE	**Component-specific variable:** Therapeutic use of stimulant medications containing Amphetamine, identified by the medical code search.	Numeric	0="False" 1="True"	
TX_DEXMETHYLPHENIDATE_CODE	**Component-specific variable:** Therapeutic use of stimulant medications containing Dexmethylphenidate, identified by the medical code search.	Numeric	0="False" 1="True"	
TX_DEXTROAMPHETAMINE_CODE	**Component-specific variable:** Therapeutic use of stimulant medications containing Dextroamphetamine, identified by the medical code search.	Numeric	0="False" 1="True"	
TX_LISDEXAMFETAMINE_CODE	**Component-specific variable:** Therapeutic use of stimulant medications containing Lisdexamfetamine, identified by the medical code search.	Numeric	0="False" 1="True"	
TX_METHYLPHENIDATE_CODE	**Component-specific variable:** Therapeutic use of stimulant medications containing Methylphenidate, identified by the medical code search.	Numeric	0="False" 1="True"	
STIM_ANY_CODE	**Component-specific variable:** Any stimulant use, non-therapeutic or therapeutic, identified by the medical code search.	Numeric	0="False" 1="True"	
OPIOID_ANY_CODE	**Component-specific variable:** Opioid use identified by the medical code search as either non-therapeutic or therapeutic.	Numeric	0="False" 1="True"	

RDC Variable Name	Description	Format	Value	Notes
OPIOID_ANY_NON_TX_CODE	**Component-specific variable:** Any non-therapeutic opioid use, that is, use of illicit opioids or misuse of prescription opioids, identified by medical code search.	Numeric	0="False" 1="True"	
OPIOID_ILLICIT_CODE	**Component-specific variable:** Use of illicit opioids such as illicitly manufactured fentanyl, and heroin identified by the medical code search.	Numeric	0="False" 1="True"	
OPIOID_MISUSE_CODE	**Component-specific variable:** Misuse of prescription opioids identified by the medical code search.	Numeric	0="False" 1="True"	
OPIOID_NON_TX_UNSP_CODE	**Component-specific variable:** Non-therapeutic opioid use identified by the medical code search, without information to classify the manner of use as either illicit or prescription misuse.	Numeric	0="False" 1="True"	

*Data collected on the uniform bill (UB-04) for institutional providers approved by the National Uniform Billing Committee and is the electronic standard for hospital billing. For more information about UB-04 see https://www.cms.gov/files/document/837I-Form-CMS-1450-MLN006926.pdf

†An electronic version of a patient's medical history maintained by the hospital over time (https://www.cms.gov/EHealthRecords/)

§Vizient is a large provider-driven, health care performance improvement organization. Vizient collects data from the hospitals prior to submitting it to NHCS. Similar to UB-04 claims but includes medication and laboratory data. For more information see: https://www.vizientinc.com/.

¶ ACEP is a United States professional organization of emergency physicians that collects UB-04 claims and EHR data to provide ED quality measures. Data collected from ACEP are similar to Vizient and EHR data, and include diagnoses, medications, laboratory results, and clinical notes. For more information visit: https://www.acep.org/

Section 6: Additional Resources

The following resources provide additional information on methodology, related data products, and the broader National Health Care Surveys program:

- **NCHS Research Data Center (RDC) Application Process**
  Information on how to apply for access to restricted-use NCHS data, including the Stimulant Use Dataset and the complete National Hospital Care Survey dataset.
  https://www.cdc.gov/rdc/application-process/index.html
- **NHCS Questionnaires, Datasets, and Documentation**
  Access point for NHCS data files, documentation, and survey instruments.
  https://www.cdc.gov/nhcs
- **NHCS Drug Use Dashboard (Preliminary Data)**
  Interactive dashboard with preliminary, unweighted estimates of drug use based on NHCS data.
  https://www.cdc.gov/nchs/hcs/data-dashboard.htm
- **Enhanced Opioid Identification Methodology Paper**
  Provides detailed information on methods used to improve identification of opioid use in NHCS data
  https://www.cdc.gov/nchs/data/series/sr_02/sr2-188.pdf
- **National Health Care Surveys Overview**
  A family of surveys that includes NHCS, with information on survey design, scope, and data collection.
  https://www.cdc.gov/nchs/hcs/index.htm

Section 7: References

1. Centers for Disease Control and Prevention. Research Data Center (RDC) [Internet]. Atlanta (GA): CDC; [cited 2026 Mar 26]. Available from: https://www.cdc.gov/rdc/index.html
2. Centers for Disease Control and Prevention (CDC). **National Hospital Care Survey (NHCS)** [Internet]. Atlanta (GA): CDC; [cited 2026 Mar 24]. Available from: https://www.cdc.gov/nchs/nhcs/index.html
3. Centers for Disease Control and Prevention (CDC). **NHCS questionnaires, datasets, and documentation: 2019, 2020, and 2021 National Hospital Care Survey RDC data dictionary** [Internet]. Atlanta (GA): CDC; [cited 2026 Mar 24]. Available from: https://www.cdc.gov/rdc/data/b1/NHCS-RDC-Data-Dictionary.pdf
4. 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
5. 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. https://doi.org/10.1055/s-0041-1731784
6. Centers for Disease Control and Prevention (CDC). **Stimulant opioid algorithm medical codes (SAS)** [Internet]. GitHub; [cited 2026 Mar 24]. Available from: https://github.com/CDCgov/stimulant_opioid_algorithm_medical_codes_SAS
7. Centers for Disease Control and Prevention (CDC). **Stimulant opioid algorithm medical codes (R)** [Internet]. GitHub; [cited 2026 Mar 24]. Available from: https://github.com/CDCgov/stimulant_opioid_algorithm_medical_codes_R
8. Centers for Disease Control and Prevention (CDC). **Stimulant opioid algorithm clinical notes** [Internet]. GitHub; [cited 2026 Mar 24]. Available from: https://github.com/CDCgov/stimulant_opioid_algorithm_clinical_notes
9. Centers for Disease Control and Prevention. 2019–2021 National Hospital Care Survey (NHCS) enhanced opioid identification dataset: research data center (RDC) documentation [Internet]. Atlanta (GA): CDC; [cited 2026 Mar 26]. Available from: https://www.cdc.gov/nchs/media/pdfs/2024/08/2021-Opioid-RDC-508.pdf

Appendix I: Clinical Note Component Variables Excluded for Poor Performance

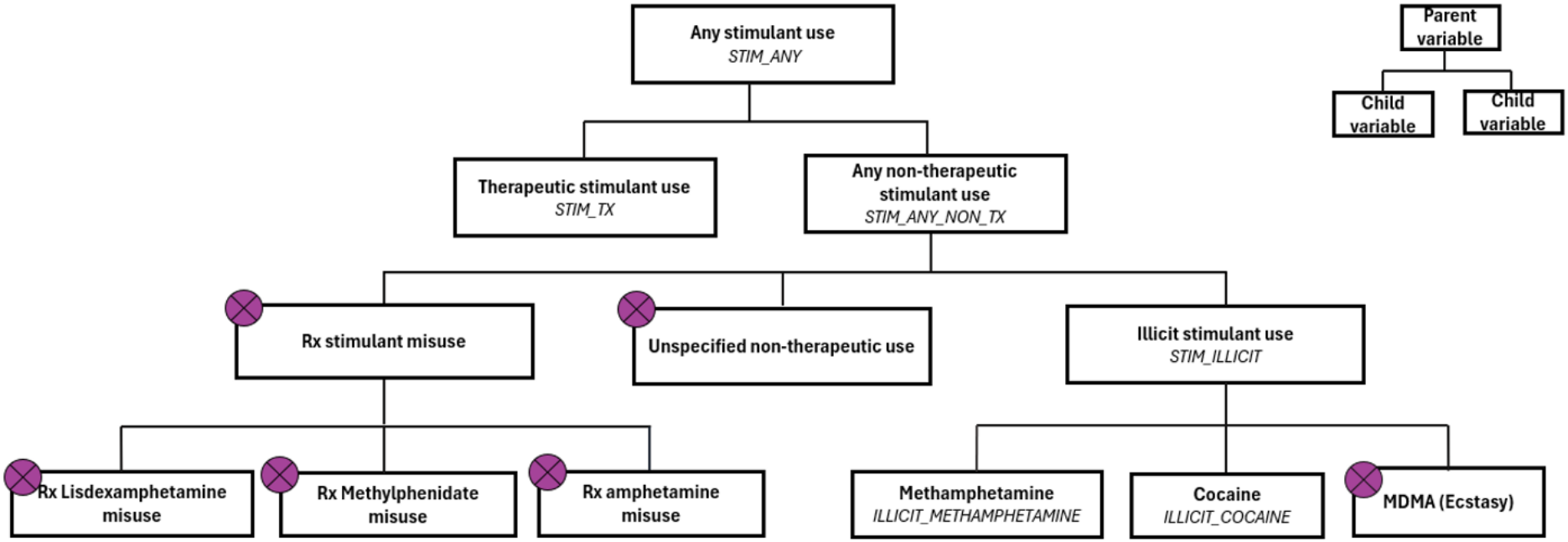
Clinical note component variables (those ending in "_NLP") reflect information derived from a subset of encounters with available clinical notes, rather than the full dataset. In addition, not all potential clinical note component variables are included in the final dataset. During development, candidate variables were evaluated using performance measures and variables that did not meet predefined performance thresholds or had insufficient numbers of positive cases were excluded. As a result, the set of _NLP variables represent both the subset of encounters with analyzable clinical notes and the subset of variables that demonstrated acceptable performance.

Figure 1 displays the hierarchy of stimulant use indicators derived from the clinical note component, beginning with the parent variable stimulant use (**STIM_ANY_NLP**) and its related child-variables. Note that in the figure, the suffix _NLP has been left out from the variable names since the figure is specific to the clinical note component of the algorithm. The figure also identifies clinical note–based variables that were evaluated but excluded from the final dataset due to insufficient performance or low numbers of positive cases. These include variables for prescription stimulant misuse (lisdexamfetamine misuse, methylphenidate misuse, and amphetamine misuse), unspecified non-therapeutic stimulant use, and illicit use of 3,4-methylenedioxymethamphetamine (MDMA, ecstasy). As described above, these variables are component-specific and represent only the clinical note component; the variables retained in the Stimulant Use Dataset should be considered in conjunction with the corresponding medical code variables when deriving final indicator values.

Figure 1 includes only stimulant use–related variables, as opioid identification is primarily used to capture stimulant and opioid co-use. For completeness, clinical note–based variables for prescription opioid misuse and unspecified non-therapeutic opioid use were also evaluated but excluded for reasons like those described above.

The 'any opioid use variable' (**OPIOID_ANY_NLP**) was excluded for different reasons. It originally combined: (1) any type of non-therapeutic opioid use (**OPIOID_ANY_NON_TX_NLP**) (2) other opioid mentions that could not be classified as non-therapeutic use, and (3) encounters identified by a similar variable from the Enhanced Opioid Identification Algorithm (**OPIOID_INVOLVEMENT_NLP**). Because it lacked a corresponding annotated standard in the Stimulant Use gold standard dataset, its performance could not be evaluated. To ensure the Stimulant Use Algorithm can be applied independently across data years, **OPIOID_ANY_NLP** was excluded.

Figure 1. Component-Specific Variables for Stimulant Use from the Clinical Note Component of the Stimulant Algorithm



⊗ Indicates variables dropped from the 'Stimulant Use Dataset' 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 notes component.

Appendix II: Indicator Counts by Identification Method

This section provides a breakdown of the number of encounters identified for each indicator by identification method. Encounters are categorized as identified by the clinical note (NLP) component only, the medical code component only, or both components. The "Total Encounters" column reflects the number of unique encounters identified across components (i.e., the union of NLP and medical code identification), with encounters identified by both counted once. For indicators without an NLP component, encounters appear only in the "Code only" column. These counts correspond directly to the totals presented in **Tables 4-7** in the main text.

To clarify how encounters are classified across identification methods and how counts in **Tables I-IV** are derived, the example below illustrates how component-specific variables contribute to a single stimulant use indicator.

Example Indicator: *illicit stimulant use*
Defined using two component-specific variables:

- **STIM_ILLICIT_CODE** (medical code component)
- **STIM_ILLICIT_NLP** (clinical note component)

Indicator definition:
STIM_ILLICIT = 1 if **STIM_ILLICIT_CODE** = 1 **OR** **STIM_ILLICIT_NLP** = 1

Figure 2. Example of indicator derivation from variables and by identification method.

ENCOUNTER_ID	NLP_ELIGIBLE	STIM_ILLICIT_CODE	STIM_ILLICIT_NLP	STIM_ILLICIT	Identification method
1001	0	1	0	1	Code only
1002	1	0	1	1	NLP only
1003	1	1	1	1	NLP and Code
1004	1	0	0	0	Not counted

Appendix III: Stimulant Use Indicators Counts by Identification Method

Tables I and **II** report the number of emergency and inpatient department encounters identified by each component of the Stimulant Algorithm, in patient records with documented stimulant use, including non-therapeutic stimulant use and therapeutic stimulant use.

Table I: Number of emergency department encounters with recorded stimulant use, by identification method: National Hospital Care Survey, 2020

Indicator	Identification Method			Total Encounters
	NLP* only	Code only	NLP and Code	
Any stimulant use	37,117	81,033	15,715	133,865
Non-therapeutic stimulant use				
Any non-therapeutic stimulant use	25,362	75,902	7,585	108,849
Illicit stimulant use				
Any Illicit stimulant	25,249	47,777	4,821	77,847
Cocaine	11,829	47,042	1,225	60,106
MDMA†	…	159	…	159
Methamphetamine	16,943	607	3,670	21,220
Prescription stimulant misuse				
Any prescription stimulant	…	1,526	…	1,526
Prescription Amphetamine	…	1,394	…	1,394
Prescription Methylphenidate	…	135	…	135
Therapeutic stimulant use				
Any therapeutic stimulant	14,053	5,303	8,223	27,579
Amphetamine	…	455	…	455
Amphetamine/Dextroamphetamine	…	7,113	…	7,113
Dexmethylphenidate	…	435	…	435
Dextroamphetamine	…	304	…	304
Lisdexamfetamine	…	2,285	…	2,285
Methylphenidate	…	3,225	…	3,225

… Category not available

*Natural language processing – encounters flagged by the clinical note component

†MDMA is 3,4-methylenedioxymethamphetamine, also commonly known as ecstasy or molly

NOTE: Groups are not mutually exclusive; a single encounter may involve use of multiple substances.

SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020

Table II: Number of inpatient encounters with documented evidence of stimulant use in the record, by identification method: National Hospital Care Survey, 2020

Indicator	Identification Method			Total Encounters
	NLP* only	Code only	NLP and Code	
Any stimulant use	2,389	52,911	3,684	58,984
Non-therapeutic stimulant use				
Any non-therapeutic stimulant use	2,384	52,542	1,351	56,277
Illicit stimulant use				
Any Illicit stimulant	2,429	34,402	880	37,711
Cocaine	1,183	34,351	163	35,697
MDMA†	…	62	…	62
Methamphetamine	1,464	-	735	2,199
Prescription stimulant misuse				
Any prescription stimulant	…	1,339	…	1,339
Prescription Amphetamine	…	1,235	…	1,235
Prescription Methylphenidate	…	106	…	106
Therapeutic stimulant use				
Any therapeutic stimulant	62	369	2,360	2,791
Amphetamine	…	250	…	250
Amphetamine/Dextroamphetamine	…	1,317	…	1,317
Dexmethylphenidate	…	30	…	30
Dextroamphetamine	…	90	…	90
Lisdexamfetamine	…	310	…	310
Methylphenidate	…	680	…	680

… Category not available
-Zero quantity
*Natural language processing – encounters flagged by the clinical note component
†MDMA is 3,4-methylenedioxymethamphetamine, also commonly known as ecstasy or molly
NOTE: Groups are not mutually exclusive; a single encounter may involve use of multiple substances.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020

Appendix IV: Opioid Use Indicators Counts by Identification Method

Tables III and **IV** report the number of emergency department and inpatient encounters identified by each component of the Stimulant Algorithm for non-therapeutic opioid use, including illicit opioid use, prescription opioid misuse, and unspecified non-therapeutic opioid use (i.e., when the medical record does not differentiate between illicit use and prescription misuse).

Table III. Number of emergency department encounters with documented evidence of non-therapeutic opioid use in the record, by identification method: National Hospital Care Survey, 2020

Indicator	NLP* only	Code only	NLP and Code	Total Encounters
Any non-therapeutic opioid use	39,312	91,096	8,762	139,170
Illicit opioid use	13,291	8,415	589	22,295
Prescription opioid misuse	…	4,781	…	4,781
Unsp non-therapeutic opioid use	…	89,856	…	89,856

… Category not available
*Natural language processing – encounters flagged by the clinical note component
NOTES: Groups are not mutually exclusive; a single encounter may involve use of multiple substances. Unsp is unspecified.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020

Table IV: Number of inpatient encounters with documented evidence of non-therapeutic opioid use in the record, by identification method: National Hospital Care Survey, 2020

Indicator	NLP* only	Code only	NLP and Code	Total Encounters
Any non-therapeutic opioid use	3,517	59,870	2,913	66,300
Illicit opioid use	1,083	1,939	58	3,080
Opioid misuse	…	1,121	…	1,121
Unsp non-therapeutic opioid use	…	61,460	…	61,460

… Category not available
*Natural language processing – encounters flagged by the clinical note component
NOTES: Groups are not mutually exclusive; a single encounter may involve use of multiple substances. Unsp is unspecified.
SOURCE: National Center for Health Statistics, National Hospital Care Survey, 2020