Identifying Stimulant and Opioid abuse in Clinical Notes

Rihem Badwe, PharmD
Nikki Adams, PhD
Division of Health Care Statistics
October 24, 2024

The findings and conclusions in this presentation are those of the authors and do not necessarily present the official position of the Centers for Disease Control and Prevention.

Agenda

- **Project Introduction/objectives**
- **Algorithm Development**
  - Case Definitions
  - Annotation
  - Code and Machine Learning/Natural Language Processing Components
  - Research File
- **Algorithm Application to External Datasets**
  - Portability testing
  - Dissemination

Acknowledgment

- Support was provided for the development of the Stimulant Algorithm by the Department of Health and Human Services' Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) under the project titled "Utilizing Natural Language Processing and Machine Learning to Enhance the Identification of Stimulant and Opioid-Involved Health Outcomes in the National Hospital Care Survey."

Background

- Substance abuse remains a significant national public health concern focused on the increase in stimulant use and co-use of stimulants and opioids.
- The National Center for Health Statistics' (NCHS) National Hospital Care Survey (NHCS) provides national estimates on hospital patient care related to drug use in the United States.
- The 2020 NHCS sample of 608 non-institutional non-federal hospitals with six or more staffed inpatient beds across the United States.

Background cont.

- NHCS is a longitudinal survey that, in 2020 alone, collected approximately 2.5 million inpatient and 7.5 million emergency department (ED) records from 205 hospitals.
- Hospitals participating in the NHCS can submit a year's worth of hospitals encounters by one of two data sources:
  - 1) Uniform Billing–04 (UB-04) administrative claims data or
  - 2) electronic health records (EHR) data.
- Submitted data for all hospitals includes medical codes (e.g., diagnosis and procedure codes), however, the free text clinical notes are only submitted by a subset of the hospitals that submit EHR data.

Project goal

- To develop an enhanced algorithm to improve identification of patients with stimulant-involved hospital visits in the 2020 National Hospital Care Survey.
- The algorithm had 2 components: the code and Natural Language Processing(NLP)/Machine Learning (ML) components.
- Since medical codes provide limited information about the context of a patient's hospital visit, the algorithm used NLP and ML techniques to analyze free text clinical notes and searches of medical codes in a patient's encounter.
- To verify the NLP methods, an annotated gold standard dataset with high quality data was developed.

FY18 & FY19 PCORTF Project Objectives

- FY18 algorithm objective was the development of an enhanced algorithm that utilizes multiple data elements in the National Hospital Care Survey (NHCS) and methods to identify presence of opioid-involved hospital encounters and opioid overdoses.

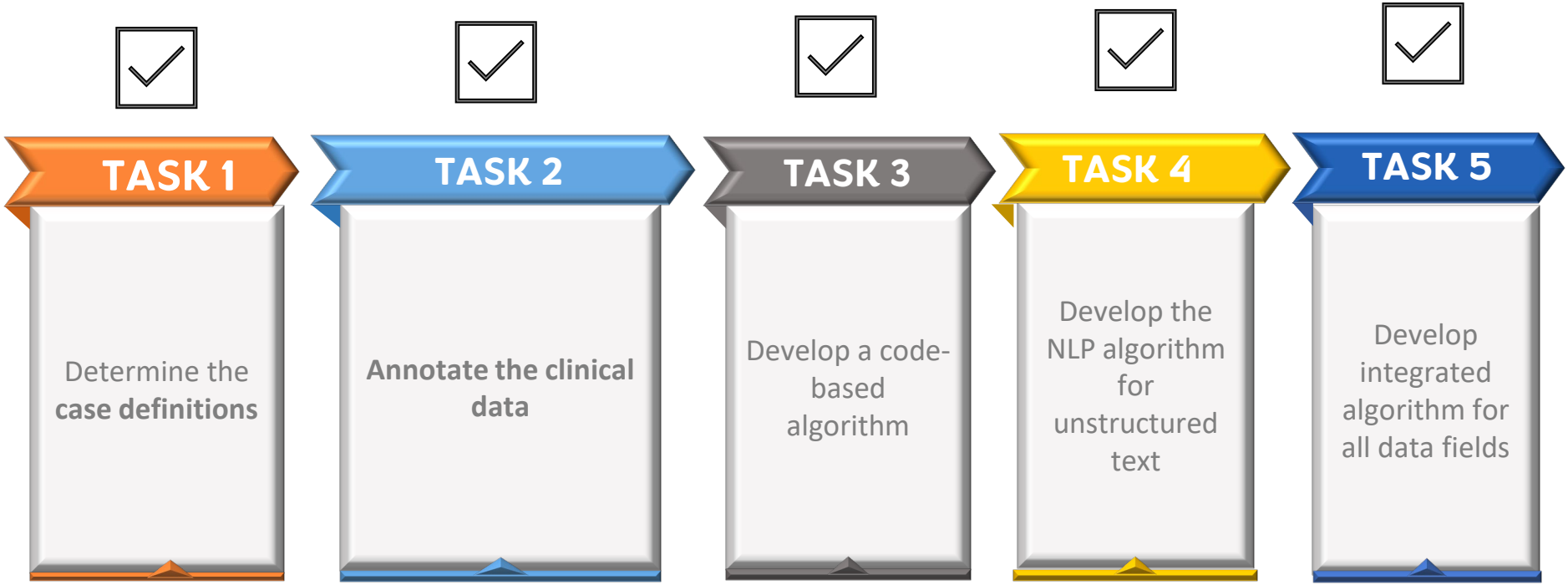
- FY19 project objective was a capstone to the PCORTF FY18 project to improve identification on co-occurring substance use disorders (SUD) and selected mental health issues (MHI) (anxiety and depressive disorders) that utilizes multiple data sources and methods. Also to conduct study to validate FY18 & FY19 algorithms to identify opioids and co-occurring disorders in NHCS data.

FY23 PCORTF Project Objectives

- The goal was to develop an algorithm using medical codes and NLP techniques to identify hospital encounters involving the use of illicit stimulants, misuse of prescription stimulants, and the co-involvement of stimulants and illicit or misused prescription opioids.

Algorithm Development

Tasks for developing the stimulant algorithm



Stimulant or Opioid-involved Case Definitions

- **Therapeutic Use:** Therapeutic use is defined as taking a stimulant or opioid medication as prescribed or directed by a healthcare provider.

- **Non-therapeutic use**: Non-therapeutic use includes illicit use, prescription misuse and unspecified non-treatment use of stimulants or opioids.

- **Illicit Use**: illicit use is defined as the use of a stimulant or opioid classified as either Schedule I medication (ex. MDMA/ecstasy or heroin) or illicitly manufactured Schedule II drugs such as (methamphetamine or fentanyl).

Stimulant or Opioid-involved Case Definitions cont.

- **Prescription misuse**: prescription misuse is defined as the use of prescription drugs in any manner not directed by a doctor, including the use of diverted prescriptions, use in greater amounts, more often, or longer than prescribed.

- **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.

Annotation

What is Annotation?

- To aid in the development of the algorithm, a “gold standard” dataset needed to be created. The gold standard established the ground truth for areas of interest for a subset of data. It can provide labeled examples of positive and negative cases that can be used:
  - To develop algorithm rules (rule-based algorithm)
  - As labeled input for supervised machine learning
  - To provide test cases with which to ultimately evaluate performance of algorithm output against expert human determinations

Annotation Pretest Process

- **WHO**: Two clinically trained annotators.

- **SAMPLE** Encounters in the 2020 NHCS dataset, identified through NLP as having ICD-10-CM codes related to stimulant or opioid use, and/or key words indicating stimulant/opioid use in the clinical notes (text fields).

- **GOALS**
  - Establish an inter-annotator reliability to ensure agreement during the annotation process.
  - Review and revise the Annotation Form to ensure all relevant information is captured.

Annotation Sample

- The aim was to collect 1,000 annotated encounters
- There were 3 groups/buckets to target for sampling:
  - Encounter notes that contain stimulant terms only (450)
  - Encounter notes that contain stimulant and opioid terms (450)
  - Encounter notes that contain neitherstimulant or opioid terms (100)
- An Annotation form was developed by the clinicians

Sample of Annotation Form Questions

Question 3A	Response
Is there evidence of illicit stimulant use in the chart?	❑ Yes (proceed below) ❑ No (proceed to question 3B)

Question 3A	Response
If yes, how many substances were mentioned?	❑ One ❑ Two ❑ Three ❑ Four

Question 3A	Response
Where was the location of patient's stimulant use in the chart? Select all that apply:	Location of this information (*select all that apply*): ❑ Labs ❑ Medications ❑ Procedures ❑ Revenue Codes ❑ Diagnoses ▪ Diagnosis #1 ▪ Diagnosis #2 ▪ Diagnosis #3 ▪ Diagnosis #4 ❑ Notes ▪ Note Event # ▪ Note Type

Sample of Annotation Form Questions Cont.

Question 3B	Response
Is there evidence of prescription stimulant misuse in the chart?	❑ Yes (proceed below) ❑ No (proceed to question 3C)

Question 3B	Response
If yes, how many substances were mentioned?	❑ One ❑ Two ❑ Three ❑ Four

Question 3B	Response
Where was the location of patient's stimulant use in the chart? Select all that apply:	Location of this information (*select all that apply*): ❑ Labs ❑ Medications ❑ Procedures ❑ Revenue Codes ❑ Diagnoses ▪ Diagnosis #1 ▪ Diagnosis #2 ▪ Diagnosis #3 ▪ Diagnosis #4 ❑ Notes ▪ Note Event # ▪ Note Type

Sample of Annotation Form Questions Cont.

Question 3B	Response
What was the clinical indicator of misuse?	❑ Explicit mention of prescription misuse ❑ Evidence of prescription diversion ❑ Mention of drug seeking behavior ❑ Written lab result ❑ Other (Specify)

Question 3C	Response
Was the nature of the patient's stimulant use described as therapeutic?	❑ Yes (proceed Below) ❑ No (proceed to question 4A)

Question 3C	Response
If yes, how many substances were mentioned?	❑ One ❑ Two ❑ Three ❑ Four

Sample of Annotation Form Questions

Question 4A	Response
Is there evidence of illicit opioid use in the chart?	❑ Yes (proceed below) ❑ No (proceed to question 3B)

Question 4A	Response
If yes, how many substances were mentioned?	❑ One ❑ Two ❑ Three ❑ Four

Question 4A	Response
Where was the location of patient's opioid use in the chart? Select all that apply:	Location of this information (*select all that apply*): ❑ Labs ❑ Medications ❑ Procedures ❑ Revenue Codes ❑ Diagnoses ▪ Diagnosis #1 ▪ Diagnosis #2 ▪ Diagnosis #3 ▪ Diagnosis #4 ❑ Notes ▪ Note Event # ▪ Note Type

Code and Machine Learning/Natural Language Processing Components

Code Component

- The code list was finalized
  - In the process of associating research file variables with codes
    - ICD-10-CM, SNOMED CT: Diagnostic data
    - RXNORM: Medication data
    - HCPCS & CPT: Procedures and services data
    - LOINC: Laboratory data

- Keyword search terms were key terms or supportive evidence found in the clinical notes to be used for the text base analysis as potential candidates for identifying cases.

Machine Learning and Natural Language Processing Component

- The NLP component was designed to extract information from the clinical notes associated with an encounter.

- EHR data were provided either directly by the participating hospital or a third party. The third party did not contain any clinical notes.

Stimulant Algorithm Research file

- The full algorithm creates a research file that will identify NHCS hospital encounters involving:
  - therapeutic stimulant use
  - misuse of prescription stimulants/opioids
  - use of illicit stimulants/opioids.

- Code component uses medical codes.
  - Flags **all** the variables in the research file

- The NLP component searches the free-text clinical notes.
  - Flags a **subset** of the variables in the research file

Products coming soon

- **Project-specific dissemination deliverables**
  - Restricted data file in the National Center for Health Statistics and Federal Research Data Center (early 2025).
  - Algorithm code on CDC GitHub repository (expected early 2025).
  - Interactive drug-related visit data with weighted estimates (expected March 2025)
- **Two NCHS Vital and Health Statistics Series 2 methodology reports**
  - Stimulant algorithm methodology (Expected publication: early 2025)
  - Methods and findings from external application of the algorithm (Expected publication: March of 2025)

Thank you for your attention and feedback!

Presenter Contact Information

- Rihem Badwe – Rihem.Badwe@cdc.hhs.gov
- Nikki Adams – Nikki.Adams@cdc.hhs.gov

For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov