

**Office of the Secretary Patient-Centered Outcome Research Trust Fund Project  
(IAA #750019PE090039)**

**National Center for Health Statistics  
Centers for Disease Control and Prevention**

**Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and  
Mortality Data: Capstone to an Existing FY18 PCORTF Project**

**FINAL REPORT**

**October 2021**

## Table of Contents

<b>1. Executive Summary</b> .....	3
<b>2. Background</b> .....	3
<b>3. Major Accomplishments</b> .....	6
<b>4. Lesson Learned</b> .....	9
<b>5. Publications and Presentations</b> .....	9
<b>6. Future Considerations</b> .....	10
<b>7. Summary</b> .....	10
<b>8. How to request linked NHCS data</b> .....	10
<b>9. References</b> .....	10

# **Identifying Co-Occurring Disorders among Opioid Users Using Linked Hospital Care and Mortality Data: Capstone to an Existing FY18 PCORTF Project**

## **FINAL REPORT**

### **1. Executive Summary**

The National Center for Health Statistics (NCHS) houses a unique data resource, the National Hospital Care Survey (NHCS), which provides statistics on health care and hospital utilization based on patient hospital inpatient and emergency department (ED) visits through the collection of administrative claims records and electronic health records (EHRs). The data are unweighted and not nationally representative. With the collection of EHR data and funding from the Patient-Centered Outcome Research Trust Fund (PCORTF), NCHS developed the Enhanced Opioid Identification Algorithm to identify opioid-involved hospital visits that utilize all available data items (1).

Using methodology similar used in the development of the Enhanced Opioid Identification Algorithm, this project developed a Co-occurring Disorders Algorithm to identify opioid-involved hospital visits with co-occurring substance use disorders (SUDs) and mental health issues (MHIs) in the 2016 NHCS data. To identify these hospital visits the algorithm analyzed medical codes (e.g., diagnoses, procedures) and used data science techniques such as regular expression operations and natural language processing (NLP) to determine the occurrence of an event by analyzing all available structured and unstructured data.

Additionally, this project involved the fielding of a validation study that involved manual abstraction of medical records from a sample of 2016 NHCS inpatient and ED encounters to assess, evaluate, and verify the performance of the Enhanced Opioid Identification Algorithm and Co-occurring Disorder Algorithm. The medical abstraction involved accessing the hospital's medical record system and completing a questionnaire that collected information on the presence of selected opioids/opioid antagonists, substance use disorders, and mental health issues.

### **2. Background**

#### **2.a. Goal**

This project aimed to develop a new algorithm that utilizes clinical data collected in the NHCS to identify opioid-involved hospital encounters with presence of co-occurring SUDs and MHIs and to conduct a validation study to assess performance of opioid use algorithms (FY18 PCORTF-funded project) and co-occurring disorder algorithms (FY19 PCORTF-funded project). This report summarizes the FY19 PCORTF project's goals, milestones, accomplishments, lessons learned, and results.

#### **2.b. National Hospital Care Survey**

The NHCS is one of the National Health Care Surveys, a family of surveys covering a wide spectrum of health care delivery settings from ambulatory and outpatient, to hospital and long-term care providers (<https://www.cdc.gov/nchs/dhcs/index.htm>). The NHCS collects hospital inpatient, ED, and outpatient department (OPD) visit level data. The data are unweighted and not

nationally representative. The NHCS is designed to collect information on the characteristics of inpatient hospitalizations and ED encounters including length of stay of inpatient encounters, diagnoses, surgical and non-surgical procedures, and patterns of hospital utilization in various regions of the country (2). The target universe for NHCS is all inpatient discharges and ED visits and OPD visits made to noninstitutional, nonfederal hospitals that have 6 or more staffed inpatient beds. Data are extracted from hospital billing or EHR systems and then transmitted electronically directly to NCHS or its designated agent. The 2016 NHCS collected EHR data in two formats: Custom Extracts and Continuity of Care Documents (CCD) generated by EHR systems. A unique feature of the survey is that it collects patient personally identifiable information (PII), which allows researchers to follow patients who have multiple hospital encounters and link patients to external data sources such as the National Death Index (NDI). In the 2016 NHCS, 158 hospitals submitted data: 89 hospitals submitted UB-04 administrative claims data, 16 hospital custom extracts of EHR data, 31 hospitals submitted EHR data in the form of CCD, and 22 hospitals submitted data via Vizient, a third-party source. The 2016 NHCS includes 7,032,304 ED and 2,591,722 inpatient encounters.

The presence of data fields varies for each of the three original sources of data. The UB-04 administrative claims data contain information on patient demographics, patient identifiers, conditions, services, and discharge status. EHR data include similar data items available in UB-04 administrative claims as well as additional items that provide more detail about a patient's hospital encounter, including medications, clinical notes, and laboratory results. Diagnosis codes (e.g., ICD-10-CM diagnoses codes) and discharge status are often missing from EHR records, but this information is often available in EHR text data fields. Vizient collects UB-04 administrative claims and obtains data on medications and laboratory tests but does not include patient identifiers to NCHS, and thus, cannot be linked to outside data sources. More information on NHCS methodology is published elsewhere (2).

The Co-occurring Disorders Algorithm was designed to complement the Enhanced Opioid Identification Algorithm, which was developed for a previous PCORTF project to identify opioid-involvement in the 2016 NHCS ED and inpatient encounters. In FY18, the Office of the Secretary-PCORTF awarded NCHS funding to develop the Enhanced Opioid Identification Algorithm to improve the identification of opioid-involved hospital encounters in the NHCS. The Enhanced Opioid Identification Algorithm identifies opioid-involved hospital encounters utilizing all available data fields in hospital billing and electronic health record (EHR) data collected in NHCS. In the 2016 NHCS, the Enhanced Opioid Identification Algorithm identified 1,370,827 opioid-involved encounters ED and inpatient encounters from 9,624,026 encounters collected (3).

## 2.c. Tasks, Objectives, and Deliverables

This section outlines the tasks, the objectives, and the deliverables.

**Table 1. Tasks, Objectives, and Deliverables**

<b>Task</b>	<b>Objective</b>	<b>Deliverables</b>
Task 1	<ul style="list-style-type: none"> <li>Development of a new set of algorithms that utilizes multiple data sources and methods to identify presence of co-occurring disorders</li> </ul>	<ul style="list-style-type: none"> <li>Report describing integrated algorithm, including criteria, medical code and search term lists, and fields to be searched. Forthcoming NCHS publication to be published in January 2022</li> </ul>
Task 2	<ul style="list-style-type: none"> <li>Conduct validation study to assess performance of opioid use algorithms (FY18 PCORTF-funded project) and co-occurring disorder algorithms</li> </ul>	<ul style="list-style-type: none"> <li>Report summarizing findings from the validation study, including methodology, algorithm performance measures, and recommended algorithm applications. Completed in October 2021.</li> </ul>
Task 3	<ul style="list-style-type: none"> <li>Apply co-occurring disorders algorithms to 2016 NHCS file</li> </ul>	<ul style="list-style-type: none"> <li>New data file containing 2016 NHCS inpatient and ED data with identification of opioid-involved encounters with co-occurring substance use disorders and mental health issues available to researchers through the NCHS and Federal Research Data Centers (RDCs). <a href="https://www.cdc.gov/nchs/data/nhcs/FY19-RDC-2021-06-01-508.pdf">https://www.cdc.gov/nchs/data/nhcs/FY19-RDC-2021-06-01-508.pdf</a></li> </ul>
Task 4	<ul style="list-style-type: none"> <li>Dissemination and promotion of validated algorithms and analytic findings to the researchers and other users</li> </ul>	<ul style="list-style-type: none"> <li>Dissemination plan that includes a citation list of research using the enhanced opioid-involved encounters with co-occurring disorders data.</li> <li>Quarterly reports to ASPE.</li> <li>Submitted final report to ASPE.</li> </ul>

### **3. Major Accomplishments**

#### **3.a. Enhanced Opioid Encounters with Co-Occurring Disorder Algorithm**

The enhanced Co-occurring Disorders Algorithm was designed to identify opioid-involved encounters with a mention of MHI and SUDs in the 2016 NHCS inpatient and ED data.

An interdisciplinary team of clinicians and researchers under the guidance of the Technical Expert Panel (TEP) selected the MHIs that are identified by the enhanced algorithm. The TEP included representatives from several federal agencies, including NCHS, Food and Drug Administration, Substance Abuse and Mental Health Services Administration, and National Institute on Drug Abuse. The algorithm identifies MHIs from a diagnosis code or diagnostic language in clinical notes. The MHIs identified by the enhanced algorithm are:

- Anxiety
- Depression
- Obsessive compulsive disorder (OCD)
- Self-harm
- Trauma and stressor related disorders

In the identification of SUDs there must be mention of the individuals who abuse a substance, are dependent on a substance, have or had a disease from substance abuse or dependence, and/or have encountered counseling or medical advice related to substance abuse or dependence. The SUDs identified by the enhanced algorithm are:

- Alcohol use disorder
- Cannabis use disorder
- Cocaine use disorder
- Hallucinogen use disorder
- Inhalant use disorder
- Nicotine use disorder
- Opioid use disorder
- Sedative use disorder
- Other stimulants use disorder
- Other psychoactive use disorder

The enhanced Co-occurring Disorders Algorithm developed by this project consists of two components. The first component utilizes data associated with medical codes and the second component utilizes NLP techniques on clinical notes. The two-component approach allowed for efficiently identifying opioid-involved encounters, utilizing all available information collected in the NHCS. The code-based component of the enhanced algorithm searched for MHI and SUD encounters using diagnosis and procedure codes. The NLP portion of the algorithm is designed to extract information from the clinical notes associated with the EHR portion of the NHCS.

### 3.b. Enhanced 2016 NHCS Datafile with Information on Opioid-Involved Encounters with Co-occurring Disorders

The enhanced Co-occurring Disorders Algorithm was used to create a new datafile available to researchers through the NCHS Research Data Center (RDC) or Federal Statistical Research Data Centers (FSRDCs). Table 2 reports the number and percentage of the 1,370,827 opioid-involved encounters in the 2016 NHCS inpatient and ED data that were identified by the enhanced Co-occurring Disorders Algorithm as having a mention of only a MHI, only a SUD, or both MHI and SUDs.

**Table 2. Number and percentage of 2016 NHCS opioid-involved encounters with Mental Health Issues and Substance Use Disorders**

	<b>Number of Encounters</b>	<b>Percent of Opioid-Involved Encounters</b>
Mental health issue – only	100,102	7.3
Substance use disorder - only	399,632	29.2
Mental health issue and substance use disorder	159,491	11.6
No mental health issue or substance use disorder	711,602	51.9

NOTES: Based on 1,370,827 opioid-involved encounters. Data are not nationally representative.  
SOURCE: 2016 National Hospital Care Survey

Table 3 reports on the number of opioid-involved hospital encounters that had specific MHIs identified in the 2016 NHCS data. Anxiety and depression were the most prevalent MHI identified and were mentioned in 11.6% and 11.2% of opioid involved encounters, respectively.

**Table 3. Number and percentage of 2016 NHCS opioid-involved encounters with specific Mental Health Issues**

<b>Mental Health Issues<sup>1</sup></b>	<b>Number of Encounters</b>	<b>Percent of Opioid-Involved Encounters</b>
Anxiety	159,476	11.6
Depression	153,548	11.2
Obsessive compulsive disorder	2,331	0.2
Self-harm	25,079	1.8
Trauma and stressor related disorders	19,392	1.4

NOTES: Based on 1,370,827 opioid-involved encounters. Data are not nationally representative.  
SOURCE: 2016 National Hospital Care Survey

<sup>1</sup> Categories are not mutually exclusive.

The number of opioid-involved encounters with a specific SUD are shown in Table 4. Nicotine use disorder was the most prevalent SUD identified.

**Table 4. Number and percentage of 2016 NHCS Opioid-Involved encounters with specific Substance Use Disorders.**

<b>Substance Use Disorders<sup>1</sup></b>	<b>Number of Encounters</b>	<b>Percent of Opioid-Involved Encounters</b>
Alcohol use disorder	62,325	4.5
Cannabis use disorder	17,306	1.3
Cocaine use disorder	22,873	1.7
Hallucinogen use disorder	1,746	0.1
Inhalant use disorder	1,417	0.1
Nicotine use disorder	482,734	35.2
Opioid use disorder	111,442	8.1
Sedative use disorder	12,883	0.9
Other stimulant use disorder	10,749	0.8
Other psychoactive use disorder	24,249	1.8

NOTES: Based on 1,370,827 opioid-involved encounters. Data are not nationally representative.

SOURCE: 2016 National Hospital Care Survey

<sup>1</sup> Categories are not mutually exclusive.

### **3.c. Validation Study for the Enhanced Opioid Identification Algorithm and the Co-occurring Disorders Algorithms**

This project included a validation study to assess and verify the performance of the Enhanced Opioid Identification Algorithm and the Co-occurring Disorders Algorithm. Nine hospitals that submitted data to the 2016 NHCS were recruited to participate in the study. Once the hospitals were recruited a sample of 1,350 encounters was selected, 150 from each sampled hospital. The budgeted time for the abstraction was limited to abstracting 100 encounters from each hospital. The sampled encounters included an oversample of 50 encounters for each hospital to ensure that a total of 100 encounters would be abstracted from each hospital.

Data abstraction started in February 2021 and ended in June 2021. Abstraction was originally scheduled to be completed in April 2021, but it was extended two months due to the COVID-19 pandemic as restrictions in put in place at participating hospitals delayed/impacted the ability to abstract data. A total of 900 encounters were abstracted. Encounters from the sampled hospitals were abstracted from the hospital’s medical record system, and clinically trained abstractors collected information on mentions of selected opioids/opioid antagonists, SUDs, and MHIs by using a questionnaire specifically designed for this project.

Of the 900 encounters abstracted, 865 were identified by the Enhanced Opioid-Identification Algorithm as opioid-involved and the remaining 35 encounters were identified as not opioid-involved. The Co-occurring Disorders Algorithm identified 320 as SUD-involved and 207 encounters as MHI-involved. Of the 865 encounters flagged by the Enhanced Opioid Identification Algorithm, abstractors found evidence of opioid-involvement in 803 encounters (92.8%). The findings from this study will be used to improve both sets of algorithms. As a next step, NHCS staff will comprehensively review all abstracted data to confirm opioid use and the presence of SUD and MHI for each encounter based on case definitions for both algorithms. Final algorithm performance metrics will also be calculated and presented in a detailed NCHS report describing methodology and how the results can be used to improve the Enhanced Opioid

Identification Algorithm and Co-occurring Disorders Algorithm. The report will include recommendations for refinements for both algorithms based on these findings. Validated algorithms will then be applied to upcoming linked hospital and mortality data to create an enhanced research dataset.

#### 4. Lesson Learned

Several critical lessons were learned from this project. First, the development of the Co-occurring Disorders Algorithm highlighted the importance of continuing to collect high quality EHR data in the NHCS. The information available in the EHR clinical notes helped identify specific SUDs and MHIs for opioid-involved encounters that could not be identified by medical codes alone. Additionally, the project showed the value of conducting a validation study to further refine and improve the performance of the Enhanced Opioid Identification Algorithm and the Co-occurring Disorders Algorithm. The information collected in the validation study will be used in future iterations of the algorithms to produce more accurate information on opioid-involved hospital encounters.

#### 5. Presentations and Dissemination Activities

The presentations that used the data produced by this project are listed below. The project was also mentioned in presentations that were focused on the development of the Enhanced Opioid Identification Algorithm that will be documented in the FY18 PCORTF final report.

- **January 8, 2020** – Carol DeFrances made a presentation entitled “Identification of Opioid Involved Health Outcomes Using Linked Hospital Care and Mortality Data” at the International Conference on Health Policy Statistics, San Diego, CA. <https://ww2.amstat.org/meetings/ichps/2020/onlineprogram/AbstractDetails.cfm?AbstractID=306607>
- **April 14, 2020** – Amy Brown and Merianne Spencer (both NCHS) presented virtually at the Rx Drug Abuse and Heroin Summit. <https://www.rx-summit.com/>
- **January 12, 2021** –Carol DeFrances and Geoff Jackson made a virtual presentation entitled, “Overview of the National Hospital Care Survey and FY18-FY19 Patient Centered Outcomes Research Trust Fund Projects” to the ASPE Provisional Drug Overdose Monthly Briefing.
- **May 17, 2021** – Carol DeFrances made a virtual presentation entitled, “Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data” during the ASPE Webinar entitled “Building the Data Infrastructure to Address the Opioid Epidemic Through the OS-PCORTF”.
- **July 16, 2021** – Nikki Adams (NCHS) presented virtually at the CORE planning team for interagency NLP and machine learning, titled “NLP Applications in NCHS’ Patient Centered Outcomes Research Trust Fund (PCORTF) FY18 and FY19 Projects on Opioid-Involved Hospitalizations”.

The code used to develop and implement the Co-occurring Disorders Algorithm will be published and made publicly available on CDC’s GitHub site (<https://github.com/CDCgov>). GitHub is a website that allows developers to post their code to be utilized by other developers and analysts. The code will be made available in the Python programming language and will

include all the medical codes and keywords used in the algorithm. In addition to the code the site will host documentation to assist users in adapting and using the code. The code will be posted by the end of 2021.

## **6. Future Considerations**

This project built upon the work completed by the FY18 PCORTF project, “Enhancing Identification of Opioid-Involved Health Outcomes Using Linked Hospital Care and Mortality Data” with the development of the Co-occurring Disorders Algorithm. The Co-occurring Disorders Algorithm successfully identified SUDs and MHIs among opioid-involved encounters in the 2016 NHCS data. This work will lead to further opportunities to continue the improved identification of opioid-involved encounters with co-occurring disorders in upcoming NHCS data releases. Additionally, the results from the validation study will be used to refine the enhanced algorithms and improve upon the reporting of hospital utilization and care for opioid-involved hospital visits.

## **7. Summary**

This project accomplished the goal of creating the Co-occurring Disorders Algorithm to enhance the identification of the opioid-involved encounters with co-occurring SUDs and MHIs providing new sources of data for the PCOR community. The methods developed and subsequent improvements to the enhanced algorithms will be utilized in future NCHS products on the identification of opioid-involved hospital visits. NCHS will continue to monitor and promote the use of the NHCS data with enhanced identification of opioid-involved visits. As more years of the NHCS data becomes available, the algorithms developed through this project will continue to be updated in new data releases and creating new resources.

## **8. How to request linked NHCS data**

The 2016 NHCS data with enhanced identification of opioid-involved hospital visits with co-occurring disorders are made available through the NCHS Federal RDCs. To access these data, researchers must submit a written proposal that will be reviewed by NCHS staff. For more information on RDC access, please see this link: <https://www.cdc.gov/rdc/>.

## **9. References**

1. National Center for Health Statistics, Linked Data on Hospitalizations, Mortality, and Drugs: Data from the National Hospital Care Survey 2016, National Death Index 2016-2017, and the Drug-Involved Mortality 2016-2017, October 2020. Non-nationally representative data. Available from: <https://www.cdc.gov/nchs/data/nhcs/Task-3-Doc-508.pdf>
2. National Center for Health Statistics. National Hospital Care Survey: Research Data Center Documentation. 2019. Available from: [https://www.cdc.gov/nchs/data/datalinkage/NHCS16\\_NDI16\\_17\\_Methodology\\_Analytic\\_Consider.pdf](https://www.cdc.gov/nchs/data/datalinkage/NHCS16_NDI16_17_Methodology_Analytic_Consider.pdf).
3. National Hospital Care Survey. Hyattsville, MD. Available from: <https://www.cdc.gov/nchs/nhcs/index.htm>.