

**Enhancing Identification of Opioid-involved Health Outcomes using
Linked Hospital Care and Mortality Data**

Summary Report on Task 1

Patient Centered Outcomes Research Trust Fund

Project Period: 2018 to 2020

National Center for Health Statistics
Division of Health Care Statistics
Office of Analysis and Epidemiology
Division of Vital Statistics

May 2019

Introduction

Researchers need comprehensive data on opioid-related emergency department (ED) visits, inpatient (IP) hospitalizations, and deaths to identify and test strategies to reduce the morbidity and mortality from misuse and overdose of opioids. In 2018, the Patient Centered Outcomes Research Trust Fund (PCORTF) awarded funding to the National Center for Health Statistics (NCHS) to help address this need. These funds support NCHS's efforts to improve public health surveillance and expand researchers' access to data on opioid-involved health outcomes.

As part of this PCORTF-funded project, NCHS is developing enhanced methods that make use of available structured and unstructured data from three data sources -- the National Hospital Care Survey (NHCS), the National Death Index (NDI), and the National Vital Statistics System--Mortality--Drug Overdose (NVSS-M-DO) file -- to identify the specific opioids involved in drug-related ED visits, IP hospitalizations, and overdose deaths. The four major objectives of this project include:

- Developing and applying text mining strategies, such as natural language processing, to written and coded data to identify the specific opioids involved in hospital encounters and drug overdose deaths;
- Linking several data sources to create an enhanced, more comprehensive data set on care and outcomes;
- Building infrastructure to report additional clinical information back to hospitals via a hospital web portal; and
- Disseminating new data files, methods, and other outputs to the research community by providing access to analytic files through the NCHS Research Data Center (RDC) Network.

The first task (Task 1) of this project is to link data from the 2014 NHCS to data from the 2014 and 2015 NDI and data from the NVSS-M-DO file, as well as demonstrate the types of analyses that can be performed with the linked data. This report provides a brief description of the data sources, the processes used to link the data sets, and the variables available in the linked data sets. Example analyses demonstrating the value of the linked data are also presented.

Data Sources Used in Task 1

National Hospital Care Survey (NHCS) - <https://www.cdc.gov/nchs/nhcs/index.htm>

The NHCS is designed to collect data on patient care in hospital-based settings to describe patterns of health care delivery and utilization in the United States (1). Settings include emergency departments (ED), inpatient (IP), and outpatient departments (OPD). Data are collected from administrative claims and/or electronic health records (EHR) from participating hospitals.

Task 1 of this project focuses on 2014 NHCS data. This data set is based on administrative claims data coded using the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM). In this data set, ICD-9-CM codes can be used to identify subsets of visits for specific conditions. For example, a recent National Health Statistics Report described the identification of substance-involved ED visits using selected ICD-9-CM codes (2).

National Death Index (NDI) - <https://www.cdc.gov/nchs/ndi/index.htm>

The NDI is a centralized database of death record information on file in jurisdictional vital records or statistics offices and includes all deaths occurring within the United States (3). This data set can be used to identify persons who have died and their causes of death. NCHS maintains this data set. Causes of death occurring in 2014 and 2015 are identified using the *International Classification of Disease, Tenth Revision* (ICD-10) for underlying and multiple causes of death (4). Drug overdose deaths are identified using underlying cause-of-death ICD-10 codes X40-X44 (unintentional), X60-X64 (intentional self-harm/suicide), X85 (homicide), and Y10-Y14 (undetermined intent). Deaths involving specific drugs are identified using multiple cause-of-death ICD-10 codes T36-T50.8.

National Vital Statistics System–Mortality–Drug Overdose (NVSS-M-DO)

<https://www.cdc.gov/nchs/nvss/deaths.htm>

The NVSS-M-DO file includes information on the specific drugs involved in a death. The methods for preparing this file (referred to as the Drugs Mentioned with Involvement [DMI] program) were developed collaboratively by NCHS and the U.S. Food and Drug Administration (FDA) and involve extracting information from the literal text on death certificates (5). The literal text is the verbatim text provided by the medical certifier, usually a medical examiner or coroner, about the causes and circumstances of the death.

The DMI methodology involves searching the literal text from three fields of the death certificate (the cause-of-death sequence from Part I, significant conditions contributing to the death from Part II, and a description of how the injury occurred) using search terms for specific drugs. Search terms include generic drug names, brand names, common usage or street names, abbreviations, metabolites, misspellings, and other variations. Each search term has been mapped to a “principal variant,” the overarching label assigned to a drug, a drug class, or exposure not otherwise specified. For example, terms such as “COCAIEN”, “COCAINE CRACK”, “COCAINE HYDROCHLORIDE”, and “COCAINETOXICITY” are all mapped to the principal variant “COCAINE”. In general, the principal variant is the generic drug name. A few search terms—mostly for combination drug products—are mapped to two or more principal variants. Use of principal variants makes it possible to generate aggregate counts for all search terms that refer to the same drug or substance. The list of search terms and principal variants used in the DMI program can be found at https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Program_Code/oea/.

In the NVSS-M-DO data file, there is a variable for each principal variant. The value for the variable reflects the number of mentions of the principal variant on the death certificate for a

given decedent. If the principal variant was not mentioned, the value is null ("."). If the principal variant was mentioned, then the value reflects the number of mentions (a number greater than 0).

Linkage Processes

Linkage of 2014 NHCS to 2014/2015 NDI

The methods for linking 2014 NHCS to 2014/2015 NDI have been described previously (6). Briefly, the linkage involves both deterministic and probabilistic linkage followed by the selection of the final pair included in the file. The steps include:

1. Deterministic linkage, by performing joins on exact Social Security Number (SSN), validated by comparison of other identifying fields.
2. Probabilistic linkage, to identify likely matches for NHCS records for which a deterministic link could not be located. Interim steps involve:
 - a. Identifying possible matched pairs,
 - b. Scoring possible matched pairs, and
 - c. Probability modeling – assigning a probability that pairs are matches.
3. Selecting pairs to include in the final file.

In Step 1, for each set of records joined on a common value of SSN, the linkage compares the first name, last name or father's surname, month of birth, day of birth, year of birth, and state of residence identifiers to validate that the paired records actually represent the same individual.

In Step 2, the linkage conducts probabilistic record linkage of NHCS patient records with NDI records. This linkage identifies likely matched pairs and estimates the probability that they represent the same patient. If the probability score is above a pre-determined threshold, the linkage retains the pair as a link.

In Step 3, the linkage creates the final file by selecting the pair for each patient with the highest probability of being a match. In all cases, if a deterministic link has been returned, this link will be assigned a probability of one and will be selected over any link returned by probabilistic linkage.

Linkage to 2014/2015 NVSS-M-DO

Only those patient records that were eligible for and successfully linked to the 2014/2015 NDI could be linked to the 2014/2015 NVSS-M-DO files. Patient records that were already linked to the 2014/2015 NDI were linked to the 2014/2015 NVSS-M-DO files using joins on exact patient ID and a unique ID consisting of the year of death, jurisdiction of death, and death certificate number.

Select Types of Variables Available¹

National Hospital Care Survey (NHCS)	
	<i>Patient</i>
	Age
	Sex
	Setting (e.g., emergency department, inpatient, or outpatient department)
	Expected source of payment
	Diagnoses
	Procedures
	Beginning service date
	Ending service date
	Length of stay
	Discharge status
	Died in the hospital
	<i>Hospital</i>
	Geographic region
	Urban/rural classification
	Type of ownership (e.g., government, non-profit)
	Type of service (e.g., general acute care, psychiatric, children's)
	Number of beds
National Death Index (NDI)	
	Underlying cause of death
	Multiple causes of death
	Date of birth
	Date of death
	Age at death
National Vital Statistics System-Mortality-Drug Overdose (NVSS-M-DO)	
	Specific substances mentioned
	Drug classes mentioned
	Non-specific substances mentioned
¹ This list provides examples of the types of information available and does not include all variables in the data sets.	

Relevant data dictionaries for the data sets used in Task 1 are available at:

- NHCS: https://www.cdc.gov/rdc/data/b1/2013_NHCS.pdf
- NDI: https://www.cdc.gov/nchs/data/datalinkage/NHCS_2014_Codebook_Mortality_Variables.pdf
- NVSS-M-DO: <https://www.cdc.gov/rdc/b1/datatype/dt1229.html>

Case Definitions

Opioid-related Visits

Opioid-related visits include NHCS records with any mention of ICD-9-CM diagnosis codes 304.00–304.02, 304.70–304.72, 305.50–305.52, 760.72, 965.00–965.02, 965.09, or 970.1 or external cause of injury codes E850.00–E850.2. These codes include codes for heroin use, abuse, or toxicity. Codes for opioids causing adverse effects in therapeutic use (E935.0–E935.2, E940.1) were not included.

Drug Overdose Deaths

Drug overdose deaths include those with underlying cause-of-death ICD-10 codes X40-X44 (unintentional), X60-X64 (intentional self-harm/suicide), X85 (homicide), and Y10-Y14 (undetermined intent).

Drug Overdose Deaths Involving an Opioid

Drug overdose deaths involving an opioid include those with underlying cause-of-death ICD-10 codes X40-X44, X60-X64, X85, and Y10-Y14 and a multiple cause-of-death code of T40.0-T40.4 or T40.6 (includes opium, heroin, other natural and semisynthetic opioids, methadone, other synthetic opioids, and unspecified narcotics).

Records in the Linked Data Sets

Description	Number of Records	Number of Unique Patients ¹
Total 2014 NHCS visits (any type, any setting [ED visits or IP hospitalizations])	5,232,415	3,406,185
Linked to 2014/2015 NDI, any cause of death	343,500	166,013
Linked to 2014/2015 NDI, cause of death is drug overdose ² (any drug)	8,458	3,536
Linked to 2014/2015 NDI, cause of death is a drug overdose involving an opioid ³	5,243	2,193
2014 NHCS opioid-related ⁴ visits (ED visits and IP hospitalizations)	48,854	36,582 ⁶
Linked to 2014/2015 NDI, any cause of death	4,708	3,225
Linked to 2014/2015 NDI, cause of death is drug overdose ² (any drug)	1,289	847
Linked to 2014/2015 NDI, cause of death is a drug overdose involving an opioid ³	940	622
2014 NHCS opioid-related ⁴ ED visits only ⁵	18,857	15,495
Linked to 2014/2015 NDI, any cause of death	1,198	933
Linked to 2014/2015 NDI, cause of death is drug overdose ² (any drug)	565	410
Linked to 2014/2015 NDI, cause of death is a drug overdose involving an opioid ³	419	313
2014 NHCS opioid-related ⁴ IP hospitalizations	29,997	24,059
Linked to 2014/2015 NDI, any cause of death	3,510	2,564
Linked to 2014/2015 NDI, cause of death is drug overdose ² (any drug)	724	542
Linked to 2014/2015 NDI, cause of death is a drug overdose involving an opioid ³	521	385
<p>¹ Excludes newborns (births in the ED or during IP hospitalization).</p> <p>² Drug overdose deaths were identified using ICD-10 underlying cause-of-death codes X40-X44, X60-X64, X85, and Y10-Y14.</p> <p>³ Drug overdose deaths involving an opioid were identified using ICD-10 underlying cause-of-death codes X40-X44, X60-X64, X85, and Y10-Y14 with a multiple cause-of-death code of T40.0-T40.4 or T40.6.</p> <p>⁴ Opioid-related visits were identified based on any mention of ICD-9-CM diagnosis codes: 304.00-304.02, 304.70-304.72, 305.50-305.52, 760.72, 965.00-965.02, 965.09, or 970.1 or external cause of injury codes E850.0-E850.2. These codes included codes for heroin use, abuse, or toxicity.</p> <p>⁵ "ED visits only" excludes 20,236 encounters that resulted in the patient being admitted from the ED to IP hospitalization.</p> <p>⁶ Patients with separate ED visits and an IP hospitalizations are only counted once in this total. They are included in the ED visit only and IP hospitalizations totals.</p> <p>NOTE: Data are not nationally representative.</p>		

Example Research Questions

There are several types of research questions that can be examined with the linked data, including “look forward” studies (e.g., outcomes for patients with an opioid-related ED visit or IP hospitalization) and “look back” studies (e.g., recent hospital care for decedents who died from a drug overdose). The following example study questions demonstrate some of the types of analyses that can be done with the linked data:

- What are patient and hospital characteristics for patients who had an opioid-related ED visit or IP hospitalization?
- What are common patterns of hospital use in the months prior to a death due to drug overdose involving an opioid?
- For patients that had an opioid-related ED visit or IP hospitalization, how do services received between those who died from an overdose and those who did not die from an overdose compare?
- How do patients with a history of repeated opioid-related ED visits or IP hospitalizations that die from an opioid overdose compare to those who did not die from an opioid overdose?

A forthcoming NCHS National Health Statistics Report will be published to provide further detail and address these research questions.

Summary

This report provides a brief overview and description of the data sources, processes used to link data, information included in the data files, and examples of the types of research questions that can be answered using the enhanced data files developed under Task 1 of the Patient Centered Outcomes Research Trust Fund project, “*Enhancing identification of opioid-involved health outcomes using linked hospital care and mortality data.*” This resource, linked data from the 2014 National Hospital Care Survey (NHCS), the 2014-2015 National Death Index (NDI), and the 2014-2015 National Vital Statistics System–Mortality–Drug Overdose (NVSS-M-DO) file, provides comprehensive information on opioid-related ED visits, IP hospitalizations, and deaths. These data are available to researchers in the NCHS RDC; more information on how to obtain access to the data can be found here: <https://www.cdc.gov/rdc/>. Such data can enable researchers to investigate questions concerning opioid-related hospital encounters, mortality, and the specific drugs involved in death and informing public health interventions to reduce morbidity and mortality due to opioids.

References

1. National Hospital Care Survey. <https://www.cdc.gov/nchs/nhcs/index.htm>

2. Brown AM, DeFrances C, Crane E, Cai R, Naeger S. Identification of substance-involved emergency department visits using data from the National Hospital Care Survey. National Health Statistics Reports; no 114. Hyattsville, MD: National Center for Health Statistics. 2018. Available from: <https://www.cdc.gov/nchs/data/nhsr/nhsr114.pdf>.
3. National Death Index. <https://www.cdc.gov/nchs/ndi/index.htm>
4. World Health Organization. International statistical classification of diseases and related health problems, tenth revision (ICD–10). 5th ed. Geneva, Switzerland. 2016. Available from: <https://icd.who.int/browse10/2016/en>.
5. Trinidad JP, Warner M, Bastian BA, et al. Using literal text from the death certificate to enhance mortality statistics: Characterizing drug involvement in deaths. National vital statistics reports; vol 65 no 9. Hyattsville, MD: National Center for Health Statistics. 2016. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_09.pdf
6. National Center for Health Statistics. Office of Analysis and Epidemiology. The Linkage of the 2014 National Hospital Care Survey to the 2014/2015 National Death Index: Methodology Overview and Analytic Considerations, April 2018. Hyattsville, Maryland. Available from: https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15_Methodology_Analytic_Consider.pdf