National Center for Health Statistics Research Data Center (NCHS RDC)

Linked Data on Hospitalizations, Mortality, and Drugs:
Data from the National Hospital Care Survey, National Death Index, and the Drug-Involved Mortality

October 2020

OVERVIEW

This document describes the research data set created by linking three data sources: (a) the 2014 National Hospital Care Survey (NHCS), (b) the 2014/2015 National Death Index (NDI), and (c) the 2014/2015 Drug-Involved Mortality restricted data (DIM). Details about the NHCS and NDI data sets and their linkage have been described elsewhere (see below). This document supplements the previous reports by focusing on the variables in the DIM file.

This document provides:

- a brief description of the three source data sets;
- information about the records and variables in the DIM data; and
- resources for additional information.

SOURCE DATA SETS

NHCS collects inpatient (IP), emergency department (ED), and outpatient department (OPD) administrative claims or electronic health record (EHR) data from participating hospitals. This data set includes information captured during a hospital visit such as diagnoses, services provided at the hospital, and discharge status. Due to the low response rate of sampled hospitals, the 2014 NHCS is unweighted and the estimates are not nationally representative. For general information about the NHCS, see: https://www.cdc.gov/nchs/nhcs/index.htm. For information on the 2014 NHCS data available in the RDC, see: https://www.cdc.gov/rdc/data/b1/2013_NHCS.pdf.

NDI includes data on all deaths occurring within the United States and identifies persons who have died, as well as their causes of death. For general information on the NDI, see: https://www.cdc.gov/nchs/ndi/index.htm. The 2014 NHCS data have been linked to the 2014 and 2015 NDI data to track mortality after hospital visits. For information on the methodology used to link NHCS and NDI data, see: https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15_Methodology_Analytic_Consider.pdf. For a list of the mortality variables in the linked NHCS/NDI data set available in the

RDC, see:
https://www.cdc.gov/nchs/data/datalinkage/NHCS_2014_Codebook_Mortality_Variables.pdf.

DIM includes information on specific drugs, drug classes, and non-specific references to drugs mentioned in the literal text on the death certificate. Identification of the specific drugs involved in a death is based on the Drug Mentioned with Involvement (DMI) methodology, which was developed collaboratively by NCHS and the U.S. Food and Drug Administration (FDA). A brief description of the DMI methodology is provided below; however researchers are encouraged to read the more detailed description at:
https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_09.pdf. Only those NHCS patient records that were eligible for and successfully linked to the 2014/2015 NDI can be linked to the 2014/2015 DIM files. Linked NHCS/NDI records are linked to the 2014/2015 DIM files using an exact match on the NHCS patient ID and a unique ID generated from the year of death, jurisdiction of death, and death certificate number.

Example of Types of Data Available

The table below contains an example of the type of data available from the three data sources.

National Hospital Care Survey (NHCS)	
Patient	
	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
	Intensive care unit stays (IP only)
	Discharge status
	Died in the hospital
Hospital	
	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
Drug-Involved Mortality (DIM)	
	Specific substances mentioned
	Drug classes mentioned
	Non-specific substances mentioned
[1]The list provides examples of the types of information available and does not include all variables in the data sets.	

The number of records available in the linked dataset for ED visits and IP hospitalizations are in the table below.

Description	Number of Records[1]
Total 2014 NHCS visits (any type, any setting [ED visit or IP hospitalization])	5,232,415
Linked to 2014/2015 NDI, any cause of death	343,500
Linked to 2014/2015 NDI, cause of death is drug overdose[2] (any drug)	8,458
Linked to 2014/2015 NDI, cause of death is a drug overdose involving an opioid[3]	5,243

[1] Excludes newborns (births in the ED or during IP hospitalization).

[2] Drug overdose deaths were identified using ICD-10 underlying cause of death codes X40-X44, X60-X64, X85, and Y10-Y14.

[3] Drug overdose deaths involving any 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.

INFORMATION ABOUT THE RECORDS AND VARIABLES IN THE DIM DATA FILE

The decedent-level file of the DIM data contains a single record for each decedent who had at least one substance mentioned in the death certificate literal text. The identification of the drugs/substances involved in the death is based on the DMI methodology (https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_09.pdf). Briefly, the DMI methodology involves searching the literal text from three fields on the standard U.S. death certificate: the causes of death from Part I, significant conditions contributing to death from Part II, and a description of how the injury occurred. The literal text is searched for mentions of drugs and other substances, as well as terms that provide context about the involvement of the drug in the death (i.e., whether the drug contributed to the death). The substances mentioned in the death certificate literal text are assumed to be involved in the death unless contextual information in the literal text suggests otherwise.

Search Terms

Drugs and substances are identified using search terms. Search terms include generic drug names, brand names, common usage or street names, abbreviations, metabolites, misspellings, and other variations. The creation of the search term list involved multiple approaches including: (a) including single-word generic names listed in the Substance Abuse and Mental Health Services Administration (SAMHSA) Drug Abuse Warning Network (DAWN) Drug Reference Vocabulary (DRV), based on the Multum Lexicon database from Cerner Multum, Inc.; (b) including the terms added by SAMHSA for substances that are misused and abused (e.g.,

illic drugs and inhalants) that were not included in the Multum Lexicon database; and (c) adding terms for drug classes, drug exposures not otherwise specified, terms containing more than one word, brand names (using the Drugs@FDA website), and obvious frequently occurring misspellings (from manual review of the literal text and from the FDA Adverse Event Reporting System data). The list of DMI search terms is available in the Excel workbook Search_Terms_and_Principal_Variants.xlsx at https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Program_Code/oae/.

Principal Variants

Each search term is 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 assigned the principal variant "COCAINE". In general, the principal variant is the generic drug name. Some search terms—mostly for combination drug products—are assigned 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. Principal variants are also categorized according to whether they refer to specific drugs or substances (e.g., methadone), classes of drugs or substances (e.g., opioids), or nonspecific references to exposures to drugs (e.g., words such as "DRUG", "MULTIDRUG", or "POLYPHARMACY").

Whenever possible, principal variants for specific substances are the unique preferred substance names listed in the joint FDA/United States Pharmacopeia (USP) Substance Registration System (SRS). Each preferred substance name is linked to a unique ingredient identifier (UNII) in the SRS, if available. The preferred name may differ from the term(s) commonly used in everyday practice or even the terms more often mentioned in literal text. For example, although the search term HEROIN is mentioned more often in literal text than the search term DIAMORPHINE, the search term HEROIN is mapped to the principal variant DIAMORPHINE. Consult the SRS for preferred names of potential substances of interest. More information about the SRS is available at https://www.fda.gov/ForIndustry/DataStandards/SubstanceRegistrationSystem-UniqueIngredientIdentifierUNII, and at https://fdasis.nlm.nih.gov/srs/. In some instances, principal variants for classes of drugs and substances are classes listed in the American Hospital Formulary System (AHFS).

Some principal variants are alternatives to or shortened versions of the SRS preferred substance name or AHFS classes. Shortened or alternative principal variants are used because the substance-related SAS variable names in the decedent-level file cannot be longer than 32 characters. For example, the SRS preferred name 2-ETHYLIDENE-1,5-DIMETHYL-3,3-DIPHENYLPYRROLIDINE is 49 characters and is abbreviated as the principal variant EDDP.

Furthermore, some principal variants are simplified versions of the SRS preferred name or AHFS class. For example, the SRS preferred name TOBACCO LEAF is simplified as the principal variant TOBACCO. The column Preferred name or class in the Excel workbook Search_Terms_and_Principal_Variants.xlsx lists the SRS preferred substance name in instances when the principal variants for specific substances are not the SRS preferred name. Researchers who use the SRS or AHFS to identify potential substances of interest should also consult the Excel workbook to ensure comprehensive capture of all mentions of interest.

Some search terms could not be mapped to a unique principal variant. When a search term may refer to two or three different substances, the search term is mapped to the principal variant AMBIGUOUS. The Excel workbook Search_Terms_and_Principal_Variants.xlsx lists the potential principal variants in the ambiguous variant variables, as well as the UNII for the potential principal variants (if available), in the columns ambiguous_variant_UNII_1, ambiguous_variant_UNII_2, and ambiguous_variant_UNII_3.

In some instances, the search terms are mapped to the principal variant UNKNOWN, indicating that the search term may refer to four or more difference substances.

The principal variants list is quite detailed (e.g., there may be unique principal variants for a "parent" drug as well as its metabolites), therefore researchers might consider grouping multiple principal variants into a larger "referent category" in order to identify all the deaths involving a particular drug. For example, the principal variant HYDROCODONE and the principal variant NORHYDROCODONE (a metabolite of hydrocodone) might be grouped together to create a referent drug category of HYDROCODONE. A recent report from NCHS provides an example of how principal variants were combined into a single referent group (https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_09-508.pdf).

Because principal variants can contain symbols and spaces, SAS programs used to analyze the decedent-level file should include the following global SAS option: OPTIONS VALIDVARNAME = ANY. This option allows for the use of non-conforming SAS variable names. In addition, refer to each principal variant using the name literal convention (i.e., refer to substance-specific variables using "substance"n (quotes required); for example, "- gamma.hydroxybutyric acid"n).

The list of DMI search terms and principal variants is available in the Excel workbook Search_Terms_and_Principal_Variants.xlsx at https://ftp.cdc.gov/pub/Health_statistics/NCHS/Program_Code/oae/.

Variables

The DIM decedent level file includes a variable for each principal variant. The value of the variable reflects the number of mentions of involvement of the drug/substance on the death certificate for that decedent.

Additional variables identify: (a) the number of unique principal variants mentioned, (b) the number of specific substances identified, (c) the number of unique drug classes identified, and (d) the number of non-specific drug references identified for a given decedent.

Variable Name	Variable description	Data Values
PATIENT_ID	**Char.** Patient identifier assigned by NCHS. Researchers linking to the 2014 NHCS should use PATIENT_ID as the common key.	>0
count_PV_involved	**Numeric.** Quantifies the number of unique principal variants involved in the death. Principal variants can include specific substances (e.g., OXYCODONE or FENTANYL), classes of drugs or substances (e.g., OPIOID), non-specific substances (e.g., DRUG, CHEMICAL, or POLYPHARMACY), ambiguous substances (i.e., possibly referring to two or three different substances), or unknown (i.e., possibly referring to 4 or more different substances. If the *count_PV_involved* variable is equal to 0, then all drugs/substances mentioned in the death certificate literal text were categorized as not being involved in the death.	≥0
count_specific_subs_involved	**Numeric.** Quantifies the number of unique specific drugs/substances (e.g., HEROIN) involved in the death.	≥0
count_classes_involved	**Numeric.** Quantifies the number of unique drug classes (e.g., OPIOID) involved in the death.	≥0
count_non_specific_subs_involved	**Numeric.** Quantifies the number of non-specific drugs/substances (e.g., DRUG, CHEMICAL, POLYPHARMACY) involved in the death.	≥0

'principal variant'	**Numeric**. The data set includes a variable for each principal variant. The value of the variable reflects the number of mentions of involvement of that drug/substance for that decedent. If the drug/substance was not categorized as having been involved in the death, the value for the variable will be null ("."). Because some of principal variant variables do not conform to ordinary SAS name standards, it is recommended the following SAS option be used: OPTION VALIDVARNAME = ANY.	>0 or .

RESOURCES FOR ADDITIONAL INFORMATION

National Hospital Care Survey. Available from: https://www.cdc.gov/nchs/nhcs/index.htm.

National Center for Health Statistics Research Data Center National Hospital Care Survey 2013-2014 Data Dictionary. Available from: https://www.cdc.gov/rdc/data/b1/2013_NHCS.pdf.

National Death Index. Available from: https://www.cdc.gov/nchs/ndi/index.htm.

The Linkage of the 2014 National Hospital Care Survey to the 2014/2015 National Death Index: Methodology Overview and Analytic Considerations. Available from: https://www.cdc.gov/nchs/data/datalinkage/NHCS14_NDI14_15_Methodology_Analytic_Consider.pdf.

Codebook for NHCS Linked to the NDI, Mortality Variables. Available from: https://www.cdc.gov/nchs/data/datalinkage/NHCS_2014_Codebook_Mortality_Variables.pdf.

Using Literal Text from the Death Certificate to Enhance Mortality Statistics: Characterizing Drug Involvement in Deaths. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr65/nvsr65_09.pdf.

Drugs Most Frequently Involved in Drug Overdose Deaths: United States, 2011–2016. Available from: https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_09-508.pdf.