

National Hospital Care Survey
Research Data Center
Documentation

Last Updated: 09/27/2023

Table of Contents

Abstract	3
1. The National Hospital Care Survey Background.....	4
2. Advantages of the National Hospital Care Survey	4
3. Hospital Frame and Sample Design.....	4
3.1 Participating Hospitals	5
4. Data Sources.....	5
5. Patient Identification	10
6. Data Elements	10
7. Supplemental Datasets.....	10
7.1 Enhanced Opioid Identification Dataset	11
7.2 Enhanced Co-Occurring Disorders Dataset.....	11
7.3 National Death Index	11
7.4 U.S. Department of Housing and Urban Development	12
7.5 Centers for Medicare & Medicaid Services	12
8. Analytic Considerations.....	12
8.1 Diagnosis Codes.....	12
8.2 Procedure Codes	13
8.3 Revenue Codes	13
8.4 Discharge Status.....	13
8.5 Emergency Department Visits Admitted as Inpatients	13
9. Additional Information and Contact Information	14

Abstract

This document provides an overview of the National Hospital Care Survey (NHCS), NHCS datafiles available in the Federal and National Center for Health Statistics Research Data Centers (RDC), and analytic considerations for NHCS data users. NHCS is a national probability sample survey from the National Center for Health Statistics that collects hospital patient and encounter data from the inpatient and emergency department settings. Data are available in the RDC for calendar years 2013-2016, and 2019-2021. Beginning in 2020, preliminary NHCS data from selected hospitals are available in the RDC. The descriptions and analytic considerations in the RDC documentation apply to all NHCS survey years unless otherwise noted. For information on the available data variables for each year, please see: <https://www.cdc.gov/rdc/b1datatype/dt1224h.htm>

Researchers interested in using NHCS data can develop a proposal, including a list of variables needed to create their research dataset, which will be delivered to researchers in the RDC upon approval. For more information on how to develop a proposal, please see: <https://www.cdc.gov/rdc/b3prosal/pp300.htm>.

1. The National Hospital Care Survey Background

The National Hospital Care Survey (NHCS) provides data on health care utilization patterns in hospital-based settings. Data collection for NHCS began in 2011, integrating two long-standing National Center for Health Statistics (NCHS) surveys:

- National Hospital Discharge Survey, the longest continuously fielded sample of inpatient care from 1965-2010; and
- National Hospital Ambulatory Medical Care Survey, which has surveyed hospital emergency departments (ED) (1992-2022), outpatient departments (OPD) (1992-2017), hospital Ambulatory Surgery Locations (2009-2017), and freestanding Ambulatory Surgery Centers (2010-2017).

The survey collects data on hospital encounters including diagnosis and procedure codes, length of patient stay, and patient demographics. NHCS also collects patient personal identifiable information (PII), allowing researchers to identify patients with multiple hospital encounters and link NHCS to external datasets. Encounter data is collected from the inpatient and ED settings, OPD data was collected from 2013-2016.

NHCS sample includes non-federal non-institutional hospitals with six or more staffed inpatient beds in the 50 states and the District of Columbia. Every three years the sample is updated to include newly opened hospitals. The 2011-2016 NHCS sample included 581 eligible hospitals. In 2019, the sample increased to 598 hospitals and 608 hospitals in 2020. However, not all sampled hospitals provide data. As a result, the response rates for the 2013-2016 and 2019 NHCS were not high enough to produce weighted national estimates. However, the response rates for 2020 and 2021 were high enough to produce national estimates for the 2020 and 2021 NHCS. Nationally representative data from the 2020 and 2021 NHCS are expected to be available very soon.

Numerous scientific publications using unweighted NHCS data for healthcare research are available, and can be found here: https://www.cdc.gov/nchs/nhcs/data_uses.htm

2. Advantages of the National Hospital Care Survey

NHCS is a unique data source that contains longitudinal data on hospital utilization and patient care. Another unique aspect of NHCS is the data are publicly available to the researchers in the Federal and NCHS RDCs. With an approved proposal, researchers can access longitudinal data on millions of inpatient and ED encounters from participating hospitals.

Due to the collection of PII, NHCS allows linkage of patients within survey years. Linkage capabilities also include the following:

- Counting the number of encounters per patient;
- Identifying the number of encounters in each hospital setting by patient;
- Identifying ED patients who were transferred to the inpatient department;
- Linking the NHCS to external datafiles that provide supplementary patient information not collected by the NHCS.

3. Hospital Frame and Sample Design

The initial frame for NHCS was constructed in 2011 using the 2010 IQVIA hospital database (previously known as SDI, Verispan, SMG, IMS Health, and IMS Government Solutions). The initial frame file consists of 6,622 non-federal and non-institutional hospitals with at least six staffed inpatient beds. From that initial file, a sample of 1,000 eligible hospitals were selected stratified by hospital bed size, type of hospital, and urban/rural designation. The initial sample of 1,000 hospitals was split into two groups of 500 – a base sample and a reserve sample. The base sample hospitals were selected for data collection and the reserve sample of hospitals were held to replenish the base sample if more hospitals were needed.

In 2013, 81 general acute hospitals with at least 500 staffed inpatient beds from the reserve sample were added to the base sample for data collection. From 2013 through 2016, 581 hospitals were eligible to participate in the survey. The 2017 sample frame was refreshed with a sample of newly constructed hospitals from a new IQVIA source file. These new hospitals (referred to as “birth hospitals”) were selected from the 2015 IQVIA file. Of the 598 hospitals in the base sample, 566 hospitals were from the 2013 augmented base sample file, and 32 hospitals were from birth hospitals from the 2015 IQVIA file. The 2017 refreshed base sample had 598 hospitals. In 2020, the base sample was increased from 598 to 608 hospitals due to the addition of newly sampled birth hospitals.

3.1 Participating Hospitals

Table 1 presents the number of hospitals, encounters, and response rates of NHCS by setting and survey year. Note, the NHCS was not fielded in 2017 due to budget limitations and data collection resumed in 2018.

Table 1. Number of Hospitals and Encounters by Department by Year

Year	Inpatient		ED		OPD		Total		Response Rate
	Hosp	Encounters	Hosp	Encounters	Hosp	Encounters	Hosp	Encounters	
2013	97	1,474,478	82	3,784,397	87	15,144,448	97	20,403,323	16.7
2014	94	1,653,622	83	4,530,360	86	19,005,777	95	25,189,759	16.4
2015	114	2,204,258	97	5,900,738	101	26,455,149	118	34,560,145	20.3
2016	145	2,591,722	124	7,032,304	128	35,692,420	150	44,572,315	25.8
2019	112	2,228,190	106	5,549,330	N/A	N/A	112	7,244,537	18.7
2020	150	2,832,062	199	7,978,530	N/A	N/A	205	10,072,917	33.7
2021	197	3,672,491	236	9,980,942	N/A	N/A	247	12,745,505	40.6

Note: Hosp is hospital, ED is emergency department, and OPD is outpatient department. Emergency Department patients transferred to the inpatient department are counted in both the inpatient and emergency department settings. NHCS collected outpatient data from 2013-2016, outpatient data was not collected in 2019-2021.

4. Data Sources

In 2011, NHCS collected Uniform Billing (UB)-04 administrative claims data from participating hospitals. UB-04 administrative claims data are the accepted electronic standard for hospital billing mandated by the Centers for Medicare & Medicaid Services (CMS) for payment of charges for Medicare and Medicaid recipients. UB-04 data collects information on patient sex and age, hospital setting, length of stay, discharge status, and up to 25 diagnosis and procedure codes.

Beginning in 2015, sampled hospitals could provide data from two additional data sources: Vizient and electronic health record (EHR) data. Vizient is a member-driven health care services company that collects encounter data from hospitals prior to submitting to data to NHCS. Vizient data are similar to UB-04 claims data in structure but include laboratory and medications data. Due to privacy concerns, Vizient data does not provide patient person identifiable information (PII), and the exact date of the hospital visit are defaulted to the first of the month.

Unlike UB-04 claims and Vizient, EHR data contain an unlimited number of diagnosis and procedure codes, laboratory, and medication data. Another benefit of EHR data is the submission of unstructured clinical notes providing researchers a unique opportunity to have additional context to patients' hospital visits. Due to privacy concerns the notes, medication, and laboratory data are not available in the RDC. However, the information collected in the clinical notes, medications, and laboratory data is used to identify and extract opioid-involved encounters. The information on opioid-involved encounters can be found in the supplemental datasets (see section 6.2 and 6.3).

Beginning in 2020, NCHS worked with the American College of Emergency Physicians (ACEP) to receive data from their hospital database for sampled non-participating hospitals. The ACEP data was originally collected to provide ED quality measures for participating hospitals. The data from ACEP are only from EDs (it does not include any inpatient data) and they do not include any hospital or patient identifiers. A summary of NHCS data sources are presented below in Table 2.

Table 2. Summary of NHCS Data Sources

Data Source	Year	Description	Considerations
UB-04 claims	2013-2016, 2019-2021	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/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/837I-FormCMS-1450-ICN006926.pdf .	<ul style="list-style-type: none"> • Administrative data used for billing rather than clinical purposes (may exclude clinically important data that was not needed for billing) • Allows tracking of ED encounters who were later admitted as inpatients • Maximum of 25 diagnosis and procedure codes provided • Revenue code indicator variables are available
Vizient	2015-2016, 2019-2021	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/ .	<ul style="list-style-type: none"> • Administrative data used for billing rather than clinical purposes. May exclude clinically important data that was not needed for billing • No PII (e.g., patient name) collected and patients cannot be linked to external data sources • The exact date of the start and end of the hospital encounter is not collected. Only the month and year of the end of the encounter is given • No data on length of stay • Maximum of 25 diagnosis and procedure codes provided • Does not allow for tracking of ED encounters who were later admitted as inpatients • Revenue code indicator variables are available

Data Source	Year	Description	Considerations
EHR (Custom Extracts and CCDs)	2015-2016	<p>An electronic version of a patient’s medical history maintained by the hospital over time see https://www.cms.gov/Medicare/E-Health/EHealthRecords/index.html.</p> <p>In 2016, EHR data was collected as Consolidated-clinical document architecture (C-CDA) and EHR Custom Extracts.</p> <p>C-CDA is a set of HL7 Clinical Document Architecture submitted using implementation templates such as Continuity of Care Documents (CCD). CCDs are an electronic document exchange standard for sharing patient, transfer, and discharge summary information. Summaries include information about current and past health status that can be shared by computer applications including web browsers and electronic medical and health record software systems. In 2016, NCHS received discharge summaries, history and physical summaries, and transfer summaries.</p> <p>EHR Custom Extract data are extracted from a hospital’s EHR according to specifications provided by NCHS. The specifications for these extracts are based on the National Health Care Surveys Implementation Guide (IG).</p>	<ul style="list-style-type: none"> • Data has clinical rather than billing focus • No limit on the number of diagnoses and procedures collected per encounter • Does not allow for tracking of ED encounters who were later admitted as inpatients • Primary diagnosis is not available • PII collected, which allows for linkage of patients to external data sources
EHR (IG)	2019-2021	<p>Beginning in 2019, the NHCS only accepted EHR in the format of National Health Care Surveys CDA Release 1.2 Implementation Guide (IG). For more information, please visit the HL7 CDA R2 IG: National Health Care Surveys (NHCS), R1 STU Release 3 - US Realm website: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=385</p> <p>The IG was developed to submit data in the exact format needed for the NHCS and other NCHS National Health Care Surveys. All EHR data are extracted from the hospitals’ EHR according to specifications provided by NCHS.</p>	<ul style="list-style-type: none"> • Data collected has a clinical not billing focus • No limit on the number of diagnoses and procedures collected • Does not allow for tracking of ED encounters who were later admitted as inpatients • Primary diagnosis is not available. • PII included, which allows for linkage to external data sources

Data Source	Year	Description	Considerations
American College of Emergency Physicians (ACEP)	2020-2021	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 includes diagnoses, medications, laboratory results, and clinical notes. For more information visit: https://www.acep.org/	<ul style="list-style-type: none">• Only contains ED data, no inpatient data was collected• No PII (e.g., patient name) collected and patients cannot be linked to external data sources

5. Patient Identification

NHCS collects PII allowing researchers to follow patients who have multiple encounters and link the survey to external datasets. Unique patients are identified by creating a Patient ID variable. To assign the Patient ID, the records are deduplicated and encounters that belong to the same person are identified. Probability-based linkage methods were used to de-duplicate the records by a person's name, date of birth, sex, medical record number, address, and Social Security Number (if available). A probability-based linkage method is necessary for patient identification because data for the same person may not be identical on all hospital records for each encounter.

6. Data Elements

After data collection and processing, the NHCS data sources are harmonized into datasets. There are several NHCS datasets in the Federal and NCHS RDCs. For each survey year, the survey collects core data of hospital, patient, and encounter level information that are available from each data source. From each data source, NHCS collects information on the hospital encounter setting, diagnoses, services, discharge status, point of origin, and hospital encounter date. Refer to the data dictionaries linked below for detailed information on the type of data elements for each year available in the RDC including variable names, variable values, and variable availability by hospital setting.

- 2013-2014: https://www.cdc.gov/rdc/data/b1/2013_NHCS.pdf
- 2015: https://www.cdc.gov/rdc/data/b1/2015_NHCS_RDC_Data_Dictionary.pdf
- 2016: https://www.cdc.gov/rdc/data/b1/2016_NHCS_DATA-DICTIONARY.pdf
- 2019-2021: <https://www.cdc.gov/rdc/data/b1/NHCS-RDC-Data-Dictionary.pdf>

7. Supplemental Datasets

The collection of PII gives researchers the ability to link NHCS data to supplemental datasets to provide additional patient information that is not collected in the survey. This includes information on patients with opioid-involved and co-occurring disorder hospital encounters that were identified by enhanced algorithms utilizing natural language processing to analyze data items not available in the RDC, such as clinical notes and medication data.

Additionally, the NCHS Data Linkage Program uses NHCS PII data to link the data to external data sources. The linkage of NHCS to other data sources increases the analytic power of NHCS by identifying patient outcomes before and after hospital visits. The datasets linked to NHCS currently include data from the National Death Index (NDI), U.S. Housing and Urban Development (HUD), and Centers for Medicare & Medicaid Services (CMS). They are all available in both the NCHS and Federal RDCs. Information on each survey year's linkage methodology, data dictionary, and analytical considerations can be found here: <https://www.cdc.gov/nchs/data-linkage/nhcs-linkage.htm>. A brief description of the supplemental datasets in the NCHS and Federal RDCs are listed below.

7.1 Enhanced Opioid Identification Dataset

In fiscal year 2018, NCHS received funding to improve the identification of opioid-involved encounters in NHCS data using funding from the OS-Patient Centered-Outcome Research Trust Fund (OS-PCORTF, <https://aspe.hhs.gov/sites/default/files/documents/8af89f6a131cdc3572c2d3a78d9abba8/os-pcortf-portfolio-report.pdf>). The Enhanced Opioid Identification Algorithm utilizes Natural Language Processing (NLP) and machine learning techniques to analyze diagnostic codes and unstructured clinical data. The unstructured clinical data include the clinical notes, medication, and laboratory results collected from EHR and Vizient data. Due to the risk of PII disclosure, medication, laboratory, and clinical notes data are not currently available in the RDC.

The Enhanced Opioid Identification Dataset identifies opioid-involved and opioid overdose ED and inpatient hospital encounters. The dataset identifies 13 commonly used opioids, generic mentions of opioids, and opioid antagonists' naloxone and naltrexone mentioned during the patient encounter. The 2016 NHCS data was the first year to have a supplemental dataset based on results from the Enhanced Opioid Identification Algorithm. Subsequent years of the survey will include similar supplemental datasets with information on opioid-involved hospital encounters. The methodology used to develop the case definitions and algorithm is described elsewhere (https://www.cdc.gov/nchs/data/series/sr_02/sr2-188.pdf).

- 2016: <https://www.cdc.gov/nchs/data/nhcs/Task-3-Doc-508.pdf>
- 2019-2020: <https://www.cdc.gov/rdc/data/b1/2020-Opioid-RDC-508.pdf>
- 2021: Coming soon!

7.2 Enhanced Co-Occurring Disorders Dataset

NCHS received additional OS-PCORTF funding in fiscal year 2019 to develop a supplemental file to the Enhanced Opioid Identification Dataset that identifies co-occurring disorders, substance use disorders, and mental health issues among patients with an opioid-involved hospital encounter. The Co-Occurring Disorder Algorithm created a dataset by utilizing coded medical data and NLP techniques to search the unstructured EHR clinical notes. The Enhanced Co-Occurring Disorders Dataset contains information on mental health issues such as anxiety, depression, and self-harm and selected substance use disorders. The methodology used to define the case definitions and algorithm development is described elsewhere (https://www.cdc.gov/nchs/data/series/sr_02/sr02-193.pdf).

- 2016: <https://www.cdc.gov/nchs/data/nhcs/FY19-RDC-2021-06-01-508.pdf> n
- 2019-2021: Coming soon!

7.3 National Death Index

The PII collected from the NHCS was sent to the NCHS's Division of Vital Statistics for linkage to the National Death Index (NDI). The NDI provides information on a decedents date of death, the underlying cause of death, and multiple causes of death. Below is a list of the NHCS survey years linked to the NDI:

- 2014 NHCS linked to the 2014-2015 NDI
- 2016 NHCS linked to the 2016-2017 NDI

Detailed information about the NDI, linkage methodology, and data dictionary can be found on the NDI homepage: <https://www.cdc.gov/nchs/ndi/index.htm>.

7.4 U.S. Department of Housing and Urban Development

Data available in the NHCS-U.S. Department of Housing and Urban Development (HUD) linked file include information on the type of housing assistance received by the patient, when the housing assistance was received, the structure of the housing, and household characteristics. Below is a list of the NHCS survey years linked to the HUD.:

- 2014 NHCS linked to the 2013-2015 HUD Housing Assistance Program Files
- 2016 NHCS linked to the 2015-2017 HUD Housing Assistance Program Files

Information on the linkage methodology, data dictionary, and analytical considerations can be found on the NHCS-HUD data linkage webpage: <https://www.cdc.gov/nchs/data-linkage/hud-restricted.htm>.

7.5 Centers for Medicare & Medicaid Services

Linking NHCS data to CMS provides information on hospital utilization and charges among elderly adults and patients who receive Medicare disability. In 2014, NHCS was linked to the Centers for Medicaid and Medicare (CMS) Medicare Master Beneficiary Summary File (MBSF). The 2016 NHCS is linked to the Medicare MBSF, Claims/Encounters, and Assessment Data. Additionally, the 2016 NHCS data is linked to CMS Transformed Medicaid Statistical Information System (T-MSIS) data. Below is a list of the NHCS survey years linked to CMS data:

- 2014 NHCS linked to the 2014-2015 CMS Medicare Master Beneficiary Summary File
- 2016 NHCS linked to the 2016-2017 CMS Medicare Enrollment, Claims, and Assessment Data
- 2016 NHCS linked to 2015-2017 CMS T-MSIS Claims Data

Information on each survey year's linkage methodology, data dictionary, and analytical considerations can be found on the NHCS-CMS data linkage webpage: <https://www.cdc.gov/nchs/data-linkage/CMS-Medicare-Restricted.htm>.

8. Analytic Considerations

The 2013-2016 and 2019 NHCS data available in the RDC are unweighted and are not nationally representative. The results and analyses produced from NHCS during those survey years data are based on participating hospitals. The 2020 and 2021 NHCS are nationally representative and the weighted file will be available in the RDC soon.

8.1 Diagnosis Codes

Due to the implementation of International Classification of Diseases, Tenth Revision (ICD-10-CM), the 2015 NHCS datafiles have a mixture of International Classification of Diseases, Ninth Revision (ICD-9-CM) and ICD-10-CM codes for diagnoses and procedures. The first three quarters of the 2015 calendar year were coded as ICD-9-CM and the last quarter was coded as ICD-10-CM. Researchers can identify which coding system was used in the datafile by using the indicator variable that identifies diagnosis code system name.

Most of the diagnoses collected in the 2016, 2019-2021 NHCS are coded as ICD-10-CM. Hospitals also submitted diagnosis codes as ICD-9-CM, Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT), and site/EHR vendor custom codes. When possible, the ICD-9-CM and SNOMED-CT diagnosis codes are translated to ICD-10-CM. The original and translated diagnosis codes are available to researchers in the RDC.

Primary diagnosis cannot be identified in EHR data sources for NHCS survey years 2015-2016 or 2019-2021. Hospitals participating in the NHCS that submitted EHR data in 2016 and 2019-2021 are missing procedure and diagnosis information for some encounters.

Further information on implementation of ICD-10-CM is available at:

<https://www.cms.gov/Medicare/Coding/ICD10/>. Further information on cross-walking ICD-9-CM and ICD-10-CM codes is available at: <https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.html>.

8.2 Procedure Codes

Ambulatory procedures are coded in Current Procedural Terminology (CPT) for services, Healthcare Common Procedure Coding System (HCPCS) for products, supplies and services and ICD-9/10 Procedure Classification System (ICD-10-PCS) for inpatient encounters.

Further information on procedure codes can be found at <https://www.aapc.com/resources/links/> and <https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/Code-Sets/index.html>.

8.3 Revenue Codes

Revenue Codes are descriptions and dollar amounts charged for hospital services provided to a patient. There are variables in the NHCS datafiles that provide information on the services provided via revenue codes. Revenue codes are only available for UB-04 claims and Vizient data.

8.4 Discharge Status

Discharge status indicates the outcome of the hospital visit. Examples of discharge statuses include discharged home, died in hospital, and transfer from the hospital to another facility. In the 2016 NHCS, hospitals that provided EHR data are missing discharge status on many hospital encounters or provided multiple discharge statuses for one encounter. Encounters in the 2013-2015, 2019, 2020, and 2021 data have one discharge status per encounter.

8.5 Emergency Department Visits Admitted as Inpatients

UB-04 claims data identifies patients presenting in the ED who were transferred then discharged from the inpatient department (ED-to-IP). To get an accurate account of ED discharge status in NHCS, inpatient records in the ED file had their discharge status changed to “admitted as an inpatient.” NHCS does not distinguish the treatment ED-to-IP patients received in the ED setting and the inpatient setting. As a result, the data collected from the ED-to-IP encounter are identical in both settings. To avoid over counting diagnoses and services provided to ED-to-IP transfers, researchers should count diagnoses and services in the inpatient or ED setting.

Beginning in 2019, ED-to-IP was included as a setting value. When analyzing individual settings (ED or inpatient), encounters with a setting value of ED-to-IP should be included in both the ED and inpatient settings.

9. Additional Information and Contact Information

For additional questions about the NHCS data please contact us at nhcs@cdc.gov.

For more information on the NHCS data collection, please visit the survey website:
<https://www.cdc.gov/nchs/nhcs/index.htm>.