

# National Hospital Care Survey

## Research Data Center

### Documentation

Last Updated: 07/09/2024

## Table of Contents

Abstract .....	3
1. The National Hospital Care Survey Background .....	3
2. Advantages of the National Hospital Care Survey .....	3
3. Hospital Frame and Sample Design .....	4
3.1 Participating Hospitals .....	4
4. Data Sources .....	5
5. Patient Identification .....	9
6. Weighting.....	9
6.1 Weighting Methodology .....	9
7. Data Elements .....	11
8. Supplemental Datasets .....	11
8.1 Enhanced Opioid Identification Dataset .....	11
8.2 Enhanced Co-Occurring Disorders Dataset.....	12
8.3 National Death Index .....	12
8.4 U.S. Department of Housing and Urban Development .....	12
8.5 Centers for Medicare & Medicaid Services.....	12
9. Analytic Considerations .....	13
9.1 Diagnosis Codes .....	13
9.2 Procedure Codes .....	13
9.3 Revenue Codes.....	14
9.4 Discharge Status.....	14
9.5 Emergency Department Visits Admitted as Inpatients.....	14
9.6 Weighted Data Presentation Standards .....	14
9.7 Weighted Data Analysis Guide.....	14
9.8 Patient encounter weight .....	14
10. Additional Information and Contact Information.....	18
Appendix I .....	19

## 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. Unweighted data are available in the RDC for calendar years 2013-2016, and 2019-2021. Beginning in 2020, unweighted preliminary NHCS data from selected hospitals are available in the RDC. Weighted nationally representative data are available for 2020. 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/b1/datatype/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/b3/prosal/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 are now available 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](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 sampling frame for NHCS was constructed in 2011 consisted of 6,622 non-institutional and non-federal hospitals with at least six staffed inpatient beds. From that initial file, a sample of 1,000 eligible hospitals were selected.

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 was held to replenish the base sample if more hospitals of certain hospital types of the fielded sample were deemed ineligible during data collection or to capture more data from a type of hospital. 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. In 2017, the sample and frame files were updated to include newly constructed hospitals from a new source file. The updating of the NHCS sample and frame occurs every three years. Due to the addition of newly sampled birth hospitals, the 2020 base sample included 608 hospitals and the frame included 6,906 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	2,781,542	200	7,960,956	N/A	N/A	205	10,010,479	33.7
2021	197	3,689,208	236	10,025,499	N/A	N/A	247	12,803,010	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 8.1 and 8.2).

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: <a href="https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/837I-FormCMS-1450-ICN006926.pdf">https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/837I-FormCMS-1450-ICN006926.pdf</a> .	<ul style="list-style-type: none"> <li>• Administrative data used for billing rather than clinical purposes (may exclude clinically important data that was not needed for billing)</li> <li>• Allows tracking of ED encounters who were later admitted as inpatients</li> <li>• Maximum of 25 diagnosis and procedure codes provided</li> <li>• Revenue code indicator variables are available</li> </ul>
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: <a href="https://www.vizientinc.com/">https://www.vizientinc.com/</a> .	<ul style="list-style-type: none"> <li>• Administrative data used for billing rather than clinical purposes. May exclude clinically important data that was not needed for billing</li> <li>• No PII (e.g., patient name) collected and patients cannot be linked to external data sources</li> <li>• 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</li> <li>• No data on length of stay</li> <li>• Maximum of 25 diagnosis and procedure codes provided</li> <li>• Does not allow for tracking of ED encounters who were later admitted as inpatients</li> <li>• Revenue code indicator variables are available</li> </ul>

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 <a href="https://www.cms.gov/Medicare/E-Health/EHealthRecords/index.html">https://www.cms.gov/Medicare/E-Health/EHealthRecords/index.html</a>.</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"> <li>• Data has clinical rather than billing focus</li> <li>• No limit on the number of diagnoses and procedures collected per encounter</li> <li>• Does not allow for tracking of ED encounters who were later admitted as inpatients</li> <li>• Primary diagnosis is not available</li> <li>• PII collected, which allows for linkage of patients to external data sources</li> </ul>
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: <a href="https://www.hl7.org/implement/standards/product_brief.cfm?product_id=385">https://www.hl7.org/implement/standards/product_brief.cfm?product_id=385</a></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"> <li>• Data collected has a clinical not billing focus</li> <li>• No limit on the number of diagnoses and procedures collected</li> <li>• Does not allow for tracking of ED encounters who were later admitted as inpatients</li> <li>• Primary diagnosis is not available.</li> <li>• PII included, which allows for linkage to external data sources</li> </ul>

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



## 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. Weighting

Beginning in 2020, the NHCS response rate was high enough to produce nationally representative estimates. National estimates are available for hospital encounters in the emergency department and inpatient department. Section 6.1 describes the methodology used to create the estimates. The nationally represented data is available as a restricted use file, which contains the entire 2020 NHCS data. Researchers also have the option of using the 2020 NHCS public use file, which contains a 5% sample of the data collected. Table 3 presents the estimates of inpatient and ED encounters for 2020 available in the RDC file. Refer to the appendix for estimates of encounters by age, sex, discharge status, and discharge month.

**Table 3. Estimated Number Encounters by Department by Year**

Year	Inpatient Department		Emergency Department	
	Encounters	Standard Error	Encounters	Standard Error
2020	36,342,461	2,173,127	114,561,914	5,631,753

### 6.1 Weighting Methodology

Weighting was done to produce hospital level, ED visit, and inpatient discharge level estimates and accounts for sampling probabilities and nonresponse. Hospital level data were collected via self-completed forms from hospital administrators. Inpatient discharges and ED visits were collected (without sampling) from the sampled hospitals via electronic files of their claims or EHR. Participating hospitals submitted their data for the entire 2020 calendar year.

All inpatient discharge and ED visit data collected for 2020 NHCS, with the exception of eight hospitals, were used to produce the inpatient and ED weighted estimates. The eight hospitals were excluded due to quality issues with the submitted data.

The production of weights to produce national estimates involved combining data from the Premier Health Database (PHD) to the 2020 NHCS. The PHD was then removed, and the weights were adjusted for only the 2020 NHCS data. The PHD is a commercially available hospital-based database that collects inpatient service-level data from partner hospitals. It stores information about hospital characteristics such as admitting/attending physician specialties, point of origin, admission type, and discharge status. It also collects encounter-level information on inpatient visits, such as age, sex, race/ethnicity, International Classification of Diseases (ICD) codes, Current Procedure Terminology (CPT) codes, and Healthcare Common Procedure Coding System (HCPCS) codes.

Additionally, the Healthcare Cost and Utilization Project (HCUP) nationwide samples, National Inpatient

Sample (NIS) and Nationwide Emergency Department Sample (NEDS) were used as construct controls for calibrating the 2020 encounter-level weights for each setting. NIS and NEDS collect discharge level data from participating state partners. These data sources collect information including patient demographics (e.g., sex, age, race, etc.), encounter characteristics (e.g., diagnoses, procedures, diagnosis-related groups (DRG), admission types, etc.), and hospital-level characteristics (e.g., hospital size and region).

- The weighting methodology for point estimation and variance estimation was preformed using the following steps:
- Computation of original NHCS weights for NHCS hospitals that reflects the original stratified sampling design and estimated response propensities from a model fitted via a likelihood-based approach.
- Computation of original Premier weights for Premier hospitals that reflects estimated participation propensities from a model fitted via a likelihood-based approach.
- For the combined data set of NHCS and Premier, creation of original combined weights as (original combined) =  $\lambda$ (original NHCS weight) for NHCS hospitals and (original combined) =  $(1-\lambda)$ (original Premier weight) for Premier hospitals.
- Creation of replicate combined weights for variance estimation:
  - Treat the combined NHCS and Premier data as a stratified sample with NHCS strata corresponding to the original design strata and with Premier serving as its own stratum.
  - Use the stratified delete-a-group jackknife (DAGJK), with  $G_A$  groups for NHCS and  $G_B$  groups for Premier, to construct  $G_A + G_B$  sets of replicate combined weights from the original combined weights.
  - Initial assignment of groups will be individual NHCS hospitals (groups of size one hospital, so that  $G_A = n_A$ ) for NHCS and random groups of expected size  $n_B/50$  Premier hospitals (so that  $G_B = 50$ ) for Premier.
- Estimates from the combined NHCS and Premier data are then computed using the original combined weights and variance estimates are computed using the replicate combined weights.
- Creation of NHCS calibrated weights:
  - Using the original combined weights and the combined NHCS and Premier data, compute key national estimates to serve as control totals.
  - Calibrate the original combined weights for the NHCS-only subset of the combined data to the key national estimates. Unlike the original NHCS weights that are hospital-specific, these calibrated weights are encounter-specific.
- Creation of replicate NHCS calibrated weights for variance estimation:
  - Using the  $g$ th set of replicate combined weights and the combined NHCS and Premier data, compute the  $g$ th set of key national estimates to serve as control totals.
  - Calibrate the  $g$ th set of replicate combined weights for the NHCS-only subset of the combined data to the  $g$ th set of key national estimates.
- Estimates from the NHCS-only data are then computed using the NHCS calibrated weights and variance estimates are computed using the replicate NHCS calibrated weights.
- Weights and replicate weights for the public use data files were obtained by modifying the NHCS calibrated weights and replicate NHCS calibrated weights. Each replicate weight was adjusted based on the 5% public use data files sampling weight across all encounters within all hospitals.

The 2020 data included data from specialty hospitals. Due to the relatively small number of specialty hospitals in the 2020 NHCS, the inpatient weights for these hospitals were calibrated independently of the acute care hospitals on the weighted survey file. These hospitals were treated as two additional strata and stratified delete-a-group jackknife (DAGJK) method was used within each stratum.

## 7. 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](https://www.cdc.gov/rdc/data/b1/2013_NHCS.pdf)
- 2015: [https://www.cdc.gov/rdc/data/b1/2015\\_NHCS\\_RDC\\_Data\\_Dictionary.pdf](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](https://www.cdc.gov/rdc/data/b1/2016_NHCS_DATA-DICTIONARY.pdf)
- 2019-2021: <https://www.cdc.gov/rdc/data/b1/2019-2020-NHCS-RDC-Data-Dictionary-508.pdf>

## 8. 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.

### 8.1 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](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>
- 2019-2021: <https://www.cdc.gov/rdc/data/b1/2021-Opioid-RDC-508.pdf>

## **8.2 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](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>
- 2019-2021: <https://www.cdc.gov/rdc/data/b1/FY19-RDC-2021-Co-Occurring-508.pdf>

## **8.3 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>.

## **8.4 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>.

## **8.5 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>.

## 9. 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, the 2020 NHCS weighted files are available in the RDC and the 2021 weighted files will be available in the RDC soon.

### 9.1 Diagnosis Codes

Due to the implementation of ICD-10-CM codes in October 2015, 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>.

### 9.2 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>.

### **9.3 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.

### **9.4 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.

### **9.5 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.

### **9.6 Weighted Data Presentation Standards**

Data users should be aware of the reliability of survey estimates, particularly smaller estimates. NCHS has published guidance for the assessment of reliability and presentation of proportions (or percentages) ([https://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_175.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf)) and for the presentation of rates and counts ([https://www.cdc.gov/nchs/data/series/sr\\_02/sr02-200.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02-200.pdf)). For presentation or publication of NHCS estimates, we recommend estimates be rounded to the nearest thousand.

These presentation standards apply to products published by NCHS. If, according to the presentation standards, an estimate is not reliable, data users should examine the confidence interval carefully before using the estimate.

### **9.7 Weighted Data Analysis Guide**

The following section provides an overview on how data users can derive visit or discharge estimates and compute variances to produce standard errors, using statistical software tools such as SAS, R, and Stata. SAS/STAT® software provides a set of procedures whose names begin with SURVEY for survey analysis. R relies on the “survey” package to conduct survey data analysis whereas Stata, uses the “svy” command. SAS, R, and Stata users can use these procedures to conduct statistical analysis on data from the 2020 NHCS public use data files.

### **9.8 Patient encounter weight**

The patient encounter weight is a vital component in the process of producing estimates from sample data and its use should be clearly understood by all data users. By aggregating the “patient encounter weights” assigned to the PUF\_ENCWGT\_BASE variable on the all inpatient discharges and ED visits for 2020, the data user can obtain the estimated total of 36,342,461 inpatient discharges (standard error of 2,173,127 inpatient discharges) and 114,561,914 ED visits (standard error of 5,631,753 ED visits) made in the United States.

Note that estimates of inpatient discharges and ED visits produced from the 2020 NHCS restricted use data files may differ somewhat from those estimates produced from the NHCS public use data files. This is because of adjustments required for the public use data files, as part of the disclosure risk mitigation process.

## **SAS Survey Procedures**

### **Specifying the VARMETHOD Option**

To correctly use the SURVEYFREQ or SURVEYMEANS procedures, the VARMETHOD= option must be added to the PROC statement to indicate the method to develop the replicate weights. For NHCS, this method is the stratified delete-a-group jackknife (DAGJK) method. The option is always VARMETHOD=JACKKNIFE.

For example, in PROC SURVEYMEANS:

```
PROC SURVEYMEANS DATA = FILE VARMETHOD=JACKKNIFE;
```

### **Declaring the REPWEIGHTS Statement**

To properly use the SURVEYFREQ or SURVEYMEANS procedures, the REPWEIGHTS options must be specified. The REPWEIGHTS statement identifies the variables containing replicate weights, the statement is—

```
REPWEIGHTS PUF_ENCWGT_1 – PUF_ENCWGT_XXX / JKCOEFS=1;
```

In the 2020 NHCS, there are 122 replicate weights on the inpatient public use data file, so “PUF\_ENCWGT\_XXX” would be “PUF\_ENCWGT\_122.” There are 100 replicate weights on the ED public use data file, so “PUF\_ENCWGT\_XXX” would be “PUF\_ENCWGT\_100.”

### **Example SAS Code**

The programs below demonstrate how to use replicate weights and calculate variance estimates using SAS SURVEYFREQ and SURVEYMEANS procedures.

#### **For categorical variables**

```
PROC SURVEYFREQ DATA=FILE VARMETHOD=JACKKNIFE;  
TABLE VAR1; * REPLACE “VAR1” WITH THE CATEGORICAL VARIABLE OF INTEREST;  
REPWEIGHTS PUF_ENCWGT_1 – PUF_ENCWGT_XXX / JKCOEFS=1; * CHANGE  
“PUF_ENCWGT_XXX” TO PUF_ENCWGT_122 FOR INPATIENT OR PUF_ENCWGT_100 FOR  
ED;  
WEIGHT PUF_ENCWGT_BASE;  
ODS OUTPUT ONEWAY=ONEWAY;  
RUN;
```

### For continuous variables

```
PROC SURVEYMEANS DATA=FILE VARMETHOD=JACKKNIFE;  
VAR VAR1; * REPLACE "VAR1" WITH THE CONTINUOUS VARIABLE OF INTEREST;  
REPWEIGHTS PUF_ENCWGT_1 – PUF_ENCWGT_XXX / JKCOEFS=1; * CHANGE  
"PUF_ENCWGT_XXX" TO PUF_ENCWGT_122 FOR INPATIENT OR PUF_ENCWGT_100 FOR  
ED;  
WEIGHT PUF_ENCWGT_BASE;  
ODS OUTPUT STATISTICS= STATISTICS;  
RUN;
```

### R Survey Procedures

The R package “survey” can be utilized for complex survey analysis (<https://cran.r-project.org/web/packages/survey/index.html>). The R programs below demonstrate how to use replicate weights and calculate variance estimates in the “survey” package.

```
install.packages("survey")  
library(survey)  
install.packages("tidyverse")  
library(tidyverse)
```

### For categorical variables:

```
# PUF_ENCWGT_1 to PUF_ENCWGT_XXX are the replicate weights  
# For inpatient, there are 122 replicate encounter weights, so XXX = 122  
# For ED, there are 100 replicate encounter weights, so XXX = 100  
# Insert filename for "FILE"
```

```
replicate_weights <- FILE%>%select('PUF_ENCWGT_1':PUF_ENCWGT_XXX')
```

```
# Create a survey design object  
# Insert filename for "FILE"  
survey_design <- FILE(ids=~1, data=FILE,  
  weight =FILE$PUF_ENCWGT_BASE,  
  repweights= replicate_weights,  
  type="other",  
  scale = 1,  
  rscales = 1,  
  mse=TRUE )
```

```
#Replace VAR1 with the categorical variable of interest  
# Conduct survey weighted frequency analysis  
survey_freq <- svytable(~VAR1, survey_design)
```

```
# View the survey frequency results  
print(survey_freq)
```

### For continuous variables

```
# PUF_ENCWGT_1 to PUF_ENCWGT_XXX are the replicate weights  
# For inpatient, there are 122 replicate encounter weights, so XXX = 122  
# For ED, there are 100 replicate encounter weights, so XXX = 100
```



```

# Insert filename for "FILE"
install.packages("survey")
library(survey)

replicate_weights <- FILE%>%select('PUF_ENCWT_1':PUF_ENCWT_XXX')

# Create a survey design object
# Insert filename for "FILE"
survey_design <- FILE(ids=~1, data=FILE,
  weight =FILE$PUF_ENCWT_BASE,
  repweights= replicate_weights,
  type="other",
  scale = 1,
  rscales = 1,
  mse=TRUE
)

#Replace VAR1 with the continuous variable of interest
survey_means <- svymean(~VAR1, survey_design)

# View the survey means results
print(survey_means)

```

### Stata Survey Procedures

The Stata programs below demonstrate how to use replicate weights and calculate variance estimates with the svyset command (<https://www.stata.com/manuals/svysvyset.pdf>).

#### For categorical variables

```

*Set survey design
svyset PUF_ID, weight(PUF_ENCWT_BASE)

*Set replicate weights, change "PUF_ENCWT_XXX" to PUF_ENCWT_122 for inpatient or
PUF_ENCWT_100 for ED
svyset [pweight=PUF_ENCWT_BASE], vce(jackknife) jkrweight(PUF_ENCWT_1-
PUF_ENCWT_XXX)

*Specify one-way tables, change "VAR1" to categorical variable of interest
svy: tab VAR1, count se
*Specify one-way tables, change "VAR1" to categorical variable of interest
svy: tab VAR1, percent

```

#### For continuous variables

```

*Set survey design
svyset PUF_ID, weight(PUF_ENCWT_BASE)

*Set replicate weights, change "PUF_ENCWT_XXX" to PUF_ENCWT_122 for inpatient or
PUF_ENCWT_100 for ED
svyset [pweight=PUF_ENCWT_BASE], vce(jackknife) jkrweight(PUF_ENCWT_1-
PUF_ENCWT_XXX)

*Specify one-way tables, change "VAR1" to continuous variable of interest
svy: mean VAR1

```

## **10. Additional Information and Contact Information**

For additional questions about the NHCS data please contact us at [nhcs@cdc.gov](mailto:nhcs@cdc.gov).

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

## Appendix I

**Table 4. Frequencies of selected variables from the weighted 2020 NHCS restricted use file.**

Variable	Restricted Use Emergency Department Data File				Restricted Use Inpatient Data File			
	Unweighted Count	Weighted			Unweighted Count	Weighted		
		Count	Std. Error	%		Count	Std. Error	%
<b>Age (in years)</b>								
Under 1	135,018	1,625,536	182,786	1.4	299,588	4,019,203	357,865	11.1
1-17	972,895	14,136,097	978,521	12.3	112,332	1,822,994	401,161	5.0
18-44	3,083,235	47,881,234	2,199,155	41.8	668,435	9,266,652	618,432	25.5
45-64	1,963,972	28,594,994	1,499,431	25.0	642,647	8,313,822	480,417	22.9
65-74	771,265	10,789,722	710,914	9.4	432,504	5,641,533	380,754	15.5
75 and over	837,490	11,455,083	961,783	10.0	496,790	7,245,679	536,847	19.9
Missing	11,069	79,247	82,724	0.1	1,584	32,578	30,783	0.1
<b>Sex</b>								
Male	3,616,311	52,083,659	2,656,643	45.5	1,214,397	16,301,395	947,177	44.9
Female	4,146,674	62,351,492	3,016,048	54.4	1,437,897	20,000,798	1,249,506	55.0
Missing	11,959	126,763	89,996	0.1	1,586	40,268	34,166	0.1
<b>Discharge status</b>								
Routine to Home	6,280,325	99,150,961	4,445,134	86.5	1,717,716	23,995,053	1,487,301	66.0
Left against medical advice	150,898	2,189,987	158,159	1.9	36,993	535,027	63,037	1.5
Transfer to short term facility	105,930	2,204,930	195,949	1.9	25,296	677,526	65,479	1.9
Transfer to long term facility	42,575	409,031	72,116	0.4	88,405	1,088,492	125,091	3.0
Home health care	121,534	1,477,416	338,695	1.3	368,387	4,525,721	465,387	12.5
Hospice care - home or medical facility	23,621	236,296	56,557	0.2	47,334	710,366	86,710	2.0
Other	450,382	4,205,577	517,510	3.7	246,651	3,482,340	268,650	9.6
Dead	41,273	380,193	31,875	0.3	72,707	922,816	86,101	2.5
Missing	558,406	4,307,523	1,642,748	3.8	50,391	405,121	217,399	1.1
<b>Discharge month</b>								
January	895,631	13,120,600	760,780	11.5	250,989	3,567,564	212,643	9.8
February	815,526	11,994,536	682,015	10.5	233,131	3,303,877	201,237	9.1
March	716,063	10,454,374	575,262	9.1	222,254	3,120,442	195,190	8.6
April	419,492	6,243,975	322,949	5.5	174,397	2,446,680	145,625	6.7
May	503,667	7,615,398	437,545	6.6	200,654	2,741,202	173,324	7.5
June	558,200	8,421,553	416,572	7.4	214,088	2,929,595	185,720	8.1
July	664,429	9,788,941	496,944	8.5	229,713	3,147,032	194,042	8.7
August	685,486	10,062,811	620,626	8.8	225,440	3,022,906	195,075	8.3
September	636,584	9,412,833	574,582	8.2	224,598	3,003,929	177,470	8.3
October	674,036	10,098,244	605,686	8.8	237,831	3,186,507	203,642	8.8
November	614,721	8,831,836	490,408	7.7	215,403	2,859,902	191,226	7.9
December	591,109	8,516,813	487,726	7.4	225,382	3,012,827	210,362	8.3