2021 NATIONAL HOSPITAL CARE SURVEY (NHCS) TECHNICAL DOCUMENTATION

For Public Use Data Files

Division of Health Care Statistics
National Center for Health Statistics
December 2024

Overview Summary

This document provides detailed information and guidance for users of the 2021 National Hospital Care Survey (NHCS) public use data files. As a principal source of information on health care utilization in the United States, NHCS collects inpatient and emergency department (ED) data from a nationally representative sample of U.S. hospitals sourced from administrative and electronic health records data. The 2021 NHCS is conducted by the National Center for Health Statistics (NCHS) and is a member of the National Health Care Surveys – a family of surveys which measure health care utilization across a variety of health care providers and settings.

Section 1 of this document includes information on the scope of the survey, its sources of data, and the confidentiality protections related to the data. Section 2 contains details on the sampling process, data collection procedures, and weighting methodology used to produce national estimates on hospital utilization. Section 3 provides information on the number of sampled hospitals that were eligible to participate in NHCS and submitted data in 2021. Section 4 details the contents of the 2021 NHCS public use data files and the edits used in the creation of the files. Section 5 contains an explanation of the procedures used to accurately produce variance estimates. National Center for Health Statistics presentation standards for proportions, counts, and rates, and their relation to NHCS data are discussed in Section 6, and the data analysis guidelines are provided in Section 7. Section 8 provides variable missingness by both inpatient and ED settings, and Section 9 provides a comparison of frequencies between the NHCS public use data files and NHCS restricted use data file, as well as to the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project (HCUP) nationwide samples. Section 10 provides a list of preferred reporting items for complex sample survey analysis for 2021 NHCS. Section 11 provides further information on the availability of the NHCS restricted use data files available in NCHS and Federal Research Data Centers.

Suggested Citation

Technical Documentation: National Center for Health Statistics. Division of Health Care Statistics. *2021 National Hospital Care Survey (NHCS) Public Use Data Files Documentation,* December 2024. Hyattsville, Maryland.

Data File: National Center for Health Statistics. Division of Health Care Statistics. 2021 National Hospital Care Survey (NHCS) public use data files. 2024. Hyattsville, Maryland.

Contact Information

Data users can find the latest information about NHCS on our website, at: https://www.cdc.gov/nchs/nhcs/index.htm. If data users have queries about the public use data files, they may send their question through email to nhcs@cdc.gov, or call us at 301-458-4321. A response to data user inquiries is generally provided in 1-2 business days.

The NHCS also has a listserv, where updates and information about the most recent NHCS news and data releases are sent out. To join the listserv, there are two options:

1. Send an email message to listserv@cdc.gov. Leave the subject line blank. In the body of the message, type or paste: **subscribe NHCS-DATA your name** where "your name" is your first and last name.
2. Fill out and submit a subscription request form through the NCHS listservs webpage (https://www.cdc.gov/nchs/products/nchs_listservs.htm). You should receive an email notification of your subscription.

Contents

Section 1 About the National Hospital Care Survey

Section 1.1 Background

The National Hospital Care Survey (NHCS) provides data on health care utilization in hospital-based settings, including inpatient departments and emergency departments (EDs). Data collection for NHCS began in 2011, integrating two long-standing National Center for Health Statistics (NCHS) surveys: the National Hospital Discharge Survey, the longest continuously fielded sample of inpatient care from 1965-2010; and the National Hospital Ambulatory Medical Care Survey (1992-2022), which was comprised of three data collection components: hospital EDs (1992-2022), outpatient departments (1992-2017), and ambulatory surgery centers/locations (2009-2017).

NHCS collects data on inpatient discharges and ED visits including information on diagnoses and procedures, length of stay, and patient demographics. NHCS aims to provide health trends and outcomes of the U.S. population's utilization of hospitals in the following ways:

- Provide nationally representative, accurate, and reliable health care data for hospital-based settings.
- Answer key questions of interest to health care professionals, researchers, and policy makers about health care quality, use of resources, and disparities of services to population subgroups.
- Monitor national trends in substance use-related emergency department visits, including opioid visits.
- Provide data that links episodes of care across hospital units, as well as to other data sources, such as the National Death Index.
- Contribute to a stronger public health infrastructure that helps address current and future public health threats.

For 2021, the NHCS sample included 608 non-institutional, non-federal hospitals with six or more staffed inpatient beds in the 50 U.S. states and the District of Columbia. In 2021, there were 228 hospitals that provided data on inpatient discharges and ED visits. Overall, 179,026 inpatient discharges and 482,126 ED visits were selected to create the 2021 NHCS public use data files.

Section 1.2 Data Sources

NHCS receives data from billing and electronic health record sources. This includes Uniform Billing (UB)-04 administrative claims data, the accepted electronic standard for hospital billing as mandated by the

Centers for Medicare & Medicaid Services (https://www.nubc.org/). Additionally, hospitals can submit data from Vizient, a member-driven health care services company that collects encounter data from hospitals prior to submitting data to NHCS (https://www.vizientinc.com/). Vizient data are similar to UB-04 administrative claims data, but do not contain patient identifiers. Alternatively, hospitals may submit electronic health record (EHR) data, which contain an unlimited number of medical diagnosis and procedure codes, laboratory and medication data, and unstructured clinical notes. NHCS only accepts EHR data in the format of HL7 CDA® R2 Implementation Guide: National Health Care Surveys Release 1, DSTU Release 1.2 – US Realm (http://www.hl7.org/implement/standards/product_brief.cfm?product_id=385).

Starting in 2020, NHCS also received data from the hospital database of the American College of Emergency Physicians (ACEP) for sampled non-participating hospitals (https://www.acep.org/). The data from ACEP are only from EDs and do not include any inpatient data, nor does the data include any hospital or patient identifiers.

Section 1.3 Data Confidentiality

NCHS and its agents take the security and confidentiality of NHCS data very seriously. Strict laws have been implemented to establish minimum Federal standards for safeguarding the privacy of individually identifiable health information. Assurance of confidentiality is provided to all hospitals according to Section 308(d) of the Public Health Services Act [42 United States Code 242m (d)]. Strict procedures according to Section 3572 of the Confidential Information Protection and Statistical Efficiency Act (44 U.S.C. 3561-3583) are utilized to prevent disclosure of personal identifiable information in NHCS data. All information which could identify a participating hospital is confidential and seen only by persons associated with NHCS and is not disclosed or released to others for any other purpose. Prior to the release of public use data files, NCHS conducts extensive disclosure risk analysis to minimize the chance of inadvertent disclosure. As a result, selected characteristics and/or data elements may have been masked on the public use data files to minimize the potential risk of disclosure. Masking was performed in such a way to cause minimal impact on the data. See Section 4: Data Processing for more information on which variables in the public use data files were impacted.

The protocol for NHCS has been approved by the NCHS Research Ethics Review Board annually since the survey's establishment (2011).

Section 2 Methodology

Section 2.1 Brief Overview

The 2021 NHCS used a national probability sample of hospitals to collect data on inpatient discharges and ED visits to develop the public use data files. The 2021 NHCS public use data files sample was designed to provide estimates for encounters in both inpatient and ED settings. NHCS data are extracted from hospital billing or EHR systems and transmitted electronically to NCHS for processing. These data are sourced from non-institutional, non-federal hospitals with six or more staffed inpatient beds in the 50 U.S. states and District of Columbia.

Section 2.2 Hospital Frame and Sample Design

The initial sampling frame for NHCS was constructed in 2011 and 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 was 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 2021 base sample included 608 hospitals and the frame included 6,906 hospitals.

Section 2.3 NHCS Public Use Data Files Sample Design

While the NHCS restricted use data file includes every hospital encounter record submitted to NHCS for the survey year, the 2021 NHCS public use data files consist of a 5% sample of NHCS encounter data from each setting (i.e., both inpatient and ED). In 2021, NHCS collected 3,672,491 inpatient and 9,980,942 ED records. Systematic random sampling was used to select the samples for the public use data files of inpatient and ED records.

This 5% sample of NHCS records was selected for the public use data files instead of the full listing of records for the following reasons: (1) inclusion of all records from inpatient discharges and ED visits from

all sampled hospitals would pose unacceptable disclosure risks for the hospitals; and (2) NHCS collects a lot of records that can be difficult for statistical software to process efficiently, so public use data files were reduced to a manageable size for data users.

Section 2.3.1 Sampling Inpatient Discharge Records

A targeted number of records was determined by taking 5% of the inpatient records (n=179,026). The sampling interval was the total number of inpatient records divided by the targeted sample. The sampling interval used to select the public use data file records in the 2021 NHCS was 20. For each hospital, a random number was selected between 0 and the sampling interval (TE). For each discharge, a random number was selected between 1 and the number of records submitted within each hospital. Inpatient discharges were then sorted by the following variables within a sampled hospital:

Admission Type → Age → Sex → Length of Stay (LOS) → Expected Source of Payer → Discharge Date → Random Discharge Number

Next from the ordered array of records for hospital (j), the records whose assigned "array sequential" numbers are equal to or first after the following numbers were selected for the public use data file sample:

$$R_j + L \times TE$$

Where:

R_j = random number between 0 and TE

L = 0, 1, 2, ….

TE = sampling interval

The sampled records were then sequentially numbered in the order in which they were selected from the hospital's submitted records.

Section 2.3.2 Sampling ED Visit Records

For ED visits (n=482,126), like inpatient discharges, random numbers were assigned 1 through the number of records submitted by a hospital. Within each hospital, events were sorted as follows:

Hospital ID → Age → Sex → Discharge Status → Expected Source of Payment → Encounter End Date → Randomly assigned number for selection

Then within date of visit, the ED visits were sorted to the numbers randomly assigned at the start of the process. The sorted events were then sequentially numbered from each hospital in the sorted order.

Section 2.4 Data Collection Procedures

In 2021, there were multiple methods for hospitals to submit hospital billing or EHR data. Hospitals could directly submit UB-04 administrative claims or EHR data. Additionally, data could be provided by third-party entities, like Vizient or ACEP. Participating hospitals were required to submit one type of data (e.g., UB-04 administrative claims or EHR, not both). Once data were collected, several steps were required for data processing. Specifications for checking, configuring, and transmitting the data files were developed by NCHS. Once NCHS received the data files, they were processed to harmonize and integrate the data to create a database with similar variables across sources.

Section 2.5 Weighting

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 in-scope hospitals defined earlier. Inpatient discharges and ED visits were collected (without sampling) from the sampled hospitals via electronic files of their claims or EHR. Participating hospitals submitted all of their data for 2021. While NHCS restricted use data files include all (or 100%) of records sent, the public use data files include a 5% sample of those records.

All inpatient discharge and ED visit data collected for 2021, except for 19 hospitals, were used to produce the inpatient and ED weighted estimates. The 19 hospitals excluded were done so 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 2021 NHCS. The PHD was then removed, and the weights were adjusted for only the 2021 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 2021 encounter-level weights for each setting. NIS collects discharge level and NEDS collects emergency visit 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). One thing of note is that we did not receive 2021 HCUP data. Due to this, we instead adjusted the 2020 HCUP data based on information reported on HCUP's website about 2021 estimates. We did a 3% increase on the ED data and for IP we used the same distributions that were in the HCUP data.

The weighting methodology for point estimation and variance estimation consisted of 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) = λ(original NHCS weight) for NCHS hospitals and (original combined) = (1-λ)(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 2021 data included a small number of both NHCS specialty hospitals and Premier specialty hospitals. Due to the relatively small number of specialty hospitals in the 2021 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 eight weight categories and the DAGJK method was used within each weight category.

Section 3 Sample Size, Eligibility, and Response Rate

An eligible responding NHCS hospital must have provided data for at least half the months the hospital was in-scope during the survey year. A hospital is considered a partial respondent if they provided less than half of the months for which the hospital was in scope for the NHCS in the survey year. For example, a hospital that was in-scope for all 12 months during the year must have provided at least 6 months of data.

Of the 247 participating hospitals that were included in the 2021 NHCS database sample, 228 (92.3%) were selected to create the public use data files sample. 19 hospitals that submitted data in 2021 were not included in the public use data files selection due to quality issues identified with the submitted data. From the 228 hospitals, 5% of all records were selected by setting for the public use data files. Overall, 179,026 inpatient discharges and 482,126 ED visits were selected. ED patients transferred to the inpatient department are counted in both the inpatient and ED settings. Table 3.1 presents the number of hospitals, encounters, and response rates of NHCS by setting and type of sample population. See Table 3.1 below for more information.

Table 3.1. Number of hospitals, encounters, and unweighted response rates by setting and file type, National Hospital Care Survey, 2021

	Inpatient		Emergency Department		Total		
	Hospitals	Encounters	Hospitals	Encounters	Hospitals	Encounters	Unweighted Response Rate[1]
Restricted Use Data File	197	3,672,491	236	9,980,942	247	12,745,505	40.6
Public Use Data File	191	179,026	209	482,126	228	661,152	N/A

N/A is not applicable.

[1] Response rate calculation is the number of hospitals that provided data (247) divided by the number of sampled hospitals (608), multiplied by 100.

Section 4 Data Processing

Certain NHCS data elements with missing values were imputed prior to creation of the public use data files. This included imputing values for patients with missing sex, age, and length of stay information. For the 2021 NHCS public use data files, 600 (0.34%) inpatient discharges and 1,575 (0.33%) ED visits contained at least one imputed value for age, sex, and/or length of stay. The data included in the public use data files underwent additional processing to prepare them for release. In addition to imputing missing data, suppression rules such as masking were applied for some records to protect patient confidentiality. Other variables such as age or length of stay were either top-coded or bottom-coded in accordance with NCHS confidentiality requirements; this is noted for specific variables outlined in this section.

Section 4.1 Diagnosis Data

NHCS collects data from participating hospitals from various sources. The diagnosis coding systems varies by data source. Diagnoses from UB-04 administrative claims data are exclusively coded in *International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)*. EHR submitting hospitals provided diagnoses in SNOMED Clinical Terms (CT) and ICD-10-CM. In the creation of a harmonized and integrated hospital database the SNOMED CT diagnosis codes were translated to ICD-10-CM where applicable. Translation from SNOMED CT to ICD-10-CM was the only modification to the diagnosis codes. On the public use data files, medical diagnosis codes were limited to ICD-10-CM diagnosis codes.

An ICD-10-CM code can have a maximum of 7 characters and is organized by chapters from A to Z. For the 2021 NHCS public use data files, ICD-10-CM codes have been truncated to four characters to minimize disclosure risks. Rare diagnoses that fail to meet the Rule of 5 have been truncated to two characters. Diagnosis codes are presented as original codes provided by the hospital. While the codes have been truncated, the diagnosis codes are never updated or revised to a different code that would result in a change to the original diagnosis for a visit or hospitalization. To maintain integrity of the data, any codes that appear to be invalid are kept as is.

Duplicate 4-character ICD-10-CM codes were removed for each unique encounter on both the inpatient and ED public use data files. Diagnosis codes were limited to 30 unique records per encounter (variables DX1 through DX30 in the public use data files). A flag was created for encounters with 31 or more

diagnosis codes, prior to removal of diagnosis codes above the 30-diagnosis code threshold (variable DX_GT30 in the public use data files).

To account for diagnosis codes that may have been excluded after the removal of diagnosis codes above 30 per encounter, category flags from the Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP) Clinical Classification Software Refined (CCSR) for ICD-10-CM diagnosis codes were created to provide information captured in the removed diagnoses. The AHRQ HCUP CCSR aggregates ICD-10-CM diagnosis codes into clinically meaningful categories. This is intended to be used analytically to examine patterns of health care among several factors.

The AHRQ HCUP CCSR categories are used to provide information from all ICD-10-CM diagnosis codes collected in NHCS, including diagnosis codes that were removed from the public use data files because an encounter had more than 30 diagnoses. All ICD-10-CM diagnosis codes for records in the inpatient and ED public use data files were mapped to a clinical category as defined by the CCSR and a CCSR category flag was created. The 20 CCSR categories with the highest counts were included on the inpatient and ED public use data files, respectively.

CCSR v2024.1 version was used for creating the CCSR categories. More information on the CCSR for ICD-10-CM diagnoses, including documentation, can be found here: https://hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp

Section 4.2 Procedure Data

The coding system used for classifying procedures that occurred during the hospital visit depends on the hospital setting. Inpatient procedures are captured by the *International Classification of Diseases, Tenth Revision, Procedure Coding System (ICD-10-PCS)*. Procedures occurring in the ED are captured using Current Procedural Terminology (CPT®) which is created and maintained by the American Medical Association. CPT codes have been excluded from the ED public use data file at this time due to limitations of approved use in the CPT data user's agreement. Therefore, only ICD-10-PCS codes are available on the inpatient public use data file; procedure codes are unavailable on the ED public use data file.

For the NHCS inpatient public use data file, ICD-10-PCS codes have been truncated to three characters or less to minimize disclosure risks. Rare procedure codes are removed from the PUF files. Duplicate 3-character ICD-10-PCS codes were removed from the inpatient discharges on the public use data file. Procedure codes were limited to a maximum of 10 per encounter (listed in the public use data file as

variables PD1 through PD10). A flag was created for inpatient discharges with 11 or more procedure codes, prior to removal of procedure codes above the 10-procedure code threshold (variable PD_GT10 in the public use data file).

Section 4.3 Data Quality

Data users and researchers should be aware of the possibility that the data may contain implausible combinations (e.g., a male patient with a uterine cancer diagnosis) due to coding errors during the collection process. To preserve the data's integrity, errors within the data have been retained without modification or correction. As a result, data users are urged to exercise discretion in determining how they address such records during their analyses.

Section 4.4 Patient Age

Patients with missing age underwent an imputation process that assigned an age based on donor records with a similar first listed diagnosis. Imputed age values comprised 0.21% of inpatient discharges and 0.29% of ED visits. To reduce disclosure risk, encounters of patients ages 86 and older were top coded to 86 years.

Section 4.5 Newborn Age

A newborn flag identifying encounters of infant patients under 1 month was added to both the inpatient and ED public use data files.

Section 4.6 Sex

Patients with missing sex underwent an imputation process that assigned sex based on donor records with similar first listed diagnosis. Imputed sex values comprised 0.23% of inpatient discharges and 0.29% of ED visits.

Section 4.7 Discharge Status

To minimize disclosure risk, four categories ("other discharge not specified," "admitted as inpatient," "court/law enforcement," "other health care facility") for discharge status were collapsed into the "other" category. The "missing" category is composed of three collapsed categories ("Invalid," "Unknown," and "Missing").

Section 4.8 Length of Stay

Length of stay (LOS) is only available for inpatient discharges and is calculated based on the start and end dates of the encounter. If a patient has a missing length of stay due to missing encounter date information, their length of stay value was imputed based on a donor record with a similar procedure as the encounter with a missing length of stay. The percentage of inpatient discharges with an imputed LOS value is 0.09%. To minimize disclosure risk, encounters for patients with inpatient LOS durations over 14 days were top coded to 15 days. Additionally, encounters for patients with LOS of 30 or more days were flagged (variable LOS_30Days in the public use data file).

Section 4.9 Discharge Month

To minimize risk of disclosure, exact dates are not provided and instead only the month of the inpatient discharge and ED visit are provided on the inpatient and ED public use data files.

Section 5 Standard Errors and Variance Estimation

Standard error is primarily a measure of the sampling variability that occurs by chance because only a sample of hospitals are in NHCS, rather than the entire universe of hospitals. Standard errors and other measures of sampling variability are best determined by using a statistical software package that considers the sample designs of surveys to produce such measures.

Propensity modeling was used to develop the base weight through variable selection and harmonization. Replicate calibrated weights were created to produce variance estimates, using the stratified delete-a-group jackknife (DAGJK) method. The DAGJK method resulted in the creation of 100 replicated calibrated weights for the ED public use data file and 130 replicated calibrated weights for the inpatient public use data file to match key national estimates.

The SAS code provided below can be used to generate weighted estimates with standard errors:

```
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_100 FOR ED OR PUF_ENCWGT_130 FOR INPATIENT;
WEIGHT PUF_ENCWGT_BASE;
ODS OUTPUT ONEWAY= ONEWAY;
RUN;
```

See Section 7 for further guidance on how to apply weights and calculate relative standard errors to generate national estimates.

Section 5.1 Subpopulation Analysis – Subsetting Data

For data users who may have a subpopulation of interest, such as a particular age group or sex, a domain analysis, also known as a subgroup or subpopulation analysis must be performed.

For some variance estimation methods, the entire set of data containing the appropriate weights for a particular survey year must be used to obtain the correct variance estimates. Therefore, it is not recommended to drop observations from the dataset when subsetting data, as it may affect variance estimation.

Instead, the estimation procedure must indicate which records are in the subgroup of interest. For example, when examining female patients aged 35 and over, the entire dataset of examined individuals, (both male and female patients of all reported ages) must be read into the statistical software program.

The STAT and DOMAIN statements in the SAS survey procedure, or comparable statements in other programs (SUBSET in R; SUBPOP or over in Stata) must be used to indicate the subgroup of interest (i.e., females aged 35 and over).

Depending on the specifications of a data user's statistical software of choice, an indicator variable created by the data user prior to the procedure may facilitate the identification of the subgroup in the procedure statements.

Section 6 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) and for the presentation of rates and counts (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.

Section 7 Data Analysis Guidance

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 2021 NHCS public use data files.

Section 7.1 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. The statistics contained on the public use data files reflect only a sample of encounters; a 5% sample of the NHCS data collected from sampled hospitals, not a complete count of all encounters that occurred in the United States. Each inpatient discharge and ED visit record in the public use data files represents one patient encounter in the sample of 179,026 inpatient discharges and 482,126 ED visits. To obtain national estimates from the 5% sample, each record is assigned an inflation factor called the "patient encounter weight" (variable PUF_ENCWGT_BASE in the public use data files).

By aggregating the "patient encounter weights" assigned to the PUF_ENCWGT_BASE variable on the 179,026 inpatient discharges and 482,126 ED visits for 2021, the data user can obtain the estimated total of 35,613,997 inpatient discharges (standard error of 1,193,378 inpatient discharges) and 118,327,593 ED visits (standard error of 2,459,558 ED visits) made in the United States.

Note that estimates of inpatient discharges and ED visits produced from the 2021 NHCS public use data files may differ somewhat from those estimates produced from the NHCS restricted use data files. This is because of adjustments required for the public use data files, as part of the disclosure risk mitigation process. Certain variables were masked on some records for confidentiality purposes. Other variables were top and/or bottom coded in accordance with NCHS confidentiality requirements.

The tables in Section 9 compare aggregate unweighted and weighted data for selected variables between the 2021 NHCS public use data files and restricted use data file by setting.

Section 7.2 SAS Survey Procedures

Section 7.2.1 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. Thus, the option is always VARMETHOD=JACKKNIFE.

For example, in PROC SURVEYMEANS:

PROC SURVEYMEANS DATA = FILE VARMETHOD=JACKKNIFE;

Section 7.2.2 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. When using NHCS data, the statement is—

REPWEIGHTS PUF_ENCWGT_1 – PUF_ENCWGT_XXX / JKCOEFS=1;

There are 130 replicate weights on the inpatient public use data file, so “PUF_ENCWGT_XXX” would be “PUF_ENCWGT_130.” There are 100 replicate weights on the ED public use data file, so “PUF_ENCWGT_XXX” would be “PUF_ENCWGT_100.”

Section 7.2.3 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_130 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_130 FOR INPATIENT OR PUF_ENCWGT_100 FOR ED;
WEIGHT PUF_ENCWGT_BASE;
ODS OUTPUT STATISTICS= STATISTICS;
RUN;
```

Section 7.3 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)
```

For categorical variables:

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

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

For continuous variables:

```
# Replace “VAR1” with the continuous variable of interest
# Replace “FILE” with filename
# Conduct survey weighted mean analysis
survey_means <- svymean(~VAR1, FILE)

# View the survey means results
print(survey_means)
```

Section 7.4 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_ENCWGT_BASE)
```

*Set replicate weights, change “PUF_ENCWGT_XXX” to PUF_ENCWGT_130 for inpatient or PUF_ENCWGT_100 for ED
svyset [pweight=PUF_ENCWGT_BASE], vce(jackknife) jkrweight(PUF_ENCWGT_1-PUF_ENCWGT_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_ENCWGT_BASE)

*Set replicate weights, change “PUF_ENCWGT_XXX” to PUF_ENCWGT_130 for inpatient or PUF_ENCWGT_100 for ED
svyset [pweight=PUF_ENCWGT_BASE], vce(jackknife) jkrweight(PUF_ENCWGT_1-PUF_ENCWGT_XXX)

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

Section 8 Survey Content

For the 2021 NHCS public use data files, 202 variables were included in the inpatient public use data file and 159 variables were included in the ED public use data file. In the inpatient and ED public use data files, 51 (25.2% and 32.1% respectively) variables include medical diagnoses information. In the inpatient public use data file, 11 (5.4%) variables include medical procedure information, and two variables measure LOS (1.0%). The ED public use data file does not include information on procedures or LOS.

Please refer to the 2021 NHCS public use data files codebooks for detailed information on the variables including variable names, variable type, variable descriptions, and variable values.

Section 8.1 Variable Missingness Rate by Setting

In the ED public use data file, 29 variables (18.2%) had an unweighted missingness rate that was greater than 5%. In the inpatient public use data files, 39 variables (19.3%) had an unweighted missingness rate that was greater than 5%. See Section 8.1.1 and 8.1.2 for additional information.

However, it is expected that most of the diagnosis variables after the first-listed diagnosis and procedure variables will have a high missingness percentage. This is because not all encounters are expected to include more than one or multiple procedures or diagnoses listed. Excluding the non-first-listed diagnosis and procedure variables and the AHRQ CCSR variables, there is only one inpatient variables (variable PD_GT10 in the public use data file) and two ED variables (variables DISCHARGE_STATUS and DX_GT30 in the public use data file) that have a nonresponse percentage above 5%.

Section 8.1.1 Variable Missingness – Among All Inpatient Discharges

The variables in the table below had an unweighted variable missingness percentage greater than 5%. As noted above, it is expected that procedure and diagnoses missingness for some variables would be greater than 5% given that not all encounters are expected to include more than one or multiple procedures or diagnoses listed. Regardless, these are all included below to assist data users in using the NHCS public use data files. The denominator for the percent missing is all visits on the inpatient public use data file.

Table 8.1 Percent missing (unweighted) for variables in the inpatient public use date file with a missingness greater than 5%

Variable Name	Variable Description	Percent Missing
PD1	Procedure #1 (ICD-10-PCS), procedure code	36.8
PD2	Procedure #2 (ICD-10-PCS), procedure code	59.9
PD3	Procedure #3 (ICD-10-PCS), procedure code	75.5
PD4	Procedure #4 (ICD-10-PCS), procedure code	85.3
PD5	Procedure #5 (ICD-10-PCS), procedure code	91.2
PD6	Procedure #6 (ICD-10-PCS), procedure code	94.6
PD7	Procedure #7 (ICD-10-PCS), procedure code	96.7
PD8	Procedure #8 (ICD-10-PCS), procedure code	97.9
PD9	Procedure #9 (ICD-10-PCS), procedure code	98.7
PD10	Procedure # 10 (ICD-10-PCS), procedure code	99.2
PD_GT10	Procedure Greater than 10 flag; associated with encounters that have 11 or more procedures	36.8
DX3	Diagnosis #3 (ICD-10-CM), diagnosis code	8.92
DX4	Diagnosis #4 (ICD-10-CM), diagnosis code	13.0
DX5	Diagnosis #5 (ICD-10-CM), diagnosis code	17.3
DX6	Diagnosis #6 (ICD-10-CM), diagnosis code	21.6
DX7	Diagnosis #7 (ICD-10-CM), diagnosis code	26.0
DX8	Diagnosis #8 (ICD-10-CM), diagnosis code	30.6
DX9	Diagnosis #9 (ICD-10-CM), diagnosis code	35.2
DX10	Diagnosis #10 (ICD-10-CM), diagnosis code	41.1
DX11	Diagnosis #11 (ICD-10-CM), diagnosis code	56.3
DX12	Diagnosis #12 (ICD-10-CM), diagnosis code	67.8
DX13	Diagnosis #13 (ICD-10-CM), diagnosis code	70.2
DX14	Diagnosis #14 (ICD-10-CM), diagnosis code	72.5
DX15	Diagnosis #15 (ICD-10-CM), diagnosis code	74.8
DX16	Diagnosis #16 (ICD-10-CM), diagnosis code	76.8
DX17	Diagnosis #17 (ICD-10-CM), diagnosis code	78.8
DX18	Diagnosis #18 (ICD-10-CM), diagnosis code	80.8
DX19	Diagnosis #19 (ICD-10-CM), diagnosis code	82.6
DX20	Diagnosis #20 (ICD-10-CM), diagnosis code	84.4
DX21	Diagnosis #21 (ICD-10-CM), diagnosis code	85.9
DX22	Diagnosis #22 ICD-10-CM), diagnosis code	87.4
DX23	Diagnosis #23 (ICD-10-CM), diagnosis code	88.7
DX24	Diagnosis #24 (ICD-10-CM), diagnosis code	90.0
DX25	Diagnosis #25 (ICD-10-CM), diagnosis code	91.4
DX26	Diagnosis #26 (ICD-10-CM), diagnosis code	92.5
DX27	Diagnosis #27 (ICD-10-CM), diagnosis code	93.4
DX28	Diagnosis #28 (ICD-10-CM), diagnosis code	94.1
DX29	Diagnosis #29 (ICD-10-CM), diagnosis code	94.8

Variable Name	Variable Description	Percent Missing
DX30	Diagnosis #30 (ICD-10-CM), diagnosis code	95.4

Section 8.1.2 Variable Missingness – Among All ED Visits

The variables in the table below had an unweighted variable missingness percentage greater than 5%.

The denominator for the percent missing is all visits on the ED public use data file.

Table 8.2 Percent missing (unweighted) for variables in the emergency department public use date file with a missingness greater than 5%

Variable Name	Variable Description	Percent Missing
DX2	Diagnosis #2 (ICD-10-CM), diagnosis code	20.9
DX3	Diagnosis #3 (ICD-10-CM), diagnosis code	39.5
DX4	Diagnosis #4 (ICD-10-CM), diagnosis code	53.4
DX5	Diagnosis #5 (ICD-10-CM), diagnosis code	63.8
DX6	Diagnosis #6 (ICD-10-CM), diagnosis code	71.0
DX7	Diagnosis #7 (ICD-10-CM), diagnosis code	76.3
DX8	Diagnosis #8 (ICD-10-CM), diagnosis code	80.3
DX9	Diagnosis #9 (ICD-10-CM), diagnosis code	83.6
DX10	Diagnosis #10 (ICD-10-CM), diagnosis code	86.8
DX11	Diagnosis #11 (ICD-10-CM), diagnosis code	93.8
DX12	Diagnosis #12 (ICD-10-CM), diagnosis code	97.2
DX13	Diagnosis #13 (ICD-10-CM), diagnosis code	97.7
DX14	Diagnosis #14 (ICD-10-CM), diagnosis code	98.1
DX15	Diagnosis #15 (ICD-10-CM), diagnosis code	98.4
DX16	Diagnosis #16 (ICD-10-CM), diagnosis code	98.7
DX17	Diagnosis #17 (ICD-10-CM), diagnosis code	98.9
DX18	Diagnosis #18 (ICD-10-CM), diagnosis code	99.1
DX19	Diagnosis #19 (ICD-10-CM), diagnosis code	99.3
DX20	Diagnosis #20 (ICD-10-CM), diagnosis code	99.4
DX21	Diagnosis #21 (ICD-10-CM), diagnosis code	99.6
DX22	Diagnosis #22 (ICD-10-CM), diagnosis code	99.6
DX23	Diagnosis #23 (ICD-10-CM), diagnosis code	99.7
DX24	Diagnosis #24 (ICD-10-CM), diagnosis code	99.8
DX25	Diagnosis #25 (ICD-10-CM), diagnosis code	99.8
DX26	Diagnosis #26 (ICD-10-CM), diagnosis code	99.9
DX27	Diagnosis #27 (ICD-10-CM), diagnosis code	99.9
DX28	Diagnosis #28 (ICD-10-CM), diagnosis code	99.9
DX29	Diagnosis #29 (ICD-10-CM), diagnosis code	99.9
DX30	Diagnosis #30 (ICD-10-CM), diagnosis code	99.9

Section 9 Data Comparison

Section 9.1 Public Use Data Files and Restricted Use Data File

Of the 247 participating hospitals that were included in the 2021 NHCS restricted use data file, 228 (92.3%) were selected to create the public use data files sample. The 2021 inpatient public use data file contains 179,026 inpatient discharges, for a weighted total of 35,613,997 inpatient discharges (standard error of 1,193,378 inpatient discharges). The 2021 ED public use data file contains 482,126 ED visits, for a weighted total of 118,327,593 ED visits (standard error of 2,459,558 ED visits).

The 2021 NHCS restricted use data file contains unweighted data from 197 hospitals that submitted 3,672,491 inpatient discharges, for a weighted total of 35,567,888 inpatient discharges (standard error of 1,171,087 inpatient discharges). The 2021 NHCS restricted use data file contains data from 236 hospitals that submitted 9,980,942 ED visits, for a weighted total of 118,329,910 ED visits (standard error of 2,463,300 ED visits). A comparison of weighted frequencies for inpatient discharges and ED visits in the public use data files and restricted use data file is presented in Table 9.1 and Table 9.2.

Table 9.1. Comparison of frequencies for inpatient discharges on the public use data file (weighted n=35,613,997) and restricted use data file (weighted n=35,567,888), National Hospital Care Survey, 2021

Variable	Public Use Data File				Restricted Use Data File			
		Weighted				Weighted		
	Unweighted Count	Count	Std. Error	%	Unweighted Count	Count	Std. Error	%
Age (in years)								
Under 1	19,747	3,815,479	259,638	10.7	401,525	3,866,344	257,139	10.9
1-17	8,250	2,046,385	390,134	5.7	157,992	2,000,804	391,404	5.6
18-44	45,815	8,865,600	481,503	24.9	916,692	8,892,783	488,395	25.0
45-64	42,202	8,098,818	314,837	22.7	843,965	8,061,270	321,153	22.7
65-74	29,437	5,639,300	182,466	15.8	588,655	5,638,060	182,685	15.9
75 and over	33,507	7,140,740	277,797	20.1	670,334	7,100,840	230,576	20.0
Missing	68	7,676	6,415	~0.0	1,372	7,788	6,491	~0.0
Sex								
Male	80,646	15,818,011	487,869	44.4	1,616,351	15,824,167	472,496	44.5
Female	98,125	19,756,035	736,416	55.5	1,962,755	19,734,541	712,200	55.5
Missing	255	39,951	18,305	0.1	1,429	9,179	6,431	~0.0
Discharge status								
Routine to Home	115,709	23,808,058	1,091,249	66.9	2,315,694	23,852,456	1,073,536	67.1
Left against medical advice	2,498	524,729	65,431	1.5	51,910	531,791	40,056	1.5
Transfer to short term facility	1,642	647,180	58,596	1.8	32,987	676,039	25,108	1.9
Transfer to long term facility	5,495	891,130	62,331	2.5	110,659	918,381	56,889	2.6
Home health care	23,032	4,418,805	385,159	12.4	463,409	4,422,191	355,762	12.4
Hospice care - home or medical facility	3,202	824,626	162,221	2.3	60,432	682,744	66,645	1.9
Other	18,698	3,349,657	131,163	9.4	370,057	3,318,129	108,982	9.3
Dead	4,727	923,158	45,768	2.6	94,703	931,537	37,941	2.6
Missing	4,023	226,655	59,421	0.6	80,684	234,621	58,703	0.7
Discharge month								
January	14,252	2,867,486	141,888	8.1	285,698	2,912,880	112,765	8.2
February	13,495	2,667,729	113,763	7.5	269,651	2,654,992	103,784	7.5
March	14,945	3,197,857	285,676	9.0	297,830	3,003,303	126,671	8.4
April	14,944	2,980,178	151,709	8.4	298,595	3,013,505	119,676	8.5
May	15,377	3,037,529	195,875	8.5	307,064	3,052,455	128,490	8.6
June	15,158	2,854,636	125,511	8.0	304,173	2,967,132	104,981	8.3
July	15,641	3,113,648	131,672	8.7	312,514	3,079,661	105,708	8.7
August	15,601	3,205,817	139,637	9.0	312,699	3,174,802	128,980	8.9
September	15,109	3,104,419	130,363	8.7	300,044	3,076,588	115,854	8.6
October	15,070	2,895,693	138,238	8.1	304,538	2,974,355	125,052	8.4
November	14,537	2,809,394	193,464	7.9	290,786	2,756,075	135,304	7.7
December	14,897	2,879,609	154,405	8.1	296,943	2,902,140	154,279	8.2

Table 9.2. Comparison of frequencies for ED visits on the public use data file (weighted n=118,327,593) and restricted use data file (weighted n=118,329,910), National Hospital Care Survey, 2021

Variable	Public Use Data File				Restricted Use Data File			
		Weighted				Weighted		
	Unweighted Count	Count	Std. Error	%	Unweighted Count	Count	Std. Error	%
Age (in years)								
Under 1	9,471	1,884,416	122,996	1.6	189,386	1,882,355	119,816	1.6
1-17	67,006	15,832,945	581,542	13.3	1,345,420	15,930,116	580,123	13.5
18-44	188,717	49,304,224	1,183,682	41.7	3,769,021	49,184,254	1,182,163	41.6
45-64	117,419	28,564,480	654,581	24.1	2,348,352	28,572,131	654,098	24.1
65-74	48,000	11,190,632	286,288	9.5	959,944	11,201,126	282,800	9.5
75 and over	51,035	11,512,598	368,619	9.7	1,020,763	11,524,737	374,830	9.7
Missing	478	38,299	38,048	~0.0	9,471	35,191	36,606	~0.0
Sex								
Male	221,739	52,707,990	1,029,662	44.5	4,434,921	52,759,810	1,029,151	44.6
Female	259,880	65,573,724	1,459,667	55.4	5,197,752	65,529,298	1,462,487	55.4
Missing	507	45,880	36,301	~0.0	9,684	40,802	36,624	~0.0
Discharge status								
Routine to Home	409,968	106,795,032	2,306,754	90.3	8,199,592	106,774,296	2,302,554	90.2
Left against medical advice	14,828	3,257,639	149,519	2.8	296,188	3,267,289	151,085	2.8
Transfer to short term facility	6,375	2,094,938	91,909	1.8	127,294	2,092,140	93,087	1.8
Transfer to long term facility	2,545	293,476	21,130	0.2	51,539	298,798	21,050	0.3
Home health care	7,740	1,338,414	147,483	1.1	154,774	1,343,608	143,030	1.1
Hospice care - home or medical facility	1,396	184,099	25,421	0.2	28,085	191,596	24,811	0.2
Other	28,421	3,358,491	222,403	2.8	568,117	3,358,760	223,638	2.8
Dead	2,465	309,810	11,736	0.3	48,955	308,930	11,030	0.3
Missing	8,388	695,696	365,193	0.6	167,813	694,493	365,167	0.6
Discharge month								
January	35,322	8,571,230	272,274	7.2	705,982	8,564,162	273,118	7.2
February	31,836	7,562,716	243,812	6.4	637,955	7,606,503	232,818	6.4
March	36,407	8,939,606	337,004	7.6	729,332	8,951,241	335,417	7.6
April	38,332	9,286,132	334,641	7.8	764,717	9,336,371	329,079	7.9
May	41,106	10,113,007	281,043	8.5	823,426	10,126,766	279,389	8.6
June	41,634	10,263,261	237,588	8.7	832,268	10,173,476	236,458	8.6
July	44,228	10,942,189	335,051	9.2	882,400	10,907,625	322,088	9.2
August	44,420	11,084,287	355,004	9.4	889,257	11,090,053	355,390	9.4
September	42,034	10,576,891	339,531	8.9	841,243	10,607,375	328,916	9.0
October	42,151	10,421,523	329,816	8.8	841,473	10,411,538	328,160	8.8
November	40,571	9,956,285	318,833	8.4	812,561	9,972,540	316,080	8.4
December	44,085	10,610,466	427,513	9.0	881,743	10,582,260	427,645	8.9

Section 9.2 Public Use Data Files, Restricted Use Data File, and HCUP Nationwide Samples

As stated in Section 2.5, the HCUP nationwide samples, the NIS and the NEDS were used as construct controls for calibrating the 2021 encounter-level weights for each setting. NIS and NEDS collect hospital discharge data from participating state partners. These data sources collect patient demographics (e.g., sex, age, race, etc.), encounter characteristics (e.g., diagnoses, procedures, DRGs, admission types, etc.), and hospital-level characteristics (e.g., hospital size and region).

Statewide data organizations voluntarily participate in NEDS by providing HCUP with data from ED visits that may or may not have resulted in a hospital admission. It is estimated that the 2021 NEDS accounts for 84% of the U.S. population and 83.4% of ED visits reported in the American Hospital Association Annual Survey Database. Records are selected for NEDS from two existing HCUP databases: the State Emergency Department Databases (SEDD) and the State Inpatient Databases (SID).

NIS is a sample of all-payer discharges from U.S. community hospitals, defined as "all non-Federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions." It's estimated that the 2021 NIS covers 97% of the U.S. population and 96% of discharges from U.S. community hospitals. NIS is sampled from the SID.

A comparison of weighted total inpatient discharges on the 2021 NHCS public use data file, NHCS restricted use data file, and HCUP NIS is in Table 9.3. A comparison of weighted totals for ED visits in the public use data file, restricted use data file, and NEDS is presented in Table 9.4.

Table 9.3 Comparison of weighted total inpatient discharges on the 2021 NHCS public use data file, NHCS restricted use data file, and the HCUP NIS

Data File	Inpatient Discharges	Standard Error
National Hospital Care Survey (NHCS) public use data file	35,613,997	1,193,378
National Hospital Care Survey (NHCS) restricted use data file	35,567,888	1,171,087
Healthcare Cost and Utilization Project National Inpatient Sample (HCUP NIS)	36,093,550	574,743

Table 9.4. Comparison of weighted total emergency department visits on the 2021 NHCS public use data file, NHCS restricted use data file, and HCUP NEDS

Data File	Emergency Department Visits	Standard Error
National Hospital Care Survey (NHCS) public use data file	118,327,593	2,459,558
National Hospital Care Survey (NHCS) restricted use data file	118,329,910	2,463,300
Healthcare Cost and Utilization Project Nationwide Emergency Department Sample (HCUP NEDS)	126,968,321	2,598,966

Section 10 Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA) Document for the 2021 National Hospital Care Survey Public Use Data File

Table 10.1 below provides a Preferred Reporting Items for Complex Survey Analysis (PRICSSA) document (Seidenberg, Moser, & West 2023) for users of the 2021 NHCS public use data. This information may be helpful to users when analyzing the 2021 NHCS.

Table 10.1 Preferred Reporting Items for Complex Sample Survey Analysis

Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA)	Description
Name of survey	National Hospital Care Survey
Data collection mode	Administrative data collection of hospital UB-04 administrative claims or EHR data
Target population	Hospital inpatient discharges and ED visits to nonfederal noninstitutional hospitals with six or more staffed inpatient beds in the United States
Populations excluded	Hospital inpatient discharges and ED visits to federal hospitals, institutional hospitals, or hospitals with less than six staffed inpatient beds
Sample design	Stratified random sample
Variance and standard error estimation	Jackknife variance estimation
Base Weight	PUF_ENCWGT_BASE
Replicate Weights	Inpatient: PUF_ENCWGT_1 – PUF_ENCWGT130 ED: PUF_ENCWGT_1 – PUF_ENCWGT100
Presentation standards	Proportions or percentages: https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf Rates and counts: https://www.cdc.gov/nchs/data/series/sr_02/sr02-200.pdf
Unweighted total sample size	179,026 inpatient discharges 482,126 ED visits
Weighted total sample size	35,613,997inpatient discharges 118,327,593ED visits
Response rate (unweighted)	40.6%
Location of example code	See **Section 7**

Section 11 Research Data Center

NCHS operates the Research Data Center (RDC) to allow researchers access to restricted-use data. The RDC is responsible for protecting the confidentiality of survey respondents, study subjects, and institutions while providing access to restricted-use data for statistical purposes. The 2021 NHCS restricted use data file, which contains inpatient and ED data excluded from the public use data file due to disclosure risks and NCHS confidentiality requirements, can be accessed through the Federal and NCHS RDC. In addition, the 2021 NHCS restricted use data file will be linked to other vital and administrative records such as the National Death Index (NDI). This allows the linked data to expand the analytic utility of NHCS data and provides the opportunity to conduct a vast array of studies focused on the associations between a variety of health factors, health care utilization and mortality.

For information on how to access the 2021 NHCS restricted use data file through the RDC, please see: https://www.cdc.gov/rdc/data/b1/NHCS-RDC-Data-Dictionary.pdf

Section 12 References

Agency for Healthcare Research and Quality (AHRQ). Clinical Classifications Software (CCS) for ICD-10-CM/PCS. Healthcare Cost and Utilization Project (HCUP). Available at: https://hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp. Accessed October 11, 2024.

American College of Emergency Physicians (ACEP). Available at: https://www.acep.org. Accessed October 11, 2024.

Health Level Seven International. HL7 CDA® R2 Implementation Guide: National Health Care Surveys (NHCS), R1 STU Release 3.1 - US Realm. Available at: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=385. Accessed October 11, 2024.

Lumley T (2023). "survey: analysis of complex survey samples." R package version 4.2. Available at: https://cran.r-project.org/web/packages/survey/index.html.

National Center for Health Statistics. Division of Health Care Statistics. National Center for Health Statistics Research Data Center National Hospital Care Survey 2019,2020 and 2021 Restricted Use File Data Dictionary. Hyattsville, Maryland. Available at: https://www.cdc.gov/rdc/data/b1/NHCS-RDC-Data-Dictionary.pdf

National Uniform Billing Committee (NUBC). Available at: https://www.nubc.org. Accessed October 11, 2024.

Parker JD, Talih M, Irimata KE, Zhang G, Branum AM, Davis D, et al. National Center for Health Statistics data presentation standards for rates and counts. National Center for Health Statistics. Vital Health Stat 2(200). 2023. DOI: https://dx.doi.org/10.15620/cdc:124368.

Parker JD, Talih M, Malec DJ, et al. National Center for Health Statistics data presentation standards for proportions. National Center for Health Statistics. Vital Health Stat 2(175). 2017. Available at: https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf.

Seidenberg AB, Moser RP, West BT. Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA). *Journal of Survey Statistics and Methodology* 2023;11(4):743-757.

StataCorp. 2023. Stata 18 Survey Data Reference Manual. College Station, TX: Stata Press. Available at: https://www.stata.com/manuals/svysvyset.pdf.

Vizient Inc. Available at: https://www.vizientinc.com. Accessed October 11, 2024.

Appendix A Unweighted frequencies for Inpatient discharges and ED visits

Appendix Table A.1. Unweighted frequencies for inpatient discharges on the public use data file, National Hospital Care Survey, 2021 (n=179,026)

Variable	Description	Count	%
Year	Survey year		
2021		179,026	100.0
Sex	Patient sex		
1	Male	80,646	45.1
2	Female	98,125	54.8
-9	Missing	255	0.1
Newborn	Newborn flag		
1	Yes	18,736	10.5
2	No	160,290	89.5
LOS_30DAYS	Length of stay in hospital (days) greater than 30 days		
1	Yes	3,280	1.8
2	No	175,745	98.2
-9	Missing	1	~0.01
Discharge_month	Discharge month		
1	January	14,252	8.0
2	February	13,495	7.5
3	March	14,945	8.4
4	April	14,944	8.4
5	May	15,377	8.6
6	June	15,158	8.5
7	July	15,641	8.7
8	August	15,601	8.7
9	September	15,109	8.4
10	October	15,070	8.4
11	November	14,537	8.1
12	December	14,897	8.3
Discharge_status	Discharge status	,	
1	Routine to Home	115,709	64.6
2	Left against medical advice	2,498	1.4
3	Transfer to short term facility	1,642	0.9
4	Transfer to long term facility	5,495	3.1
5	Home health care	23,032	12.9
6	Hospice care - home or medical facility	3,202	1.8
7	Other	18,698	10.4
8	Dead	4,727	2.6
-9	Missing	4,023	2.3
PD_GT10	Greater than 10 procedures associated with the encounters		
1	Yes	1,079	0.6
2	No	112,096	62.6
-9	Missing	65,851	36.8
DX_GT30	Encounters that had more than 30 diagnoses		
1	Yes	7,282	4.1
2	No	166,579	93.1
-9	Missing	5,165	2.9

Variable	Description	Count	%
CCSR_CIR007	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR007 [Essential hypertension]		
1	Yes	42,128	23.5
2	No	131,733	73.6
-9	Missing	5,165	2.9
CCSR_CIR008	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR008 [Hypertension with complications and secondary hypertension]		
1	Yes	37,222	20.8
2	No	136,639	76.3
-9	Missing	5,165	2.9
CCSR_CIR011	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR011 [Coronary atherosclerosis and other heart disease]		
1	Yes	26,992	15.1
2	No	146,869	82.0
-9	Missing	5,165	2.9
CCSR_CIR017	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR017 [Cardiac dysrhythmias]		
1	Yes	25,526	14.3
2	No	148,335	82.9
-9	Missing	5,165	2.9
CCSR_CIR019	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR019 [Heart failure]		
1	Yes	27,138	15.2
2	No	146,723	82.0
-9	Missing	5,165	2.9
CCSR_DIG004	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR DIG004 [Esophageal disorders]		
1	Yes	26,220	14.7
2	No	147,641	82.5
-9	Missing	5,165	2.9
CCSR_END003	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR END003 [Diabetes mellitus with complication]		
1	Yes	24,744	13.8
2	No	149,117	83.3
-9	Missing	5,165	2.9
CCSR_END005	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR END005 [Diabetes mellitus, Type 2]		
1	Yes	33,819	18.9
2	No	140,042	78.2
-9	Missing	5,165	2.9
CCSR_END009	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR END009 [Obesity]		
1	Yes	28,123	15.7
2	No	145,738	81.4
-9	Missing	5,165	2.9
CCSR_END010	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR END010 [Disorders of lipid metabolism]		
1	Yes	45,303	25.3
2	No	128,558	71.8
-9	Missing	5,165	2.9

Variable	Description	Count	%
CCSR_END011	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR END011 [Fluid and electrolyte disorders]		
1	Yes	45,490	25.4
2	No	128,371	71.7
-9	Missing	5,165	2.9
CCSR_FAC009	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR FAC009 [Implant, device or graft related encounter]		
1	Yes	24,009	13.4
2	No	149,852	83.7
-9	Missing	5,165	2.9
CCSR_FAC016	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR FAC016 [Exposure, encounters, screening or contact with infectious disease]		
1	Yes	83,030	46.4
2	No	90,831	50.7
-9	Missing	5,165	2.9
CCSR_FAC021	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR FAC021 [Personal/family history of disease]		
1	Yes	60,365	33.7
2	No	113,496	63.4
-9	Missing	5,165	2.9
CCSR_FAC025	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR FAC025 [Other specified status]		
1	Yes	75,205	42.0
2	No	98,656	55.1
-9	Missing	5,165	2.9
CCSR_GEN002	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR GEN002 [Acute and unspecified renal failure]		
1	Yes	26,889	15.0
2	No	146,972	82.1
-9	Missing	5,165	2.9
CCSR_GEN003	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR GEN003 [chronic kidney disease]		
1	Yes	27,031	15.1
2	No	146,830	82.0
-9	Missing	5,165	2.9
CCSR_INF003	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR INF003 [Bacterial infections]		
1	Yes	23,274	13.0
2	No	150,587	84.1
-9	Missing	5,165	2.9
CCSR_PRG023	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR PRG023 [Complications specified during childbirth]		
1	Yes	14,462	8.1
2	No	159,399	89.0
-9	Missing	5,165	2.9
CCSR_RSP012	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR RSP012 [Respiratory failure; insufficiency; arrest]		
1	Yes	23,754	13.3
2	No	150,107	83.9

Variable	Description	Count	%
-9	Missing	5,165	2.9

Appendix Table A.2. Unweighted frequencies for ED visits on the public use data file, National Hospital Care Survey, 2021 (n=482,126)

Variable	Description	Count	%
Year	Survey year		
2021		482,126	100.0
Sex	Patient sex		
1	Male	221,739	46.0
2	Female	259,880	53.9
-9	Missing	507	0.1
Newborn	Newborn flag		
1	Yes	1,179	0.2
2	No	480,947	99.8
Discharge_month	Discharge month		
1	January	35,322	7.3
2	February	31,836	6.6
3	March	36,407	7.6
4	April	38,332	8.0
5	May	41,106	8.5
6	June	41,634	8.6
7	July	44,228	9.2
8	August	44,420	9.2
9	September	42,034	8.7
10	October	42,151	8.7
11	November	40,571	8.4
12	December	44,085	9.1
Discharge_status	Discharge status		
1	Routine to Home	409,968	85.0
2	Left against medical advice	14,828	3.1
3	Transfer to short term facility	6,375	1.3
4	Transfer to long term facility	2,545	0.5
5	Home health care	7,740	1.6
6	Hospice care - home or medical facility	1,396	0.3
7	Other	28,421	5.9
8	Dead	2,465	0.5
-9	Missing	8,388	1.7
DX_GT30	Encounters that had more than 30 diagnoses		
1	Yes	137	0.03
2	No	474,106	98.3
-9	Missing	7,883	1.6
CCSR_CIR007	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR007 [Essential hypertension]		
1	Yes	71,454	14.8
2	No	402,789	83.5
-9	Missing	7,883	1.6
CCSR_CIR008	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR008 [Hypertension with complications and secondary hypertension]		
1	Yes	26,288	5.5
2	No	447,955	92.9
-9	Missing	7,883	1.6
CCSR_CIR011	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR011 [Coronary atherosclerosis and other heart disease]		
1	Yes	20,542	4.3

Variable	Description	Count	%
2	No	453,701	94.1
-9	Missing	7,883	1.6
CCSR_CIR012	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR012 [Nonspecific chest pain]		
1	Yes	34,788	7.2
2	No	439,455	91.2
-9	Missing	7,883	1.6
CCSR_CIR019	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR CIR019 [Heart failure]		
1	Yes	19,320	4.0
2	No	454,923	94.4
-9	Missing	7,883	1.6
CCSR_END005	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR END005 [Diabetes mellitus, Type 2]		
1	Yes	41,513	8.6
2	No	432,730	89.8
-9	Missing	7,883	1.6
CCSR_END010	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR END010 [Disorders of lipid metabolism]		
1	Yes	41,101	8.5
2	No	433,142	89.8
-9	Missing	7,883	1.6
CCSR_END011	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR END011 [Fluid and electrolyte disorders]		
1	Yes	29,975	6.2
2	No	444,268	92.2
-9	Missing	7,883	1.6
CCSR_FAC016	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR FAC016 [Exposure, encounters, screening or contact with infectious disease]		
1	Yes	91,928	19.1
2	No	382,315	79.3
-9	Missing	7,883	1.6
CCSR_FAC021	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR FAC021 [Personal/family history of disease]		
1	Yes	73,118	15.2
2	No	401,125	83.2
-9	Missing	7,883	1.6
CCSR_FAC025	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR FAC025 [Other specified status]		
1	Yes	112,344	23.3
2	No	361,899	75.1
-9	Missing	7,883	1.6
CCSR_GEN003	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR GEN003 [chronic kidney disease]		
1	Yes	19,670	4.1
2	No	454,573	94.3
-9	Missing	7,883	1.6
CCSR_INF012	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR INF012 [Coronavirus disease – 2019 (COVID-19)]		
1	Yes	22,243	4.6
2	No	452,000	93.8

Variable	Description	Count	%
-9	Missing	7,883	1.6
CCSR_INJ017	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR INJ017 [Superficial injury; contusion, initial encounter]		
1	Yes	21,559	4.5
2	No	452,684	93.9
-9	Missing	7,883	1.6
CCSR_MBD024	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR MBD024 [Tobacco-related disorders]		
1	Yes	49,008	10.2
2	No	425,235	88.2
-9	Missing	7,883	1.6
CCSR_MUS010	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR MUS010 [Musculoskeletal pain, not low back pain]		
1	Yes	39,589	8.2
2	No	434,654	90.2
-9	Missing	7,883	1.6
CCSR_SYM004	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR SYM004 [Nausea and vomiting]		
1	Yes	31,399	6.5
2	No	442,844	91.9
-9	Missing	7,883	1.6
CCSR_SYM006	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR SYM006 [Abdominal pain and other digestive/abdomen signs and symptoms]		
1	Yes	54,521	11.3
2	No	419,722	87.1
-9	Missing	7,883	1.6
CCSR_SYM010	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR SYM010 [Nervous system signs and symptoms]		
1	Yes	31,114	6.5
2	No	443,129	91.9
-9	Missing	7,883	1.6
CCSR_SYM013	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR SYM013 [Respiratory signs and symptoms]		
1	Yes	41,474	8.6
2	No	432,769	89.8
-9	Missing	7,883	1.6