National Hospital Care Survey

Research Data Center

Documentation

Last Updated: 1/28/19

Table of Contents

Abstract

The purpose of this document is to provide an overview of the National Hospital Care Survey (NHCS), the files available through the Research Data Center (RDC), and associated analytic considerations. NHCS is a national probability sample survey of hospitals that collects hospital encounter data. The survey is conducted by the National Center for Health Statistics (NCHS). Data are available in the RDC for calendar years 2013, 2014, 2015, and 2016. The descriptions and analytic considerations documented in the RDC documentation apply to all data years unless otherwise noted.

Data users should develop a proposal, including a list of needed variables. After a proposal is approved, a file will be developed for the analyst's use. For more information on how to develop a proposal, please see: https://www.cdc.gov/rdc/b3prosal/pp300.htm.

For information on the available data variables for each year, please see: https://www.cdc.gov/rdc/b1datatype/dt1224h.htm

The National Hospital Care Survey Background

The National Hospital Care Survey (NHCS) provides data on patterns of health care in hospital-based settings and on hospital encounters in Inpatient (IP), Emergency (ED), and Outpatient departments (OP).

The universe for NHCS consists of hospitals with six or more beds staffed for inpatient care in the 50 states and the District of Columbia, which are non-federal and non-institutional. NHCS started data collection in 2011, integrating two long-standing National Center for Health Statistics surveys:

- National Hospital Discharge Survey - the longest continuously fielded sample of inpatient care from 1965-2010; and
- National Hospital Ambulatory Medical Care Survey - surveying hospital EDs and OPDs since 1992, hospital Ambulatory Surgery Locations since 2009, and freestanding Ambulatory Surgery Centers since 2010.

In 2011, NHCS started with only the collection of sampled hospital's UB-04 claims records. The claims data are strictly defined by National Uniform Billing Committee and collect up to 25 diagnosis and procedures codes. Starting in 2015, NHCS transitioned into collecting additional sources of data such as Electronic Health Records (EHR) and Vizient. EHR data contain unlimited diagnosis and procedure information and without the billing motivated manipulation. Another benefit of EHR data is that the data contain the clinical notes recorded by the clinician during the hospital visit, providing information on the context of the visit and specific drug mentions. Vizient data are similar to UB-04 claims data but contain information on medications and laboratory testing results. While Vizient data collect similar information as UB-04, patient identifiers are not included, and the dates of the encounter are masked.

The 2013-2016 NHCS data available in the RDC are based on a sample of 581 hospitals drawn from the universe of eligible hospitals. However, not all 581 hospitals provided data. Response rates are not high enough to produce weighted national estimates. Data are also available for inpatient and ambulatory settings. However, in a given year, a hospital may not have provided data for all three settings. Ambulatory data are divided into ED and OPD. Definitions of the settings are below.

- Inpatient: Hospital department where a patient can stay overnight in the hospital while under treatment. Hospitalizations in the inpatient department are referred to as discharges.
- Ambulatory: Ambulatory departments are hospital departments where a patient is not able to stay overnight while under treatment. Ambulatory includes both outpatient and emergency departments. OPDs are hospital departments designed for patients that receive treatment but are not admitted for overnight care. EDs provide unscheduled outpatient services to patients whose conditions require immediate care, and are staffed 24 hours a day. On- and off-site EDs that are open fewer than 24 hours are included if staffed by the hospital. Patients presenting in the ED may subsequently be admitted as inpatients.

For an example of how to use the non-weighted data for healthcare research refer to the NCHS publications National Hospital Care Survey Demonstration Projects: Traumatic Brain Injury (available here: https://www.cdc.gov/nchs/data/nhsr/nhsr097.pdf) and National Hospital Care Survey Demonstration Projects: Pneumonia Inpatient Hospitalizations and Emergency Department Visits (available here: https://www.cdc.gov/nchs/data/nhsr/nhsr116.pdf).

Hospital Frame and Sample Design

The initial frame for the NHCS was constructed in 2011 from the 2010 IMS Government Solutions (previously SDI, Verispan, SMG, IMS Health) hospital database. The 2010 hospital database is a file of hospitals and facilities used for marketing. Not all facilities in the 2010 hospital database met NHCS eligibility requirements. An NHCS eligible hospital is defined as a non-federal and non-institutional hospital with at least six staffed inpatient beds. From that initial frame file of 6,622 hospitals, a sample of 1,000 hospitals was selected. The sample was selected based on a stratified design of hospital bed size, type of hospital, and urban/rural designation. The initial sample of 1,000 hospitals was split into two groups of 500 hospitals - a base sample and a reserve sample. The base sample hospitals were sent in the field for data collection. The reserve sample was selected to replenish the base sample if more hospitals were needed. The base and reserve samples were selected using the same stratified sampling methodology.

In 2013, 81 general acute hospitals with at least 500 staffed inpatient beds from the reserve sample were added to the base sample for data collection. From 2013 through 2016, 581 hospitals were eligible to participate in the survey and/or complete the NHCS Annual Hospital Interview, which is offered to all hospitals in the sample regardless of participation status

Data File Structure

File structure differs between 2013/2014 and 2015/2016. For data from 2013 and 2014, each record represents a unique encounter for a unique person and each record includes the encounter's diagnoses, procedures and reason for visit. For 2015 and 2016, the main file can include demographics, visit characteristics and revenue codes, and each record represents a unique encounter for a unique patient. However, starting in 2015, diagnoses, procedures and reason for visit are each in separate files. These 3 files are not on an encounter basis – rather, each diagnosis, procedure or reason for visit is one record. The files may then be linked on ENCOUNTER_ID to the file containing demographics and encounter characteristics.

Participating Hospitals

The number of hospitals that have provided data by hospital department over the lifespan of NHCS is presented in the table below. The table below shows the number of hospitals and patient encounters collected by year and hospital department in NHCS from 2013-2016.

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

Year	IP Hosp	IP Encounters	ED Hosp	ED Encounters	OPD Hosp	OPD Encounters	Total Hosp	Total Encounters
2013	97	1,474,478	82	3,784,397	87	15,144,448	97	20,403,323
2014	94	1,653,622	83	4,530,360	86	19,005,777	95	25,189,759
2015	114	2,204,258	97	5,900,738	101	26,455,149	118	34,560,145
2016	145	2,591,722	124	7,032,304	128	35,692,420	150	44,572,315

National Hospital Care Survey Data Elements

NHCS collects information on hospital setting for each encounter (IP, ED, and OPD), diagnoses, services, discharge status, point of origin, start and end date of the hospital encounter, and patient identifiers. There are also additional data elements captured that are only available for hospital encounters in IP or Ambulatory departments. For example, NHCS includes information on type of admission for IP encounters. Refer to the data dictionary for detailed information on the type of data elements available in the RDC including variable names, variable values, and important information on availability by data source (e.g., claims or EHR) and hospital setting.

Advantages of the National Hospital Care Survey

Due to the collection of personally identifiable information (PII), NHCS allows linkage of records within and across the data files and to outside databases. Linkage capabilities include:
- Within a hospital setting to count the number of encounters by patient;
- Across hospital settings to identify the number of encounters in each setting by patient;
- Across hospital settings to identify the number of encounters where a patient's encounter started in the ED and was admitted as an IP; and
- Linkage on a person basis to the National Death Index (NDI) (https://www.cdc.gov/nchs/ndi/index.htm) (2014 and 2016 NHCS). 2014 NHCS data are matched to the 2014 and 2015 NDI. The 2016 NHCS data are matched to the 2016 and 2017 NDI.
- Linkage on a person basis to the Centers for Medicare and Medicaid Services (CMS) Master Beneficiary File (https://www.cdc.gov/nchs/ndi/index.htm) 2014 NHCS is matched to the 2014 CMS Master Beneficiary File.

Data Sources

From 2011 to 2014, NHCS only collected UB-04 claims data from hospitals. Starting in 2015 hospitals also had the option of providing data from two different data sources: Vizient and EHR. The table below provides an overview of each of the three data sources. Vizient data are similar to claims data in structure but include separate laboratory and medications files. Vizient data have some limitations due to privacy concerns, Vizient data do not provide patient PII, nor the exact date of the hospital visit. Additionally, all Vizient hospital dates of visit are defaulted to the first of the month. EHR data are normalized to the same structure as claims and Vizient. A data element on the file identifies the data source. The separate laboratory and medications files for Vizient and EHR are not currently available in the RDC. For more information on the difference between the data sources available in 2015 and 2016, please refer to the table on the following page.

The table below contains a summary of each data source available in the RDC and known differences between the data sources. The data issues are also described in more detail further in the document. Further information about UB-04 claims can be found in the next section.

Data Source	Year	Description	Considerations
UB-04 claims	2013-2016	Data collected on the uniform bill (UB-04) for institutional providers approved by the National Uniform Billing Committee. UB-04 is the electronic standard for hospital billing. For more information on the UB-04, see: https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/837I-FormCMS-1450-ICN006926.pdf.	• Administrative data used for billing rather than clinical purposes. Thus, may exclude information not needed for billing, but clinically important. • Able to track ED encounters admitted as inpatients. • Maximum of 25 diagnosis and procedure codes provided. • Revenue code flags are available.
Vizient	2015-2016	Vizient is a member-driven health care services company. Vizient collects information from the hospitals prior to submitting it to NHCS. It is similar to UB-04 claims but includes some medications and laboratory data. For more information, please visit: https://www.vizientinc.com/.	• Administrative data used for billing rather than clinically. Thus, may exclude information not needed for billing, but clinically important. • No PII included (e.g. patient name) and thus cannot be linked to outside data sources. • The exact date of the start and end of the hospital encounter is not collected. Only the month of the end of the visit is given. • No data on length of stay. • Maximum of 25 diagnosis and procedure codes provided. • Not able to track ED encounters admitted as inpatients. • Revenue code flags are available.

Data Source	Year	Description	Considerations
EHR	2015-2016	An electronic version of a patient's medical history maintained by the hospital over time. (https://www.cms.gov/Medicare/E-Health/EHealthRecords/index.html) In 2016 there were two types of EHR data collected: Consolidated-clinical document architecture (C-CDA) and EHR Custom Extracts. C-CDA is a set of HL7 Clinical document architecture. Includes 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 the most commonly needed information about current and past health status that can be shared by all computer applications, including web browsers, electronic medical and health record software systems. In 2016, NCHS received discharge summaries, history and physical summaries, and transfer summaries. EHR Custom Extract records are extracted from the hospitals' EHR according to specifications provided by NCHS. The specifications are based on the National Health Care Surveys Implementation Guide.	• Clinical rather than billing focus. • No limit on the number of diagnoses and procedures collected. • Not able to track ED encounters admitted as inpatients. • Primary diagnosis is not available.

Data Collection and Processing Summary

Claims de-duplication

Using UB-04 claims data presented a challenge, since one encounter, both for inpatient discharges and ambulatory visits, can have multiple claims. Therefore, de-duplication processes needed to be developed. The initial de-duplication was performed at the hospital level, using processes to identify duplicate claims for the same encounter, inpatient or ED, within a hospital. In order to develop the de-duplication method for a hospital, claims were grouped in each of these three ways: (1) by Patient Control Number (PCN); (2) by Beginning Date of Encounter + Medical Record Number (MRN); and (3) by Beginning Date of Encounter + Date of Birth + Patient Name. The purpose of this processing was to evaluate whether PCN could be used to accurately identify duplicate claims for the same encounter, controlling for matching data elements for beginning date of encounter, MRN, date of birth, and patient name. If the number of duplicate groups of claims identified by PCN was close to the other counts produced, PCN was used to de-duplicate the claims. However, if the values of these variables in a group of duplicates was not similar, then the hospital's claims were further assessed through a manual review of the duplicate groups and a final determination of the de-duplication method to be used at that hospital was made.

Outpatient claims splitting:

Although the vast majority of ambulatory visits involve a single day, many cases of ambulatory claims spanning two or more dates were detected among ambulatory claims. Multiple date claims were examined, and some cases had evidence that they were justifiably multiple day visits. Three types of cases were identified as probably being valid multiple day visits to the ambulatory department: (1) an ED visit could span several days, and even a short ED visit that started in the late evening of one day could easily continue into the next day and would appear as a two-day visit; (2) a patient who receives "observation services" is periodically monitored by hospital staff to determine the need for possible admission, and this period of monitoring could involve several days; and (3) patients undergoing ambulatory surgery might be kept in the ambulatory center for longer periods of time pre- or post-surgery.

A claim for multiple dates which should not have been treated as a single visit spanning multiple days was split into two or more sub-claims. Examples of multiple date claims that should have been split were repeated treatments over a period of time, e.g., for weekly speech therapy or regular dialysis, for which some hospitals preferred to submit one claim covering a period of time. Any claim with non-consecutive dates in the ED or the OPD was split into two or more claims for single day stays (or for consecutive day stays). All consecutive day claims were then separated into one claim per day assuming there were no evidence of a justifiable longer stay – such as visits to the ED, for patients receiving observation services, or for ambulatory surgery: any consecutive day visits with evidence of any of these three conditions were not split.

ED visits admitted as inpatients

ED visits that were admitted as an inpatient did not have a separate ED record, therefore the inpatient record was duplicated in the ED file. However, the duplicated ED records maintain the inpatient discharge status. In order to get an accurate account of ED discharge status, inpatient records in the ED file had the discharge status changed to "admitted as an inpatient." Additionally, in order not to over count services provided, services on the duplicated ED records are only counted in the inpatient setting.

Patient identification

After de-duplication of hospital records took place, a probability-based record linkage method was used to identify patients. In the first round of patient identification, two records were compared by name (first, last, middle initial), date of birth, sex, hospital identifier, medical record number, Social Security Number (SSN), and ZIP code. If there was sufficient agreement of data elements between the two records, the records were retained as a pair (i.e., record pair) for further analysis. The second round compared the record pairs, controlling for agreement in the hospital identifier and MRN (Pass 1), SSN if reported (Pass 2), and for agreement in sex, year and month of birth, Soundex (a phonetic coding system designed to suppress spelling variations) of last name, and state abbreviation (Pass 3). The record pairs compared in the three passes had match weights assigned to 11 matching variables. The match weights were likelihood ratio scores based on the probability of agreement in the records retained as record pairs and the probability of agreement in the records that were not included in any record pairs.

Two additional reviews were conducted for records pairs of children aged less than ten years old at the time of discharge. The first review targeted newborn infants whose first names contained "BABY," "GIRL," "BOY," "FEMALE", or "MALE." The newborn pairs were then subject to one of three adjustments. (1) Pairs that

contained records where names may be "BABYGIRL" in one record and a real name, e.g., "JANE," in another were accepted as a match when the hospital identifier and MRN were the same. (2) Pairs with the same hospital identifier, service date, and patient address, but different MRNs, were identified as twin or multiple birth records and manually split. (3) Pairs with different last names and medical record numbers were manually reviewed and split if the pair was determined to be false.

Analytic considerations:

National representativeness:

All the NHCS data available in the RDC are unweighted and are not nationally representative.

Data elements:

- Diagnosis and procedure coding (2015/2016):
  - Due to required implementation of ICD-10, the 2015 files are coded in a mixture of ICD-9 and ICD-10 for both diagnoses and procedures. The first three quarters are coded in ICD-9 and the last quarter is coded in ICD-10. There is a data element on the file identifying which system was used. Further information on implementation of ICD-10 is available at: https://www.cms.gov/Medicare/Coding/ICD10/. Further information on cross-walking ICD-9 and ICD-10 codes is available at: https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.html.
  - In the 2016 files, the majority of the diagnosis codes are in ICD-10-CM. Hospitals also submitted diagnosis codes in ICD-9-CM, SNOMED-CT, FormFast, Progetto Nazional Emorragia Digestiva (PNED), and systems custom to the hospital. When possible the ICD-9-CM and the SNOMED-CT codes were translated to ICD-10-CM. The translated ICD-10-CM codes are available in the RDC.
  - Primary diagnosis cannot be identified in the hospitals that provided EHR data in 2015 and 2016.
  - In the 2016 data, EHR hospitals are missing procedure and diagnosis information for some encounters.
- Ambulatory procedures: These are coded in Current Procedural Terminology (CPT) for services, HCPCs for products, supplies and services and ICD-9/10 PCS. (https://www.aapc.com/resources/links/ and https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/Code-Sets/index.html)
- Revenue codes: Revenue Codes are descriptions and dollar amounts charged for hospital services provided to a patient. There are flags available in the RDC that provide information on the services identified via revenue codes. Revenue codes are only available for claims and Vizient data.
- Discharge status: Discharge status indicates the final outcome of the hospital visit. Examples of discharge status include home, dead, and transfer to another facility. In 2016, hospitals that provided EHR data are missing discharge status on a large number of hospital encounters. Additionally, hospitals that provided EHR Custom Extracts also provided multiple discharge statuses for one encounter.

Linkage

- The ability to <u>link individual patients across settings</u> is available for all data sources. (use PATIENT_ID).
- The ability to <u>link to outside databases</u>, such as the NDI is limited to records that include PII. Vizient data did not provide PII and thus cannot be linked to the NDI. To date, linkage to the NDI is available only for 2014 and 2016 data.
- Researchers can link patients across multiple years of data.

Additional Information and Contact Information

For additional questions or details on analytical considerations of NHCS data, please contact us at nhcs@cdc.gov.

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