

## Overview

- **Topic/focus:** Cost and Utilization, Healthcare Facilities and Services
- **Year(s):** 1988 - present
- **Source:** Agency for Healthcare Research and Quality
- **Study and sample characteristics:** Administrative medical records
- **Universe:** Hospitals in participating states
- **Variables:** Varies depending on dataset, see below for more information
- **Access:** Data use agreement is required
- **Cost:** Please check with the EDRC Coordinator
- **Key web links:** <http://hcupnet.ahrq.gov/>
- **Summary:** The Healthcare Cost and Utilization Project (HCUP, pronounced "H-Cup") is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ). HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of patient-level health care data (HCUP Partners). HCUP includes the largest collection of longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988. These databases enable research on a broad range of health policy issues, including cost and quality of health services, medical practice patterns, access to health care programs, and outcomes of treatments at the national, State, and local market levels. HCUP's objectives are to: 1) Create and enhance a powerful source of national, state, and all-payer health care data, 2) Produce a broad set of software tools and products to facilitate the use of HCUP and other administrative data, 3) Enrich a collaborative partnership with statewide data organizations aimed at increasing the quality and use of health care data, and 4) Conduct and translate research to inform decision making and improve health care delivery. Please see the Nationwide Inpatient Sample , the Kid's Inpatient Database , and the Nationwide Emergency Department Sample descriptions for more detail. Participating state-specific data is also available.

## HCUP Requirements When Handling Data

Some of these requirements were outlined in the HCUP training module, but must still be followed when releasing any study done using the data.

- Researchers must not attempt to identify providers or beneficiaries contained within the dataset.
- Researchers must not attempt to contact providers contained in the dataset about HCUP data.
- When releasing research finding using the HCUP data no subsection of the data will have less than ten members in it as to keep the identity of the members private.
- You must not divulge specifics about the data with user who does not have access to the HCUP data.

## Data Description

### National Emergency Department Sample (NEDS)

- **Years:** 2010-2012

- **Data Documentation:**
  - [Webpage Tool](#)
  - [PDF \(short version here\)](#)
  - [Additional Info](#)
  - [Reverse Lookup and Element Description](#)
- **Variables:**
  - Core: 90
  - ED: 60
  - Hospital: 14
  - IP: 37
- **SQL Schema: HCUP**
- **Contained Tables:**
- **Normalization plan:**
  - Core
    - CHRON1-15 can be moved into CHRON\_NEDS table
    - DX1-DX15 can be moved into DX table might be able to be combined with DXCCS1-DXCCS15
    - ECODE1-ECODE4 can be move into ECODE table
  - ED
    - CPT1-CPT15 can be normalized
  - Hospital
    - Table is fine as is. No normalization required.
  - IP
    - PCLASS, PR\_IP, and PRCCS\_IP should all be able to be normalized out.

## Nationwide Inpatient Sample (NIS)

- **Years:** 2010 - 2012
- **Data Documentation:**
  - [Webpage Tool](#)
  - [PDF](#)
  - [Additional Info](#)
  - [Reverse Lookup and Element Description](#)
- **Variables:**
  - Core: 142
  - DX\_PR\_GRP5: 69
  - Hospital Weights: 13
  - Severity: 33
- **SQL Schema: HCUP**
- **Contained Tables:**
- **Normalization plan:**
  - Core
    - DX and DXCCS should be able to be merged out into normalized table
    - PR and PRCCS should be able to be merged out into normalized table
    - PRDAY should be able to be to merged out along with PRCCS and PR
  - Hospital
    - Table is fine as is. No normalization required as of right now
  - DX\_PR\_GRPX
    - CRHON along with CHRONB should be able to be normalized out together
    - procedure class should also be able to be normalized out
  - Severity
    - Table looks to be fine as is. No normalization required as of right now.

## Normalized HCUP Tables

NIS\_NORM\_DIAG\_2010-2012

NIS\_NORM\_PRICEDURE\_2010-2012

NEDS\_NORM\_DIAG\_2010-2012

This table contains the ICD9 codes that exist in the NEDS Core files normalized to allow easier analysis of the codes that are assigned to patients. In the original code set the diagnosis codes were across 25 columns where not all 25 columns were filled. This this cause any analysis done on the dataset to be done one each of the 25 columns and is very inefficient on sql queries of the database.

NIS_CORE_2010-2012		NIS_NORM_DIAG_2010-2012	
KEY or KEY_NIS	Record primary key. Used to distinguish keys apart from each other. Years 2010 and 2011 use KEY as the name to this column and have ~12 digit record numbers while year 2012 changed this element to KEY_NIS and is only ~8 digit record numbers. Each record number should only appear one in the dataset as these elements are tied to record and not to patients.	KEY or KEY_NIS	Same as in the CORE file except as this is a normalization the record number is used to tie each user back to the original dataset the each record number appears multiple times (once per diagnosis code associated to each KEY or KEY_NIS.
HOSP or HOSP_NIS	Hospital identifier code that is used to distinguish one hospital from another. Ties into the HOSP file which has information about the different hospitals. This element is name HOSP in 2010 and 2011 and HOSP_NIS in 2012.	HOSP or HOSP_NIS	Same as the CORE file. This element is kept in order to easier facilitate hospital analysis.
DX1-DX25	25 columns in original dataset that could or could not be empty. ICD9 codes were also in an odd format by having an implicit decimal point after the third char (V4544 instead of V45.44)	DX_CODE	Single column containing all of the diagnosis codes for each user as a relational table.
		DX_ORDER	New row to keep information from original data on the order in which the DX appeared. AKA a value of 1 in this position shows that it original corresponded to DX1 and a value of 13 corresponding to DX13.
DXCCS1-	HCUP modification of the ICD9 codes into	DXCCS_CODE	Normalized into one

DXCCS25	their own code system. Use the public ICD9 database to decode these.		column each ICD9 code has a corresponding entry for the DXCCS which is why they are normalized together.
		DXCCS_ORDER	New row to keep information from original data on the order in which the DXCCS appeared. AKA a value of 1 in this position shows that it original corresponded to DXCCS1 and a value of 13 corresponding to DXCCS13. This is kept for both DXCCS and DX as a sort of fact check to make sure that the codes correspond.

## Sources

All information relating to data elements descriptions and database summaries have been pulled from the database providers website located here: <http://hcupnet.ahrq.gov/>