HCUP HOSPITAL IDENTIFIERS

Table of Contents

Table of Tables

HCUP HOSPITAL IDENTIFIERS

This document describes the different hospital identifiers included in the Healthcare Cost and Utilization Project (HCUP) databases and the process used to reconcile the hospital identification used by the HCUP Partner organizations and the American Hospital Association (AHA).

HCUP State Databases

The HCUP Partners determine which hospital identifiers, if any, may be released on the HCUP State Inpatient Databases (SID), State Emergency Department Databases (SEDD), and State Ambulatory Surgery and Services Databases (SASD). Up to three hospital identifiers are included in the HCUP databases:

- The HCUP Partners own number scheme for identifying hospitals and facilities (HCUP data element: DSHOSPID),

- The hospital identifier used by the AHA (HCUP data element: AHAID or IDNUMBER), and

- A unique HCUP hospital identifier (HCUP data element: HOSPID).

HCUP Nationwide Databases

Hospital identifiers vary depending on the nationwide database and the year:

- The National Inpatient Sample (NIS) and the Kids' Inpatient Sample (KID), beginning in 2012, include a unique HCUP hospital identifier, HOSP_NIS or HOSP_KID, respectively. These hospital identification numbers are reassigned each year. Therefore, it is not possible to link the NIS or KID to other HCUP or external databases or to track the same hospital across years. The use of these hospital pseudo-identifiers helps protect identities while preserving the analyst's ability to estimate hospital-level variation.
  - Prior to 2012, the NIS and KID contained the unique HCUP hospital identifier, HOSPID. The HOSPID could be used to link to other HCUP databases. The HOSPID could also be used to track the same hospital across years.
- The Nationwide Readmissions Database (NRD) contains HOSP_NRD for all years. This is similar to HOSP_NIS and HOSP_KID where the hospital identification numbers are reassigned each year Therefore, it is not possible to link to other HCUP or external databases or to track the same hospital across years.
- The Nationwide Ambulatory Surgery Sample (NASS) contains HOSP_NASS for all years. This is similar to HOSP_NIS and HOSP_KID where the hospital identification numbers are reassigned each year. Therefore, it is not possible to link to other HCUP or external databases or to track the same hospital across years.
- The Nationwide Emergency Department Sample (NEDS) contains HOSP_ED for all years. HOSP_ED identifies ED records that are associated with the same hospital-owned ED. It does not link to other HCUP databases or external databases. However, it can be used to identify the same hospital across years. Once a HOSP_ED value is assigned to an emergency department, the same value is retained each year.

Once the data source's identification of the hospital is reconciled with the identification of the hospital in the AHA, data from the AHA Annual Survey are used to:

- Identify facilities that are defined by the AHA as "community hospitals;"

- Identify community hospitals listed in the AHA Annual Survey for which no inpatient data were supplied by the data source for the SID.

- Classify each community hospital into a stratum for sampling for the HCUP nationwide databases;

- Add information about the hospital to the nationwide and State databases
  - For State databases that allow hospital identifiers to be released, a AHA Linkage File is available. This file contains a small number of data elements that allow researchers to link the HCUP State databases to the AHA Annual Survey (which must be purchased separately) to supplement the HCUP State databases with hospital characteristics. Users can merge the data elements of interest from the AHA Annual Survey Databases to the HCUP AHA Linkage Files using AHAID, and then subsequently link to the HCUP State databases by using DSHOSPID.
  - For the nationwide databases, limited hospital characteristics are available on the files. These include census division and region of the hospital, urban-rural designation, teaching status, bed size, control/ownership of the hospital, and trauma center level (NEDS only).
    - Prior to 2012, HOSPID could be used to link the NIS or the KID to the AHA Annual Survey to supplement with additional hospital information (such as its county FIPS code). This is not possible starting in data year 2012.

Note: Generally, AHA identifiers can be determined for community and non-community hospitals that are registered with the AHA. AHA identifiers are not available for non-hospital-owned ambulatory surgery centers.

RECONCILING HOSPITAL IDENTIFIERS

The goal is to identify an appropriate AHA hospital identifier for each source hospital in a given year.

To begin, relevant data elements are extracted from State data and from the AHA data for each year. The two elements extracted from the State data are:

- the source-specific hospital identifier (DSHOSPID), and

- a count of the hospital's inpatient records for each quarter and for the year in the State Inpatient Databases (SID).

Linkage of Source and AHA Hospital Identifiers

First, the hospital identifiers used by the data source (herein referred to as DSHOSPID) are linked to the relevant data elements extracted from the AHA Annual Survey of Hospitals data. AHA identifiers include all hospitals in the state, not just the community hospitals.

A SAS merge step is used to link the DSHOSPIDs to AHA identifiers. The specific variables used in the merge depend on the information provided by the data source. In order of preference, these variables are:

- hospital name, city, and zip code;
- hospital name; or
- any other unique variable that is available, e.g., Medicare provider number.

The AHA and the data source often use different methods to represent components of a hospital's name (e.g., "Community General Hospital" may be represented as "Community General Hosp" by the AHA and as "Community Gen Hosp" by the data source). Hence, before the SAS merge, the AHA and the source's hospital names are transformed into a uniform link variable, which imposes similar methods of abbreviations, lowercase and uppercase letters, and different characters. This effectively reduces the number of non-merges that occur simply because of different methods of representing the hospital name's components.

Three types of linkages result from this step, as shown in Table 1.

Table 1. Linkages Between AHA and Data Source Identifiers

Row	AHA Identifier	DSHOSPID	Link?
1	present	present	yes
2	present	absent	no
3	absent	present	no

Approximately 80 percent of DSHOSPIDs link to AHA identifiers in this step. (These successful links are represented by Row 1 in Table 1). The other 20 percent of DSHOSPIDs (Rows 2 and 3) must be linked manually, using the process described below.

Prior experience has shown that a large majority of the hospitals failing to link in this step will usually be due to:

- closures,
- openings (new hospitals),
- mergers,
- demergers,
- dates of changes in the hospital structure that differ between the data source and the AHA Annual Survey, and
- levels of aggregation that differ between the data source and the AHA Annual Survey (e.g., the data source treats two separate facilities as two hospitals, while the AHA Annual Survey treats the two facilities as a single hospital, or vice versa).

Resolution of Unmatched Hospitals

The goal in this step is to identify an appropriate AHA hospital identifier for each source hospital that was not matched above.

Several external sources of information are used to reconcile the unmatched AHA hospitals (Row 2) and the unmatched source hospitals (Row 3). These are:

- *Source Documentation:* This information, received from the data source, usually contains a list of hospitals, their cities (specific addresses are not always included), and the source's hospital identifier. This documentation is the primary source for finding missing information (e.g., specific names and addresses) for an unmatched hospital.

- *AHA Guide:* The *AHA Guide*, an annual hard-copy volume published by the AHA, provides a wealth of information about registered hospitals. The *AHA Guide* includes an entire section on multihospital health-care systems that identifies the hospitals included in specific multihospital systems. The *AHA Guide* also provides information about individual hospitals (organized by state, and within each state, by city), which includes:

  - Information also available from the AHA Annual Survey data; for example:
    - type of service (general medical/surgical, rehabilitation);
    - average lengths of stay (long- or short-term);
    - type of ownership; and
    - numbers of beds, admissions, births, etc.

  - Information about hospitals embedded within the organizational structure of another hospital; for example:
    - *Binghamton, Broome County, NY*
    - *United Health Services (includes Binghamton General Hospital, Mitchell Ave. ...; Wilson Memorial Hospital, ... Harrison St. ...)*

  - Information about changes in a hospital's name; for example:
    - *Dobbs Ferry, Westchester County, NY*
    - *Community Hospital at Dobbs Ferry (formerly Dobbs Ferry Hospital)*

  - References to a new hospital name or new location; for example:
    - *Delhi, Delaware County, NY*
    - *Lindsay and Olive B. O'Connor Hospital. See Mary Imogene Bassett Hospital, Cooperstown.*

- *AHA Summaries:* The AHA Summary of Registered Hospitals and the AHA Summary of Nonregistered Hospitals, which are usually delivered with the annual *AHA Guide*, document additions and deletions reflected in the hospitals' responses to the AHA Annual Survey.

- *Record counts generated from the supplied inpatient data:*
  The number of discharges reported in the inpatient data is compared to the number of discharges reported to the AHA. While this information is rarely definitive in linking source identifiers to AHA identifiers, it is sometimes useful in identifying a link to an AHA hospital, and provides a means of validating linkages obtained by other means. This information is especially useful in distinguishing a single hospital from two combined hospitals.

When it is not clear what needs to be done to a hospital or group of hospitals, an AHRQ analyst is consulted.

When the reconciliation process is complete:

- An AHA identifier (AHAID and IDNUMBER) has been assigned to all hospitals registered with the AHA,

- all hospitals with a non-missing AHA identifier have been assigned an HCUP hospital identifier (HOSPID), and

- all hospitals with a non-missing AHA identifier have only one FIPS county code assigned.

Hospitals composed of multiple facilities in different locations are assigned the FIPS county code of the major facility, as defined by the AHA.

The HCUP data element AHAID contains the hospital number used on AHA Annual Survey of Hospitals data file available through the AHA. The HCUP data element IDNUMBER contains the same identifier with the leading "6" removed.

RULES FOR RESOLVING PROBLEM HOSPITALS

Following are the rules used for resolving problem hospitals. In these examples, the HCUP hospital identifier (HOSPID) starts with "SS" to indicate the state FIPS code:

The HCUP hospital identifier (HOSPID) reflects the AHA view of a hospital and is a randomly assigned number based on the AHA hospital identifier (AHAID and IDNUMBER). If the data source reports the data from facilities that are combined in the AHA hospital definition, the IDNUMBER and the HOSPID will be the same for all the facilities. In the following example, three different source identifiers are considered to be part of one facility as defined by the AHA:

Table 2. Example of Different Source Identifiers Mapped to One AHA Identifier

Year	Data Source	AHA IDNUMBER	HOSPID
2018	165 (Acute Care Unit)	910140	SS089
2018	165P (Psychiatric Unit)	910140	SS089
2018	166S (Swing Bed Unit)	910140	SS089

Openings

The AHA AHAID/IDNUMBER and HCUP HOSPID are assigned to a newly opened hospital only when the hospital has first been recognized by the AHA for a particular survey year, even if the data source supplies data for an earlier time period. For example, the data source supplied data for a hospital starting in 2016, but the AHA first recognized the hospital in 2018:

Table 3. Example of Hospital Openings

Year	Data Source	AHA IDNUMBER	HOSPID
2016	86-0601625		
2017	86-0601625		
2018	86-0601625	860001	SS014
2019	86-0601625	860001	SS014

Closures

When a hospital closes (in the AHA's view), the AHAID/IDNUMBER and HCUP HOSPID are carried forward if there are inpatient data available from the data source. In this example, the AHA considered the hospital closed in 2017, but the data source still supplied data:

Table 4. Example of Hospital Closures

Year	Data Source	AHA IDNUMBER	HOSPID
2015	047	450520	SS171
2016	047	450520	SS171
2017	047	450520 (closed)	SS171

Mergers

When two or more hospitals merge (in the AHA's view), the IDNUMBER (along with the AHAID and HOSPID) of the merged entity is assigned to all its component hospitals even if they continue reporting separately to the state. In this example, two hospitals have different source identifiers, but starting in 2017 are considered one facility by the AHA because of a merger:

Table 5. Example of Hospital Mergers

Year	Data Source	AHA IDNUMBER	HOSPID
2016	036	450400	SS091
2017	036	450002 (merger)	SS013
2018	036	450002 (merger)	SS013
2019	036	450002 (merger	SS013
2016	126	451750	SS169
2017	126	450002 (merger)	SS013
2018	126	450002 (merger)	SS013

| | 2019 | 126 | 450002 (merger) | SS013 |

Demergers

When hospitals demerge (in the AHA's view), the component hospitals are assigned a new AHAID/IDNUMBER or the one they previously had. The HCUP HOSPID follows the AHA IDNUMBER, so that the HOSPID changes if the AHAID/IDNUMBER changes and the HOSPID is reused if the AHAID/IDNUMBER is reused. In this example, a hospital demerges in 2016 into two facilities:

Table 6. Example of Hospital Demergers

Year	Data Source	AHA IDNUMBER	HOSPID
2015	562	220515 (merger)	SS051
2016	562	220547 (demerger)	SS026
2017	562	220547 (demerger)	SS026
2015	561	220515 (merger)	SS051
2016	561	221240 (demerger)	SS037
2017	561	221240 (demerger)	SS037

Changes in Hospital Characteristics

Note: The following decision is made only after AHRQ is consulted. If during HCUP processing of the data, summary statistics on the distribution of length of stay look questionable for a community hospital (e.g., the mean length of stay is considerably greater than 30 days), the AHA community flag is investigated. If the AHA community flag was imputed from previous years because a hospital did not report to the AHA, and the data source can confirm that the facility is no longer a community hospital, the AHA identifier is still assigned to the facility and the community flag is imputed. In this example, the facility was considered a non-community hospital starting in 2017:

Table 7. Example of Changes in Hospital Characteristics

Year	Data Source	AHA IDNUMBER	HOSPID	Community Flag
2015	86-0201864	860575	SS090	1
2016	86-0201864	860575	SS090	1
2017	86-0201864	860575	SS101	0 (changed)
2018	86-0201864	860575	SS101	0 (changed)