

HEALTHCARE COST AND UTILIZATION PROJECT — HCUP
A FEDERAL-STATE-INDUSTRY PARTNERSHIP IN HEALTH DATA
Sponsored by the Agency for Healthcare Research and Quality

INTRODUCTION TO
THE HCUP STATE INPATIENT DATABASES (SID)

These pages provide only an introduction to the SID package.

Full documentation is provided online at the HCUP User Support website:

www.hcup-us.ahrq.gov

Issued January 2025

Agency for Healthcare Research and Quality
Healthcare Cost and Utilization Project (HCUP)
5600 Fishers Lane
Mail Stop 7W25B
Rockville, MD 20857

SID Data and Documentation Distributed through the HCUP Central Distributor

Website: www.hcup-us.ahrq.gov
Phone: (866) 290-HCUP (4287) (toll free)
Email: hcup@ahrq.gov

Table of Contents

SUMMARY OF DATA USE LIMITATIONS	2
HCUP DATA USE AGREEMENT REQUIREMENTS	3
HCUP CONTACT INFORMATION	3
ABSTRACT	4
INTRODUCTION TO THE HCUP STATE INPATIENT DATABASES (SID)	5
OVERVIEW OF THE SID	5
How the HCUP SID Differ from State Data Files.....	6
What Types of Hospitals Are Included in the SID?.....	6
How to Identify Hospitals in the SID.....	7
What is the File Structure of the SID in the 2019–2023 Files?.....	7
What is the File Structure of the SID in the 2016–2018 Files?.....	9
What is the File Structure of the SID in the 2015 Files?	10
What is the File Structure of the SID in the 2005–2014 Files?.....	13
What is the File Structure of the SID in the 1998–2004 Files?.....	15
What is the File Structure of the SID in the 1995–1997 Files?.....	16
What is the File Structure of the SID in the 1990–1994 Files?.....	17
GETTING STARTED	19
SID Data Files.....	19
SID Programs, Documentation, and Tools.....	19

**HCUP STATE INPATIENT DATABASES (SID)
SUMMARY OF DATA USE LIMITATIONS**

******* REMINDER *******

All users of the SID must take the online Data Use Agreement (DUA) training course, and read and sign a Data Use Agreement. Details and links may be found on the following page.

Authorized users of HCUP data agree to the following restrictions:^a

- Will not use the data for any purpose other than research, analysis, and aggregate statistical reporting.
- Will not re-release any data to unauthorized users.
- Will not redistribute HCUP data by posting on any website or publishing in any other publicly accessible online repository. If a journal or publication requests access to data or analytic files, I will cite restrictions on data sharing in the Data Use Agreement and direct them to AHRQ HCUP (www.hcup-us.ahrq.gov) for more information on accessing HCUP data.
- Will not identify or attempt to identify any individual, including by the use of vulnerability analysis or penetration testing. Methods that could be used to identify individuals directly or indirectly shall not be disclosed or published.
- Will not report any statistics where the number of observations (i.e., individual discharge records) in any given cell of tabulated data is less than or equal to 10 (≤ 10).
- Will not publish information that could identify individual establishments (e.g., hospitals), and will not contact establishments.
- Will not use the data concerning individual establishments for commercial or competitive purposes affecting establishments, or to determine rights, benefits, or privileges of establishments.
- Will not use the data for criminal and civil litigation, including expert witness testimony or for law enforcement activities.
- Will acknowledge in reports that data from the “Healthcare Cost and Utilization Project (HCUP)” were used, including names of the specific databases used for analysis.^b

Any violation of the limitations in the data use agreement is punishable under Federal law by a fine, up to five years in prison, or both. Violations may also be subject to penalties under State statutes.

^a This is a summary of key terms of the Data Use Agreement for HCUP State Databases, please refer to the DUA for full terms and conditions.

^b Suggested citations for the HCUP databases are provided in the Requirements for Publishing with HCUP Data available at www.hcup-us.ahrq.gov/db/publishing.jsp.

HCUP DATA USE AGREEMENT REQUIREMENTS

All HCUP data users, including data purchasers and collaborators, must complete the online HCUP Data Use Agreement (DUA) Training Tool, and read and sign the HCUP Data Use Agreement. Proof of training completion and signed Data Use Agreements must be submitted to the HCUP Central Distributor.

Data purchasers will be required to provide their DUA training completion code and will execute their DUAs electronically as a part of the online ordering process. The DUAs and training certificates for collaborators and others with access to HCUP data should be submitted directly to the HCUP Central Distributor using the contact information below.

The on-line DUA training course is available at: www.hcup-us.ahrq.gov/tech_assist/dua.jsp.

The **HCUP Data Use Agreement for the State Databases** is available on the HCUP User Support (HCUP-US) website.

- PDF file, 325 KB: www.hcup-us.ahrq.gov/team/StateDUA.pdf

HCUP CONTACT INFORMATION

HCUP Central Distributor and HCUP User Support

Information about the content of the HCUP databases is available on the HCUP User Support (HCUP-US) website (www.hcup-us.ahrq.gov).

If you have questions, please review the HCUP Frequently Asked Questions located at www.hcup-us.ahrq.gov/tech_assist/faq.jsp.

If you need further technical assistance, please contact the HCUP Central Distributor and User Support team at:

Phone: (866) 290-HCUP (4287) (toll free)

Email: hcup@ahrq.gov

We would like to receive your feedback on the HCUP data products.

Please send user feedback to hcup@ahrq.gov.

HEALTHCARE COST AND UTILIZATION PROJECT—HCUP
A FEDERAL-STATE-INDUSTRY PARTNERSHIP IN HEALTH DATA
Sponsored by the Agency for Healthcare Research and Quality

**The Agency for Healthcare Research and Quality and
the staff of the Healthcare Cost and Utilization Project (HCUP) thank you for your
interest in the HCUP State Inpatient Databases (SID)**

HCUP State Inpatient Databases (SID)

ABSTRACT

The State Inpatient Databases (SID) are part of the Healthcare Cost and Utilization Project (HCUP), sponsored by the Agency for Healthcare Research and Quality (AHRQ).

The HCUP State Inpatient Databases (SID) are a powerful set of hospital databases from data organizations in participating States.

- The SID contain the universe of the inpatient discharge abstracts from participating States that are translated into a uniform format to facilitate multistate comparisons and analyses.
- Together, the SID encompass almost 97 percent of all U.S. hospital discharges. Some States include discharges from specialty hospitals, such as acute psychiatric hospitals.
- The SID contain a core set of clinical and nonclinical information on all visits, regardless of the expected payer, including but not limited to Medicare, Medicaid, private insurance, self-pay, or those billed as 'no charge'.
- In addition to the core set of uniform data elements common to all SID, some include other elements such as the patient's race.

Researchers and policymakers use SID to investigate questions unique to one State; to compare data from two or more States; to conduct market-area variation analyses; and to identify State-specific trends in inpatient care utilization, access, charges, and outcomes.

The individual State databases are in the same HCUP uniform format and represent 100 percent of records processed by AHRQ. However, the participating data organizations control the release of specific data elements. AHRQ is currently assisting the data organizations in the release of the 1990–2023 SID.

The SID can be linked to hospital-level data from the American Hospital Association (AHA) Annual Survey of Hospitals and county-level data from the Bureau of Health Professions' Area Resource File, except in States that do not allow the release of hospital identifiers.

Thirty-four of the data organizations participating in HCUP have agreed to release their 2023 SID files through the HCUP Central Distributor under the auspices of AHRQ. Uses are limited to research and aggregate statistical reporting.

INTRODUCTION TO THE HCUP STATE INPATIENT DATABASES (SID)

OVERVIEW OF THE SID

The Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID) consist of individual data files from 49 participating data organizations. In general, the SID contain the universe of that State's hospital inpatient discharge records. They are composed of annual, State-specific files that share a common structure and common data elements. Most data elements are coded in a uniform format across all States. In addition to the core set of uniform data elements, the SID include State-specific data elements or data elements available only for a limited number of States. The uniform format of the SID helps facilitate cross-State comparisons. In addition, the SID are well suited for research that requires complete enumeration of hospitals and discharges within market areas or States.

Thirty-four of the 49 data organizations that participate in HCUP have agreed to release their 2023 SID files through the HCUP Central Distributor under the auspices of AHRQ. The individual State databases are in the same HCUP uniform format and represent 100 percent of records processed by AHRQ. However, the participating data organizations control the release of specific data elements.

SID data sets are currently available for multiple States and years. Each release of the SID includes:

- Data in American Standard Code for Information Interchange (ASCII) format on a digital versatile disc with read-only memory (DVD-ROM).
- Patient-level hospital discharge abstract data for 100 percent of discharges from hospitals in participating States.
- American Hospital Association (AHA) Linkage File to link the SID to data from the AHA Annual Survey of Hospitals. This is only available for those States that allow the release of hospital identifiers.

The SID are calendar year files for all data years except 2015. Because of the transition to ICD-10-CM/PCS¹ on October 1, 2015, the 2015 SID are split into two parts. Nine months of the 2015 data with ICD-9-CM² codes (discharges from Jan 1, 2015 – September 30, 2015) are in one set of files labeled Q1Q3. Three months of 2015 data with ICD-10-CM/PCS codes (discharges from October 1, 2015 – December 31, 2015) are in a separate set of files labeled Q4. More information about the changes to the HCUP databases for ICD-10-CM/PCS and use of data across the two coding systems may be found on the HCUP User Support website under ICD-10-CM/PCS Resources (www.hcupus.ahrq.gov/datainnovations/icd10_resources.jsp).

SID documentation and tools—including file specifications, programming source code for loading ASCII data into SAS (SAS Institute Inc.; Cary, NC), SPSS (IBM Corp.; Armonk, NY), and Stata (StataCorp; College Station, TX), and value labels—are available online at the HCUP User Support website (www.hcup-us.ahrq.gov).

Starting with the 2006 SID, the AHA Linkage files are available via the HCUP User Support website (www.hcup-us.ahrq.gov). The AHA Linkage files may not be available when the discharge-level database is released.

¹ ICD-10-CM/PCS: International Classification of Diseases, 10th Edition, Clinical Modification/ Procedure Coding System

² ICD-9-CM: International Classification of Diseases, Ninth Edition, Clinical Modification

How the HCUP SID Differ from State Data Files

The SID available through the HCUP Central Distributor differ from the data files available from the data organizations in the following ways:

- Data elements available on the files
- Coding of data elements

Because the data organizations dictate the data elements that may be released through the HCUP Central Distributor, the data elements on the SID are a subset of the data collected by the corresponding data organizations. HCUP uniform coding is used on most data elements on the SID. A few State-specific data elements retain the original values provided by the respective data organizations.

What Types of Hospitals Are Included in the SID?

The types of hospitals included in the SID depend on the information provided by the data organizations and how the files were handled during HCUP processing. Most State government data organizations provide information on all acute care hospitals in the respective State. Private data organizations are often restricted to member hospitals and may not provide information on all hospitals in their State.

Beginning with the 1994 SID, all hospitals reported by the data organizations were retained in the SID files. Discharges from facilities such as psychiatric facilities, alcohol and drug dependency facilities, and State, Federal, and Veterans Affairs hospitals will be in the SID, if reported by the data source. Prior to 1994, only discharges from community hospitals were retained in the SID.

Community hospitals, as defined by the AHA, include “all nonfederal, short-term, general and other specialty hospitals, excluding hospital units of institutions.” Included among community hospitals are academic medical centers and specialty hospitals such as obstetrics, gynecology, otolaryngology, short-term rehabilitation, orthopedic, and pediatric hospitals. Noncommunity hospitals include Federal hospitals (e.g., Veterans Affairs, Department of Defense, and Indian Health Service hospitals), long-term hospitals, psychiatric hospitals, alcohol/chemical dependency treatment facilities, and hospital units within institutions such as prisons.

Some community hospitals may not be included in the SID because their data were not provided by the data source. To identify community hospitals, the SID must be linked to the AHA Annual Survey of Hospitals by the AHA hospital identifier.

Tables showing the number of hospitals in the SID can be found online at the HCUP User Support website (www.hcup-us.ahrq.gov). The tables present the hospitals by the number of:

- Community hospitals
- Hospitals not classified as community
- Community hospitals not in the SID

Information contained in the AHA Annual Survey of Hospitals was used to determine if a hospital was a community hospital. Some hospitals could not be categorized as community or noncommunity hospitals because they could not be matched with AHA information. This occurs when a hospital was closed in a previous year or when the hospital does not report to the AHA.

How to Identify Hospitals in the SID

Up to three hospital identifiers are on the SID:

- Some data organizations allow the original hospital identifier (DSHOSPID) to be included on the SID. If available on the SID, this identifier is coded for all hospitals and may distinguish different units within a hospital.
- Some data organizations allow the AHA hospital identifier (AHAID) to be included on the SID. This variable enables the SID to be linked to the AHA Annual Survey of Hospitals that contains information on hospital characteristics. The AHA hospital identifier is coded for most hospitals. The AHA hospital identifier is missing if the hospital is not registered with the AHA or if the source-provided information could not be matched to the AHA.
- Some data organizations allow the HCUP-specific hospital identifier (HOSPID) to be included on the SID. HOSPID is coded for all hospitals with a nonmissing AHA hospital identifier.

What is the File Structure of the SID in the 2019–2023 Files?

Based on the availability of data elements across States, data elements included in the 2019–2023 SID are structured as follows:

- Core file
- Charges file
- AHA Linkage file
- Diagnosis and Procedure Groups file

The **Core file** is a discharge-level file that contains:

- Core data elements that form the nucleus of the SID
- State-specific data elements intended for limited use

Core data elements meet at least one of the following criteria:

- Are available from all or nearly all data sources
- Lend themselves to uniform coding across sources
- Are needed for traditional applications (e.g., length of stay, patient age)

State-specific data elements meet at least one of the following criteria:

- Are available from a limited number of sources
- Do not lend themselves to uniform coding across sources
- Are not needed for traditional applications

The Core file is a discharge-level file with one observation per discharge abstract.

The **Charges file** contains detailed charge information. There are three kinds of Charges files:

- 1) *Summarized detail* in which charge information is summed within the revenue center. This type of Charges file includes one record per discharge abstract. Each record contains three corresponding arrays with the following information:
 - a. Revenue center (REVCDn)
 - b. Total charge for the revenue center (CHGn)
 - c. Total units of service for the revenue center (UNITn)

For example, if a patient had five laboratory tests, REVCD1 would include the revenue code for laboratory, CHG1 would include the total charge for the five tests, and UNIT1 would be five. To combine data elements between this type of Charges file and the Core file, merge

the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 2) *Collapsed detail* in which charge information is summed across revenue centers. This type of Charges file includes one record per discharge abstract. Each record contains an array of collapsed charges (CHGn) that are predefined by the data organization that provided the data.

Consider the example of a patient that had five laboratory tests from different revenue centers in the range of 300 to 319. CHG1, which was predefined as Laboratory Charges for revenue centers 300–319, would include the total charge for the five tests, but there is no detail on which specific revenue centers were used. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 3) *Line item detail* in which a submitted charge pertains to a specified revenue center and there may be multiple charges reported for the same revenue center. This type of Charges file includes multiple records per discharge abstract. Each record includes the following information for one service:

- a. Revenue center (REVCODE)
- b. Charge (CHARGE)
- c. Unit of service (UNITS)
- d. Day of service (SERVDAY) for some files

For example, if a patient had five laboratory tests, there are five records in the Charges file with information on the charge for each laboratory test. Information from this type of Charges file may be combined with the Core file by the unique record identifier (KEY), but there is not a one-to-one correspondence of records.

Refer to the Description of Data Elements online at the HCUP User Support website (www.hcup-us.ahrq.gov) for more information on the charge information from the different States.

The **AHA Linkage file** contains AHA linkage data elements that allow the SID to be used in conjunction with the AHA Annual Survey of Hospitals data files. These files contain information about hospital characteristics and are available for purchase through the AHA. Because the data organizations in participating States determine whether the AHA linkage data elements may be released through the HCUP Central Distributor with the SID, not all SID include AHA linkage data elements.

The AHA Linkage file is a hospital-level file with one observation per hospital or facility. To combine the discharge-level files with the hospital-level file (AHA Linkage file), merge the files by the hospital identifier provided by the data source (DSHOSPID), but be careful of the different levels of aggregation. For example, the Core file may contain 5,000 discharges for DSHOSPID "A," but the Hospital file contains only one record for DSHOSPID "A."

Starting with the 2006 SID, the AHA Linkage files are available via the HCUP User Support website (www.hcup-us.ahrq.gov). The AHA Linkage files may not be available when the discharge-level database is released.

Diagnosis and Procedure Groups File is a discharge-level file that contains data elements from AHRQ software tools. They are designed to facilitate the use of the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM/PCS) diagnostic and procedure information in the HCUP databases. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (KEY) provides the linkage between the Core files and the Diagnosis and Procedure Groups files. These files are available beginning with the 2019 SID.

What is the File Structure of the SID in the 2016–2018 Files?

Based on the availability of data elements across States, data elements included in the 2016–2018 SID are structured as follows:

- Core file
- Charges file
- AHA Linkage file

Unavailable with the 2016–2018 SID are two file types that had been included with the SID in prior data years: the Diagnosis and Procedure Groups file and the Disease Severity file. The data elements included in those two files were derived from AHRQ software tools. If you are interested in applying the AHRQ software tools to the ICD-10-CM/PCS data in the 2016–2018 SID, beta versions of the AHRQ software tools are available on the [HCUP User Support website](http://www.hcup-us.ahrq.gov/tools_software.jsp) at www.hcup-us.ahrq.gov/tools_software.jsp. Also available is a tutorial on how to apply the AHRQ software tools to the HCUP databases at www.hcup-us.ahrq.gov/tech_assist/tutorials.jsp.

The **Core file** is a discharge-level file that contains:

- Core data elements that form the nucleus of the SID
- State-specific data elements intended for limited use

Core data elements meet at least one of the following criteria:

- Are available from all or nearly all data sources
- Lend themselves to uniform coding across sources
- Are needed for traditional applications (e.g., length of stay, patient age)

State-specific data elements meet at least one of the following criteria:

- Are available from a limited number of sources
- Do not lend themselves to uniform coding across sources
- Are not needed for traditional applications

The Core file is a discharge-level file with one observation per discharge abstract.

The **Charges file** contains detailed charge information. There are three kinds of Charges files:

- 1) *Summarized detail* in which charge information is summed within the revenue center. This type of Charges file includes one record per discharge abstract. Each record contains three corresponding arrays with the following information:
 - a. Revenue center (REVCDn)
 - b. Total charge for the revenue center (CHGn)
 - c. Total units of service for the revenue center (UNITn)

For example, if a patient had five laboratory tests, REVCD1 would include the revenue code for laboratory, CHG1 would include the total charge for the five tests, and UNIT1 would be five. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 2) *Collapsed detail* in which charge information is summed across revenue centers. This type of Charges file includes one record per discharge abstract. Each record contains an array of collapsed charges (CHGn) that are predefined by the data organization that provided the data.

Consider the example of a patient that had five laboratory tests from different revenue centers in the range of 300 to 319. CHG1, which was predefined as Laboratory Charges for revenue centers 300–319, would include the total charge for the five tests, but there is no detail on which specific revenue centers were used. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 3) *Line item detail* in which a submitted charge pertains to a specified revenue center and there may be multiple charges reported for the same revenue center. This type of Charges file includes multiple records per discharge abstract. Each record includes the following information for one service:
 - a. Revenue center (REVCODE)
 - b. Charge (CHARGE)
 - c. Unit of service (UNITS)
 - d. Day of service (SERVDAY) for some files

For example, if a patient had five laboratory tests, there are five records in the Charges file with information on the charge for each laboratory test. Information from this type of Charges file may be combined with the Core file by the unique record identifier (KEY), but there is not a one-to-one correspondence of records.

Refer to the Description of Data Elements online at the HCUP User Support website (www.hcup-us.ahrq.gov) for more information on the charge information from the different States.

The **AHA Linkage file** contains AHA linkage data elements that allow the SID to be used in conjunction with the AHA Annual Survey of Hospitals data files. These files contain information about hospital characteristics and are available for purchase through the AHA. Because the data organizations in participating States determine whether the AHA linkage data elements may be released through the HCUP Central Distributor with the SID, not all SID include AHA linkage data elements.

The AHA Linkage file is a hospital-level file with one observation per hospital or facility. To combine the discharge-level files with the hospital-level file (AHA Linkage file), merge the files by the hospital identifier provided by the data source (DSHOSPID), but be careful of the different levels of aggregation. For example, the Core file may contain 5,000 discharges for DSHOSPID “A,” but the Hospital file contains only one record for DSHOSPID “A.”

Starting with the 2006 SID, the AHA Linkage files are available via the HCUP User Support website (www.hcup-us.ahrq.gov). The AHA Linkage files may not be available when the discharge-level database is released.

What is the File Structure of the SID in the 2015 Files?

The file structure of the 2015 SID is similar to previous years (and future years) in terms of how data elements are split across multiple data files, but differs from others years because the records within the 2015 files have been separated into two sets of files based on the discharge date because of the transition from reporting medical diagnoses and inpatient procedures using ICD-9-CM to the ICD-10-CM/PCS code sets.³

³ ICD-9-CM: International Classification of Diseases, Ninth Edition, Clinical Modification; ICD-10-CM/PCS: International Classification of Diseases, 10th Edition, Clinical Modification/ Procedure Coding System

The 2015 SID are split into two separate sets of files based on the discharge date and different coding schemes:

- If the record was discharged between January 1, 2015, and September 30, 2015, it is retained in the quarter 1 to quarter 3 (Q1–Q3) files and includes ICD-9-CM data
- If the record was discharged between October 1, 2015, and December 31, 2015, it is retained in the quarter 4 (Q4) files and includes ICD-10-CM/PCS data

Almost all of the diagnosis and procedure-related data elements that are based on ICD-10-CM/PCS data have been renamed with the prefix of I10 to distinguish them from the ICD-9-CM-based data element. Exceptions include data elements that are based on third-party proprietary software such as the Diagnosis Related Groups (DRGs) and the All Patient Refined DRG (APR-DRG).

Based on the availability of data elements across States, data elements included in the 2015 SID are structured as follows:

- Core files, one for Q1–Q3 and one for Q4
- Charges files, one for Q1–Q3 and one for Q4
- AHA Linkage file (one file because this is a hospital-level file, instead of a discharge-level file)
- Diagnosis and Procedure Groups files, one for Q1–Q3 and one for Q4
- Disease Severity Measures files, one for Q1–Q3 and one for Q4

The **Core file** is a discharge-level file that contains:

- Core data elements that form the nucleus of the SID
- State-specific data elements intended for limited use

Core data elements meet at least one of the following criteria:

- Are available from all or nearly all data sources
- Lend themselves to uniform coding across sources
- Are needed for traditional applications (e.g., length of stay, patient age)

State-specific data elements meet at least one of the following criteria:

- Are available from a limited number of sources
- Do not lend themselves to uniform coding across sources
- Are not needed for traditional applications

The Core file is a discharge-level file with one observation per discharge abstract.

The **Charges file** contains detailed charge information. There are three kinds of Charges files:

- 1) *Summarized detail* in which charge information is summed within the revenue center. This type of Charges file includes one record per discharge abstract. Each record contains three corresponding arrays with the following information:
 - a. Revenue center (REVCD_n)
 - b. Total charge for the revenue center (CHG_n)
 - c. Total units of service for the revenue center (UNIT_n)

For example, if a patient had five laboratory tests, REVCD1 would include the revenue code for laboratory, CHG1 would include the total charge for the five tests, and UNIT1 would be five. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 2) *Collapsed detail* in which charge information is summed across revenue centers. This type of Charges file includes one record per discharge abstract. Each record contains an array of

collapsed charges (CHGn) that are predefined by the data organization that provided the data.

Consider the example of a patient that had five laboratory tests from different revenue centers in the range of 300 to 319. CHG1, which was predefined as Laboratory Charges for revenue centers 300–319, would include the total charge for the five tests, but there is no detail on which specific revenue centers were used. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 3) *Line item detail* in which a submitted charge pertains to a specified revenue center and there may be multiple charges reported for the same revenue center. This type of Charges file includes multiple records per discharge abstract. Each record includes the following information for one service:
 - a. Revenue center (REVCODE)
 - b. Charge (CHARGE)
 - c. Unit of service (UNITS)
 - d. Day of service (SERVDAY) for some files

For example, if a patient had five laboratory tests, there are five records in the Charges file with information on the charge for each laboratory test. Information from this type of Charges file may be combined with the Core file by the unique record identifier (KEY), but there is not a one-to-one correspondence of records.

Refer to the Description of Data Elements online at the HCUP User Support website (www.hcup-us.ahrq.gov) for more information on the charge information from the different States.

The **AHA Linkage file** contains AHA linkage data elements that allow the SID to be used in conjunction with the AHA Annual Survey of Hospitals data files. These files contain information about hospital characteristics and are available for purchase through the AHA. Because the data organizations in participating States determine whether the AHA linkage data elements may be released through the HCUP Central Distributor with the SID, not all SID include AHA linkage data elements.

The AHA Linkage file is a hospital-level file with one observation per hospital or facility. To combine the discharge-level files with the hospital-level file (AHA Linkage file), merge the files by the hospital identifier provided by the data source (DSHOSPID), but be careful of the different levels of aggregation. For example, the Core file may contain 5,000 discharges for DSHOSPID “A,” but the Hospital file contains only one record for DSHOSPID “A.”

Starting with the 2006 SID, the AHA Linkage files are available via the HCUP User Support website (www.hcup-us.ahrq.gov). The AHA Linkage files may not be available when the discharge-level database is released.

The **Diagnosis and Procedure Groups File** is a discharge-level file that contains data elements from AHRQ software tools. In Q1–Q3, they are designed to facilitate the use of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic and procedure information in the HCUP databases. In Q4, they are designed to facilitate the use of the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM/PCS) diagnostic and procedure information in the HCUP databases. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (KEY) provides the linkage between the Core files and the Diagnosis and Procedure Groups files.

Disease Severity Measures File is a discharge-level file that contains information from the AHRQ Comorbidity Software. Information from these severity files is to be used in conjunction with the Inpatient Core files. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (KEY) provides the linkage between the Core files and the Disease Severity Measures files.

What is the File Structure of the SID in the 2005–2014 Files?

Based on the availability of data elements across States, data elements included in the 2005–2014 SID are structured as follows:

- Core file
- Charges file
- AHA Linkage file
- Diagnosis and Procedure Groups file
- Disease Severity Measures file

The **Core file** is a discharge-level file that contains:

- Core data elements that form the nucleus of the SID
- State-specific data elements intended for limited use

Core data elements meet at least one of the following criteria:

- Are available from all or nearly all data sources
- Lend themselves to uniform coding across sources
- Are needed for traditional applications (e.g., length of stay, patient age)

State-specific data elements meet at least one of the following criteria:

- Are available from a limited number of sources
- Do not lend themselves to uniform coding across sources
- Are not needed for traditional applications

The Core file is a discharge-level file with one observation per discharge abstract.

The **Charges file** contains detailed charge information. There are three kinds of Charges files:

- 1) *Summarized detail* in which charge information is summed within the revenue center. This type of Charges file includes one record per discharge abstract. Each record contains three corresponding arrays with the following information:
 - a. Revenue center (REVCD_n)
 - b. Total charge for the revenue center (CHG_n)
 - c. Total units of service for the revenue center (UNIT_n)

For example, if a patient had five laboratory tests, REVCD1 would include the revenue code for laboratory, CHG1 would include the total charge for the five tests, and UNIT1 would be five. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 2) *Collapsed detail* in which charge information is summed across revenue centers. This type of Charges file includes one record per discharge abstract. Each record contains an array of collapsed charges (CHG_n) that are predefined by the data organization that provided the data.

Consider the example of a patient that had five laboratory tests from different revenue centers in the range of 300 to 319. CHG1, which was predefined as Laboratory Charges for revenue centers 300–319, would include the total charge for the five tests, but there is no detail on which specific revenue centers were used. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 3) *Line item detail* in which a submitted charge pertains to a specified revenue center and there may be multiple charges reported for the same revenue center. This type of Charges file

includes multiple records per discharge abstract. Each record includes the following information for one service:

- a. Revenue center (REVCODE)
- b. Charge (CHARGE)
- c. Unit of service (UNITS)
- d. Day of service (SERVDAY) for some files

For example, if a patient had five laboratory tests, there are five records in the Charges file with information on the charge for each laboratory test. Information from this type of Charges file may be combined with the Core file by the unique record identifier (KEY), but there is not a one-to-one correspondence of records.

Refer to the Description of Data Elements online at the HCUP User Support website (www.hcup-us.ahrq.gov) for more information on the charge information from the different States.

The **AHA Linkage file** contains AHA linkage data elements that allow the SID to be used in conjunction with the AHA Annual Survey of Hospitals data files. These files contain information about hospital characteristics and are available for purchase through the AHA. Because the data organizations in participating States determine whether the AHA linkage data elements may be released through the HCUP Central Distributor with the SID, not all SID include AHA linkage data elements.

The AHA Linkage file is a hospital-level file with one observation per hospital or facility. To combine the discharge-level files with the hospital-level file (AHA Linkage file), merge the files by the hospital identifier provided by the data source (DSHOSPID), but be careful of the different levels of aggregation. For example, the Core file may contain 5,000 discharges for DSHOSPID "A," but the Hospital file contains only one record for DSHOSPID "A."

Starting with the 2006 SID, the AHA Linkage files are available via the HCUP User Support website (www.hcup-us.ahrq.gov). The AHA Linkage files may not be available when the discharge-level database is released.

Diagnosis and Procedure Groups File is a discharge-level file that contains data elements from AHRQ software tools. They are designed to facilitate the use of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic and procedure information in the HCUP databases. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (KEY) provides the linkage between the Core files and the Diagnosis and Procedure Groups files. These files are available beginning with the 2005 SID.

Disease Severity Measures File is a discharge-level file that contains information from the AHRQ Comorbidity Software. Information from these severity files is to be used in conjunction with the Inpatient Core files. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (KEY) provides the linkage between the Core files and the Disease Severity Measures files. These files are available beginning with the 2005 SID.

What is the File Structure of the SID in the 1998–2004 Files?

Based on the availability of data elements across States, data elements included in the 1998–2004 SID are structured as follows:

- Core file
- Charges file
- AHA Linkage file

The **Core file** is a discharge-level file that contains:

- Core data elements that form the nucleus of the SID
- State-specific data elements intended for limited use

Core data elements meet at least one of the following criteria:

- Are available from all or nearly all data sources
- Lend themselves to uniform coding across sources
- Are needed for traditional applications (e.g., length of stay, patient age)

State-specific data elements meet at least one of the following criteria:

- Are available from a limited number of sources
- Do not lend themselves to uniform coding across sources
- Are not needed for traditional applications

The Core file is a discharge-level file with one observation per discharge abstract.

The **Charges file** contains detailed charge information. There are two kinds of Charges files:

- 1) *Summarized detail* in which charge information is summed within the revenue center. This type of Charges file includes one record per discharge abstract. Each record contains three corresponding arrays with the following information:
 - a. Revenue center (REVCDn)
 - b. Total charge for the revenue center (CHGn)
 - c. Total units of service for the revenue center (UNITn)

For example, if a patient had five laboratory tests, REVCD1 would include the revenue code for laboratory, CHG1 would include the total charge for the five tests, and UNIT1 would be five. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 2) *Collapsed detail* in which charge information is summed across revenue centers. This type of Charges file includes one record per discharge abstract. Each record contains an array of collapsed charges (CHGn) that are predefined by the data organization that provided the data.

Consider the example of a patient that had five laboratory tests from different revenue centers in the range of 300 to 319. CHG1, which was predefined as Laboratory Charges for revenue centers 300–319, would include the total charge for the five tests, but there is no detail on which specific revenue centers were used. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

Refer to the Description of Data Elements online at the HCUP User Support website (www.hcup-us.ahrq.gov) for more information on the charge information from the different States.

The **AHA Linkage file** contains AHA linkage data elements that allow the SID to be used in conjunction with the AHA Annual Survey of Hospitals data files. These files contain information about hospital characteristics and are available for purchase through the AHA. Because the data organizations in participating States determine whether the AHA linkage data elements may be released through the HCUP Central Distributor with the SID, not all SID include AHA linkage data elements.

The Core and Charges files are discharge-level files with one observation per abstract. The same record is represented in each file, but contains different data elements. To combine data elements across discharge-level files, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

The AHA Linkage file is a hospital-level file with one observation per hospital or facility. To combine discharge-level files with the hospital-level file (AHA Linkage file), merge the files by the hospital identifier provided by the data source (DSHOSPID), but be careful of the different levels of aggregation. For example, the Core file may contain 5,000 discharges for DSHOSPID "A," but the Hospital file contains only one record for DSHOSPID "A."

What is the File Structure of the SID in the 1995–1997 Files?

Based on the availability of data elements across States, data elements included in the 1995–1997 SID are structured as follows:

- Core file
- State-specific file
- AHA Linkage file

The **Core file** contains core data elements that form the nucleus of the SID. Core data elements meet at least one of the following criteria:

- Are available from all or nearly all data sources
- Lend themselves to uniform coding across sources
- Are needed for traditional applications (e.g., length of stay, patient age)

The **State-specific file** contains State-specific data elements intended for limited use. State-specific data elements meet at least one of the following criteria:

- Are available from a limited number of sources
- Do not lend themselves to uniform coding across sources
- Are not needed for traditional applications

The **AHA Linkage file** contains AHA linkage data elements that allow the SID to be used in conjunction with the AHA Annual Survey of Hospitals data files. These files contain information about hospital characteristics and are available for purchase through the AHA. Because the data organizations in participating States determine whether the AHA linkage data elements may be released through the HCUP Central Distributor with the SID, not all SID include AHA linkage data elements.

The Core and State-specific files are discharge-level files with one observation per abstract. The same record is represented in each file, but each contains different data elements. To combine data elements across discharge-level files, merge the files by the unique record identifier (SEQ_SID). There will be a one-to-one correspondence of records.

The AHA Linkage file is a hospital-level file with one observation per hospital or facility. To combine discharge-level files with the AHA Linkage file, merge the files by the hospital identifier provided by the data source (DSHOSPID), but be careful of the different levels of aggregation. For example, the Core may contain 5,000 discharges for DSHOSPID "A," but the Hospital file contains only one record for DSHOSPID "A."

What is the File Structure of the SID in the 1990–1994 Files?

Based on the availability of data elements across States, data elements included in the 1990–1994 SID are structured the same as the 1998–2004 files. This includes a maximum of three types of files:

- Core file
- Charges file
- AHA Linkage file

The **Core file** contains:

- Core data elements that form the nucleus of the SID
- State-specific data elements intended for limited use

Core data elements meet at least one of the following criteria:

- Are available from all or nearly all data sources
- Lend themselves to uniform coding across sources
- Are needed for traditional applications (e.g., length of stay, patient age)

State-specific data elements meet at least one of the following criteria:

- Are available from a limited number of sources
- Do not lend themselves to uniform coding across sources
- Are not needed for traditional applications

The **Charges file** contains detailed charge information. There are two kinds of Charges files:

- 1) *Summarized detail* in which charge information is summed within the revenue center. This type of Charges file includes one record per discharge abstract. Each record contains three corresponding arrays with the following information:
 - a. Revenue center (REVCDn)
 - b. Total charge for the revenue center (CHGn)
 - c. Total units of service for the revenue center (UNITn)

For example, if a patient had five laboratory tests, REVCD1 would include the revenue code for laboratory, CHG1 would include the total charge for the five tests, and UNIT1 would be five. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

- 2) *Collapsed detail* in which charge information is summed across revenue centers. This type of Charges file includes one record per discharge abstract. Each record contains an array of collapsed charges (CHGn) that are predefined by the data organization that provided the data.

Consider the example of a patient that had five laboratory tests from different revenue centers in the range of 300 to 319. CHG1, which was predefined as Laboratory Charges for revenue centers 300–319, would include the total charge for the five tests, but there is no detail on which specific revenue centers were used. To combine data elements between this type of Charges file and the Core file, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

Refer to the Description of Data Elements online at the HCUP User Support website (www.hcup-us.ahrq.gov) for more information on the charge information from the different States.

The **AHA Linkage file** contains AHA linkage data elements that allow the SID to be used in conjunction with the AHA Annual Survey of Hospitals data files. These files contain information about hospital characteristics and are available for purchase through the AHA. Because the data organizations in

participating States determine whether the AHA linkage data elements may be released through the HCUP Central Distributor with the SID, not all SID include AHA linkage data elements.

The Core and Charges files are discharge-level files with one observation per abstract. The same record is represented in each file, but contains different data elements. To combine data elements across discharge-level files, merge the files by the unique record identifier (KEY). There will be a one-to-one correspondence of records.

The AHA Linkage file is a hospital-level file with one observation per hospital or facility. To combine discharge-level files with the hospital-level file (AHA Linkage file), merge the files by the hospital identifier provided by the data source (DSHOSPID), but be careful of the different levels of aggregation. For example, the Core file may contain 5,000 discharges for DSHOSPID "A," but the Hospital file contains only one record for DSHOSPID "A."

GETTING STARTED

SID Data Files

SID Data Files are provided on DVD-ROMs. The number of DVD-ROMs depends on the State and year of data. To load and analyze SID data onto your computer, users will need the following:

- One to four gigabytes of space available, depending on which database you are using.
- A third-party zip utility such as ZIP Reader, 7-Zip®, SecureZIP®, WinZip®, or Stuffit Expander®.
- SAS®, SPSS®, Stata® or similar analysis software.

To extract the data files from the compressed download file, follow these steps:

- 1) Create a directory for the SID on your hard drive.
- 2) Unzip the compressed files into the new directory using a third-party zip utility. This will create compressed, encrypted data-related files in the new directory. You will be prompted to enter the encryption password (sent separately by email) to decrypt the file.

Please note that attempts to unzip encrypted files using the built-in zip utility in Windows® (Windows Explorer) or Macintosh® (Archive Utility) will produce an error message warning of incorrect password and/or file or folder errors. The solution is to use a third-party zip utility.

Third-party zip utilities are available from the following reputable vendors on their official websites.

- ZIP Reader (Windows) (free download offered by the PKWARE corporation)
 - 7-Zip® (Windows) (free download offered by 7-Zip)
 - SecureZIP® for Mac or Windows (free evaluation and licensed/fee software offered by the PKWARE corporation)
 - WinZip (Windows) (evaluation and fee versions offered by the WinZip corporation)
 - Stuffit Expander® (Mac) (free evaluation and licensed/fee software offered by Smith Micro corporation)
- 3) Unzip each of the compressed, encrypted data-related files using the passwords provided and third-party zip utility method. This will place the data-related ASCII files in this same directory by default.

SID Programs, Documentation, and Tools

The SID programs, technical documentation and HCUP tools are available online via the Databases page at the HCUP User Support website (www.hcup-us.ahrq.gov/databases.jsp). The site provides important resources for SID users, and all of the files may be downloaded free of charge. A summary is provided in Table 1.

The SID programs include SAS, SPSS, and Stata load programs containing the programming code necessary to convert SID ASCII files into SAS, SPSS, or Stata. Please note that for the 2015 SID, there will be one set of load programs for the Q1–Q3 files and another set of load programs for the Q4 files.

The SID technical documentation provides detailed descriptions of the structure and content of the SID.

The HCUP Tools include the Clinical Classifications Software Refined (CCSR) and general label and format information applicable to all HCUP databases.

Information intended to summarize key issues to be anticipated by researchers before analyzing health services outcomes in the HCUP databases that include ICD-10-CM/PCS coding is included on the HCUP User Support website (www.hcup-us.ahrq.gov/datainnovations/icd10_resources.jsp). The section discusses key differences in the structure of HCUP databases, presents preliminary coding differences that were observed in HCUP databases, and provides general guidance and forewarning to users interested in analyzing outcomes that are potentially impacted by the transition.

Table 1. SID Related Reports and Database Documentation Available on HCUP-US

<p>Description of SID Database</p> <ul style="list-style-type: none"> • SID Overview • Introduction to the SID (<i>this document</i>) • SID File Compositions—describes types of hospitals and types of records included in each SID (e.g., number of discharges and hospitals by year, number of diagnoses and procedures) • SID-Related Reports <p>Restrictions on Use</p> <ul style="list-style-type: none"> • HCUP Data Use Agreement Training • SID Data Use Agreement • Requirements for Publishing with HCUP Data <p>File Specifications and Load Programs</p> <ul style="list-style-type: none"> • File Specifications—details data file names, number of records, record length, and record layout (e.g., file size by year) • SAS Load Programs • SPSS Load Programs • Stata Load Programs <p>Data Elements</p> <ul style="list-style-type: none"> • Availability of States Across All Years • Availability of Data Elements and Descriptions by Year • Summary Statistics for All States Across All Years—lists means and frequencies on nearly all data elements <p>Additional Resources for Data Elements</p> <ul style="list-style-type: none"> • HCUP Quality Control Procedures—describes procedures used to assess data quality • HCUP Coding Practices—describes how HCUP data elements are coded • HCUP Hospital Identifiers—explains data elements that characterize individual hospitals across States and Years 	<p>ICD-10-CM/PCS Data Included in the SID Starting With 2015</p> <ul style="list-style-type: none"> • 2016 State Databases Revised File Structure and New Data Elements • Caution: 2015 SID Includes ICD-9-CM and ICD-10-CM/PCS Data <ul style="list-style-type: none"> ○ 2015 State Databases Revised File Structure and New Data Elements • Additional ICD-10-CM/PCS Resources • Tutorial for Loading HCUP Software Tools for ICD-10-CM/PCS <p>Known Data Issues</p> <ul style="list-style-type: none"> • Includes State-specific information on databases that have been updated or have known data issues <p>HCUP Tools: Labels and Formats</p> <ul style="list-style-type: none"> • DRG Formats Program—Creates SAS formats to label the values of each DRG and MDC category • HCUP Formats Program—Creates SAS formats to label the values of selected categorical data elements in HCUP files • HCUP Diagnosis and Procedure Groups Formats Program—Creates SAS formats to label the values of HCUP Diagnosis and Procedure Groups data elements, including Clinical Classifications Software Refined (CCSR) data elements • ICD-9-CM Formats Program—Creates SAS formats to label the values of ICD-9-CM Diagnoses and Procedures • ICD-10-CM Formats Program—Creates SAS formats to label the values of ICD-10-CM Diagnoses and Procedures • Severity Formats Program—Creates SAS formats to label the values of data elements in HCUP Severity Files <p>HCUP Supplemental Files</p> <ul style="list-style-type: none"> • American Hospital Association Linkage Files • Cost-to-Charge Ratio Files • Hospital Market Structure (HMS) Files • HCUP Variables for Revisit Analysis <p>Obtaining HCUP Data</p> <ul style="list-style-type: none"> • Purchase HCUP data from the HCUP Central Distributor
---	---