

**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:**

<http://www.hcup-us.ahrq.gov>

Issued May 2012

Agency for Healthcare Research and Quality  
Healthcare Cost and Utilization Project (HCUP)  
540 Gaither Road, 5<sup>th</sup> Floor  
Rockville, Maryland 20850

Phone: (866) 290-HCUP (4287)  
Fax: (301) 427-1430  
E-mail: [hcup@ahrq.gov](mailto:hcup@ahrq.gov)  
Website: <http://www.ahrq.gov/data/hcup/>

**SID Data and Documentation Distributed by:**  
HCUP Central Distributor

Phone: (866) 556-4287 (toll-free)  
Fax: (866) 792-5313  
E-mail: [HCUPDistributor@AHRQ.gov](mailto:HCUPDistributor@AHRQ.gov)



## Table of Contents

<b>SUMMARY OF DATA USE LIMITATIONS</b> .....	<b>1</b>
<b>HCUP CONTACT INFORMATION</b> .....	<b>2</b>
<b>ABSTRACT</b> .....	<b>3</b>
<b>INTRODUCTION TO THE HCUP STATE INPATIENT DATABASES (SID)</b> .....	<b>4</b>
<b>OVERVIEW OF THE SID</b> .....	<b>4</b>
How the HCUP SID Differ from State Data Files.....	5
What Types of Hospitals Are Included in the SID? .....	5
How to Identify Hospitals in the SID .....	6
What is the File Structure of the SID in the 2005-2010 Files? .....	6
What is the File Structure of the SID in the 1998-2004 Files? .....	7
What is the File Structure of the SID in the 1995-1997 Files? .....	9
What is the File Structure of the SID in the 1990-1994 Files? .....	9
<b>GETTING STARTED</b> .....	<b>11</b>
SID Data Files.....	11
SID Programs, Documentation and Tools .....	11
<b>OTHER HCUP PRODUCTS</b> .....	<b>13</b>
<b>DATABASES</b> .....	<b>13</b>
<b>HCUP CENTRAL DISTRIBUTOR</b> .....	<b>13</b>
<b>HCUP USER SUPPORT</b> .....	<b>14</b>
<b>HCUPnet</b> .....	<b>14</b>
<b>TOOLS</b> .....	<b>14</b>
<b>PUBLICATIONS</b> .....	<b>15</b>
<b>DATA USE AGREEMENT FOR STATE INPATIENT DATABASES</b> .....	<b>17</b>



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

\*\*\*\*\* REMINDER \*\*\*\*\*

**All users of the SID must sign a data use agreement. The signed data use agreements must be kept on file by the organization that purchased the SID data. †**

Authorized users of HCUP data agree to the following limitations: ‡

- Will not use the data for any purpose other than research or aggregate statistical reporting.
- Will not re-release any data to unauthorized users.
- Will not identify or attempt to identify any individual.
- Will not link HCUP data to data from another source that identifies individuals.
- Will not report information that could identify individual establishments (e.g., hospitals).
- Will not use the data concerning individual establishments for commercial or competitive purposes involving those establishments.
- Will not use the data to determine rights, benefits, or privileges of individual establishments.
- Will not identify or attempt to identify any establishment when its identity has been concealed on the database.
- Will not contact establishments included in the data.
- Will not attribute to data contributors any conclusions drawn from the data.
- Must acknowledge the "Healthcare Cost and Utilization Project, (HCUP)", as described in the Data Use Agreement, in reports.

Any violation of the limitations in the data use agreement is punishable under Federal law by a fine of up to \$10,000 and up to 5 years in prison. Violations may also be subject to penalties under State statutes.

† A copy of the Data Use Agreement is included at the end of this document and is also available online at the HCUP User Support Website: <http://www.hcup-us.ahrq.gov>. See next page for AHRQ's address.

‡ Specific provisions are detailed in the Data Use Agreement for HCUP State Inpatient Databases.

## HCUP CONTACT INFORMATION

The SID Data Use Agreement Training Tool and the Data Use Agreement are available on the AHRQ-sponsored HCUP User Support (HCUP-US) Website:

<http://www.hcup-us.ahrq.gov>

**After completing the on-line training tool, please submit signed data use agreements to HCUP at:**

Agency for Healthcare Research and Quality  
Healthcare Cost and Utilization Project (HCUP)  
540 Gaither Road, 5<sup>th</sup> Floor  
Rockville, MD 20850

Phone: (301) 427-1410

Fax: (301) 427-1430

Website: <http://www.ahrq.gov/data/hcup>

**For technical assistance,**

Visit the AHRQ-sponsored HCUP User Support Website at

<http://www.hcup-us.ahrq.gov>

Or send an e-mail to HCUP User Support at

[hcup@ahrq.gov](mailto:hcup@ahrq.gov)

Or contact the HCUP Central Distributor at

HCUP Central Distributor  
Phone: (866) 556-4287 (toll-free between the hours of 9 a.m. and 5 p.m. (ET). If the HCUP Central Distributor is not immediately available, please leave a message on the voice mail, and your call will be returned within one business day.)

Fax: (866) 792-5313

E-mail: [HCUPDistributor@AHRQ.gov](mailto:HCUPDistributor@AHRQ.gov)

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

Our Internet address for user feedback is

[hcup@ahrq.gov](mailto: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  
purchasing 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), formerly the Agency for Health Care Policy and Research.

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 in participating States, translated into a uniform format to facilitate multi-state comparisons and analyses.
- Together, the SID encompass almost 90 percent of all U.S. hospitals 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 patients, regardless of payer, including persons covered by Medicare, Medicaid, private insurance, and the uninsured.
- 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.

Twenty-seven of the Data Organizations participating in the HCUP have agreed to release their SID files through the HCUP Central Distributor under the auspices of the AHRQ. The individual state databases are in the same HCUP uniform format and represent 100% 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 1990-2010 SID.

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

Access to the SID is available through the HCUP Central Distributor. 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 Data Organizations in 44 participating States. 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.

Twenty-seven of the Data Organizations participating in the HCUP have agreed to release their state-specific files through a Central Distributor under the auspices of the AHRQ. The individual state databases are in the same HCUP uniform format and represent 100% 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 ASCII format on CD-ROM.
- Patient-level hospital discharge abstract data for 100 percent of discharges from hospitals in participating States.
- AHA Linkage File to link the SID to data from the American Hospital Association Annual Survey of Hospitals. This is only available for those states that allow the release of hospital identifiers.

SID Documentation and tools, including file specifications, programming source code for loading ASCII data into SAS and SPSS, and value labels, are available online at the HCUP User Support Website: <http://www.hcup-us.ahrq.gov>.

Starting with the 2006 SID, the AHA Linkage files will be available via the HCUP User Support Website: <http://www.hcup-us.ahrq.gov>. The AHA Linkage files may not be available as soon as the discharge-level databases.

## 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 and
- 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?

What types of hospitals are included in the SID depends 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 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 American Hospital Association (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, ear nose throat, short-term rehabilitation, orthopedic, and pediatric hospitals. Non-community hospitals include federal hospitals (Veterans Administration, Department of Defense, and Indian Health Service hospitals), long-term hospitals, psychiatric hospitals, alcohol/chemical dependency treatment facilities and hospitals 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: <http://www.hcup-us.ahrq.gov>. The tables breakdown the number of hospitals by:

- The number of community hospitals
- The number of hospitals not classified as community
- The number of 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 non-community hospitals because these hospitals could not be matched with AHA information. This occurs when a hospital 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 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 2005-2010 Files?

Based on the availability of data elements across states, data elements included in the 2005-2010 SID are structured as follows:

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

The **Core file** contains:

- core data elements that form the nucleus of the SID, and
- 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, or
- are needed for day-to-day 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, or
- are not needed for day-to-day applications.

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: revenue center (REVCN), total charge for the revenue center (CHGN), and total units of service for the revenue center (UNITN). For example, if a patient had 5 laboratory tests, REVCN1 would include the revenue code for laboratory, CHGN1 would include the total charge for the 5 tests, and UNITN1 would be 5. 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 5 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 5 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: revenue center (REVCODE), charge (CHARGE), unit of service (UNITS), and possibly day of service (SERVDAY). For example, if a patient had 5 laboratory tests, there are 5 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.

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. Since 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.

Starting with the 2006 SID, the AHA Linkage files will be available via the HCUP User Support Website: <http://www.hcup-us.ahrq.gov>. The AHA Linkage files may not be available as soon as the discharge-level databases.

**Diagnosis and Procedure Groups Files:** These discharge-level files contain data elements from AHRQ software tools designed to facilitate the use of the 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 Files:** These discharge-level files contain 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.

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 1 record for DSHOSPID "A".

### 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, and
- AHA Linkage file.

The **Core file** contains:

- core data elements that form the nucleus of the SID, and
- 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, or
- are needed for day-to-day 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, or
- are not needed for day-to-day 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: revenue center (REVCDn), total charge for the revenue center (CHGn), and total units of service for the revenue center (UNITn). For example, if a patient had 5 laboratory tests, REVCD1 would include the revenue code for laboratory, CHG1 would include the total charge for the 5 tests, and UNIT1 would be 5. 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 5 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 5 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.

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. Since 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 1 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, and
- 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, or
- are needed for day-to-day 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, or
- are not needed for day-to-day applications.

The **AHA Linkage file** contains AHA linkage data elements that allow the SID to be used in conjunction with the American Hospital Association (AHA) Annual Survey of Hospitals data files. These files contain information about hospital characteristics and are available for purchase through the AHA. Since 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 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 AHA Linkage file contains only 1 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, and
- AHA Linkage file.

The **Core file** contains:

- core data elements that form the nucleus of the SID, and
- 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, or
- are needed for day-to-day 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, or
- are not needed for day-to-day 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: revenue center (REVCDn), total charge for the revenue center (CHGn), and total units of service for the revenue center (UNITn). For example, if a patient had 5 laboratory tests, REVCD1 would include the revenue code for laboratory, CHG1 would include the total charge for the 5 tests, and UNIT1 would be 5. 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. The Maryland SEDD includes this type of Charges file.
- 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 5 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 5 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.

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. Since 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 1 record for DSHOSPID "A".

## GETTING STARTED

**SID Data Files** are provided on CD-ROMs. The number of CD-ROMs depends on the state and year of data.

**SID Programs, Documentation and Tools** for all states and all years are available online at the HCUP User Support Website at <http://www.hcup-us.ahrq.gov>.

### SID Data Files

To load SID data onto your PC, you will need between one and four gigabytes of space available, depending on which SID database you are using. Because of the size of the files, the data are distributed as self-extracting PKZIP compressed files. To decompress the data, you should follow these steps:

1. Create a directory for the state-specific SID on your hard drive.
2. Copy the self-extracting data files from the SID Data Files CD-ROM(s) into the new directory.
3. Unzip each file by running the corresponding \*.exe file:

Type the file name within DOS or click on the name within Windows Explorer.

Edit the name of the "Unzip To Folder" in the WinZip Self-Extractor dialog to select the desired destination directory for the extracted file.

Click on the "Unzip" button.

The ASCII data files will then be uncompressed into the selected destination directory. After the files are uncompressed, the \*.exe files can be deleted.

### SID Programs, Documentation and Tools

The SID programs, technical documentation and HCUP tools available online via the Databases page at the HCUP User Support Website (<http://www.hcup-us.ahrq.gov/databases.jsp>) provide important resources for SID users, and all of the files may be downloaded free of charge.

The SID programs include SAS-load and SPSS-load programs containing the programming code necessary to convert SID ASCII files into SAS or SPSS.

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

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

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

<p><b>Restrictions on the Use of the SID</b></p> <ul style="list-style-type: none"> <li>• SID Data Use Agreement</li> <li>• Requirements for Publishing with HCUP Data</li> </ul> <p><b>Description of the SID Files</b></p> <ul style="list-style-type: none"> <li>• Introduction to the SID – <i>this document</i></li> <li>• HCUP Quality Control Procedures – describes procedures used to assess data quality</li> <li>• File Composition—describes types of hospitals and types of records included in each SASD             <ul style="list-style-type: none"> <li>▪ Number of Discharges by Year</li> <li>▪ Number of Hospitals by Year</li> </ul> </li> <li>• File Specifications – details data file names, number of records, record length, and record layout             <ul style="list-style-type: none"> <li>▪ File Size by Year</li> </ul> </li> </ul> <p><b>Availability of Data Elements</b></p> <ul style="list-style-type: none"> <li>• Availability of SID data elements by Year</li> </ul> <p><b>Description of Data Elements in the SID</b></p> <ul style="list-style-type: none"> <li>• Description of Data Elements – details uniform coding and state-specific idiosyncrasies</li> <li>• Summary Statistics – lists means and frequencies on nearly all data elements</li> <li>• HCUP Coding Practices – describes how HCUP data elements are coded</li> <li>• HCUP Hospital Identifiers – explains data elements that characterize individual hospitals</li> </ul>	<p><b>Corrections to the SID</b></p> <ul style="list-style-type: none"> <li>• Maryland SID, 2006-2009</li> <li>• Nevada SID, 2009</li> <li>• North Carolina SID, 2007-2008</li> <li>• Washington SID, 2003-2006</li> </ul> <p><b>SAS and SPSS Programs</b></p> <p>Programs to load the ASCII data files into statistical software:</p> <ul style="list-style-type: none"> <li>• SAS Load Programs</li> <li>• SPSS Load Programs</li> </ul> <p><b>HCUP Tools: Labels and Formats</b></p> <ul style="list-style-type: none"> <li>• Overview of Clinical Classifications Software (CCS), a categorization scheme that groups ICD-9-CM diagnosis and procedure codes into mutually exclusive categories</li> <li>• Labels for Diagnosis Related Groups (DRG) and Major Diagnostic Categories (MDC)</li> <li>• SAS Format Library Program creates formats to label all HCUP categorical data elements</li> <li>• Labels for CCS diagnosis and procedure categories</li> <li>• Labels for ICD-9-CM Diagnoses and Procedures</li> <li>• Severity Format Program creates SAS formats to label the values data elements in the Severity File</li> </ul> <p><b>SID Related Reports</b></p> <ul style="list-style-type: none"> <li>• SID Related Reports</li> </ul> <p><b>HCUP Supplemental Files</b></p> <ul style="list-style-type: none"> <li>• American Hospital Association Linkage Files</li> <li>• Cost-to-Charge Ratio Files</li> <li>• Hospital Market Structure (HMS) Files</li> </ul>
--	---

## OTHER HCUP PRODUCTS

Information on HCUP products and services is available on the AHRQ-sponsored HCUP User Support Website at <http://www.hcup-us.ahrq.gov>.

### DATABASES

For more information on all HCUP databases, visit the HCUP-US Website at <http://www.hcup-us.ahrq.gov> or contact the HCUP Central Distributor (detailed below).

**Nationwide Inpatient Sample (NIS)** is a nationwide database of hospital inpatient stays. The NIS is the largest all-payer inpatient care database that is publicly available in the United States, containing data from 5 to 8 million hospital stays from about 1,000 hospitals sampled to approximate a 20-percent stratified sample of U.S. community hospitals. The NIS has been available since 1988. For trends analysis, it is recommended that analyses begin with the 1993 data year.

**Nationwide Emergency Department Sample (NEDS)** is a unique and powerful database that yields national estimates of emergency department (ED) visits. The NEDS is the largest all-payer ED database in the United States, containing almost 26 million (unweighted) records for ED visits for over 950 hospitals sampled to approximate a 20-percent stratified sample of U.S. hospital-based EDs.

**State Inpatient Databases (SID)** are hospital inpatient databases from Data Organizations participating in HCUP. The SID contain the universe of the inpatient discharge abstracts in the participating HCUP states, translated into a uniform format to facilitate multi-State comparisons and analyses.

**State Ambulatory Surgery Databases (SASD)** are outpatient databases from Data Organizations in participating HCUP States; these databases capture surgeries performed on the same day in which patients are admitted and released. The SASD contain the ambulatory surgery encounter abstracts in participating States, translated into a uniform format to facilitate multi-state comparisons and analyses. All of the databases include abstracts from hospital-affiliated ambulatory surgery sites. Some contain the universe of ambulatory surgery encounter abstracts for that state, including records from both hospital-affiliated and freestanding surgery centers. Composition and completeness of data files may vary from state to state.

**The State Emergency Department Databases (SEDD)** include data on all emergency department visits that do not result in an admission from Data Organizations in participating HCUP states that provide ED data. Information on patients initially seen in the emergency room and then admitted to the hospital is included in the SID. All of the databases include abstracts from hospital-affiliated emergency department sites. Composition and completeness of data files may vary from state to state.

**Kids' Inpatient Database (KID)** is a unique database of hospital inpatient stays for children. The KID has been produced every three years since 1997 and was specifically designed to permit researchers to study a broad range of conditions and procedures related to child health issues.

### HCUP CENTRAL DISTRIBUTOR

HCUP databases are available for purchase through the AHRQ-sponsored HCUP Central Distributor. All years of the NIS, NEDS, and KID are released through the HCUP Central Distributor. In addition, many of the HCUP State Partners allow the public release of the HCUP SID, SASD, and SEDD through the HCUP Central Distributor. Application Kits for purchasing the HCUP databases are available online at <http://www.hcup-us.ahrq.gov> or by contacting the HCUP

Central Distributor directly. Information on how to obtain uniformly-formatted HCUP files from states not participating in the HCUP Central Distributor is also available from the HCUP Central Distributor:

HCUP Central Distributor  
Phone: (866) 556-4287 (toll-free)  
FAX: (866) 792-5313  
E-mail: [HCUPDistributor@ahrq.gov](mailto:HCUPDistributor@ahrq.gov)

## HCUP USER SUPPORT

HCUP User Support (HCUP-US) provides technical assistance to all HCUP users and is designed to facilitate the use of HCUP data, software tools, and products. The goals of this service are to increase awareness of the strengths and uses of HCUP data and to enhance the skills of individuals using the data for research, education, and policy analysis. A user-friendly Website for HCUP-US is located at <http://www.hcup-us.ahrq.gov>. This site includes links to information on how to purchase and understand the HCUP databases, as well as links to HCUP User Support Services and an index of HCUP topics. For further information, consultants are available via both telephone and e-mail to help in planning analytic research and to offer advice about appropriate uses of HCUP data.

## HCUPnet

HCUPnet is a Web-based query tool for identifying, tracking, analyzing, and comparing statistics on hospitals at the national, regional, and state level. HCUPnet offers easy access to national statistics and trends and selected state statistics about hospital stays. This tool provides step-by-step guidance, helping researchers to quickly obtain the statistics they need. HCUPnet generates statistics using the NIS, KID, and SID for those states that have agreed to participate. In addition, HCUPnet provides Quick Statistics – ready-to-use tables on commonly requested information – as well as national statistics based on the AHRQ Quality Indicators. HCUPnet can be found at: <http://hcupnet.ahrq.gov/>.

## TOOLS

**AHRQ Quality Indicators (QIs)** are clinical performance measures for use with readily available inpatient data. Methods and software for the AHRQ Quality Indicators can be downloaded from <http://www.qualityindicators.ahrq.gov>.

The following tools can all be found at the HCUP User Support Website, Tools and Software page, at [http://www.hcup-us.ahrq.gov/tools\\_software.jsp](http://www.hcup-us.ahrq.gov/tools_software.jsp). Methods and software related to these products can be downloaded from the same Web page.

**Clinical Classifications Software (CCS)**, formerly known as the Clinical Classifications for Health Policy Research (CCHPRs), are classification systems that group ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification) diagnoses and procedures into a limited number of clinically meaningful categories. CCS is also available for ICD-10 diagnoses, Current Procedural Terminology (CPT) and HCPCS procedures, in addition to mental health and substance abuse-related ICD-9-CM diagnoses.

**Comorbidity Software** assigns variables that identify comorbidities in hospital discharge records using ICD-9-CM diagnosis codes.

**Procedure Classes** identify whether a procedure is (a) diagnostic or therapeutic, and (b) minor or major in terms of invasiveness and/or resource use.

**Cost-to-Charge Ratio (CCR) Files** are hospital-level files designed to supplement the data elements in the NIS and SID databases.

**Chronic Condition Indicator** provides users an easy way to categorize ICD-9-CM diagnosis codes into one of two categories: chronic or not chronic. The tool can also assign ICD-9-CM diagnosis codes into 1 of 18 body system categories.

**Utilization Flags** reveal additional information about use of health care services by combining information from UB-92 revenue codes and ICD-9-CM procedure codes to create flags, or indicators, of utilization. Use of procedures and services such as ICU, CCU, NICU, and specific diagnostic tests and therapies can be assessed with these Utilization Flags.

## **PUBLICATIONS**

Publications using HCUP data or describing methods for using HCUP data can be found at: <http://www.hcup-us.ahrq.gov/reports.jsp>.

**HCUP Fact Books** report aggregate statistics and detailed analyses using HCUP data. The Fact Books can be viewed online or can be requested from the AHRQ Publications Clearinghouse at (800) 358-9295. You can also send a postcard requesting these reports by writing to: AHRQ Publications Clearinghouse, P.O. Box 8547, Silver Spring, MD 20907.

**HCUP Statistical Briefs** are Web-based reports that present simple, descriptive statistics on a variety of focused topics such as hospital admissions through the emergency department, hospitalizations among the uninsured, women and heart disease, hospital stays associated with alcohol abuse, and racial and ethnic disparities in potentially preventable hospitalizations.

**HCUP Methods Series** features a broad array of methodological reports on the HCUP databases and software tools. Topics range from how to use the NIS for reporting trends, how to properly calculate variance estimates using the NIS, an evaluation of linking patients across hospital stays in the SID, evaluations of HCUP emergency department and ambulatory surgery data, an evaluation of E code reporting across the HCUP States, and creation of utilization flags based on UB-92 revenue codes.

**HCUP Database Reports** are specific to the design and use of the HCUP databases. These reports include descriptions of the design of each database, comparisons of HCUP data with other data sources, evaluations of data quality, and descriptions of database composition.

**New Findings and Publications** based on HCUP data are available at the HCUP User Support Website at <http://www.hcup-us.ahrq.gov/reports/pubsearch/pubsearch.jsp>.





## **DATA USE AGREEMENT for the State Inpatient Databases from the Healthcare Cost and Utilization Project Agency for Healthcare Research and Quality**

This Data Use Agreement (“Agreement”) implements the data protections of the Health Insurance Portability and Accountability Act (HIPAA) of 1996 (Public Law 104-191) and the Agency for Healthcare Research and Quality (AHRQ) confidentiality statute. Any individual (“data recipient”) seeking to obtain or use data in the State Inpatient Databases (SID) from the Healthcare Cost and Utilization Project (HCUP) maintained by the Center for Delivery, Organization, and Markets (CDOM) within AHRQ, must sign and submit this Agreement to AHRQ or its agent before access to the SID may be granted.

In accordance with HIPAA, the SID may only be used or disclosed in the form of a *limited data set*, as defined by the HIPAA Privacy Rule (45 CFR § 164.514(e)).

The AHRQ confidentiality statute, Section 924(c) of the Public Health Service Act (42 U.S.C. 299c-3(c)), requires that data collected by AHRQ that identify individuals or establishments be used only for the purpose for which they were supplied. Data supplied to AHRQ for HCUP and disclosed in limited data set form are identifiable under the HIPAA Privacy Rule and are provided by the data sources only for research, analysis, and aggregate statistical reporting. Therefore, data recipients may use HCUP data only for these purposes.

*No Identification of Persons*—Any effort to determine the identity of any person contained in HCUP databases (including but not limited to patients, physicians, and other health care providers), or to use the information for any purpose other than for research, analysis, and aggregate statistical reporting, would violate the AHRQ confidentiality statute, the conditions of this Agreement, and the HIPAA Privacy Rule. Recipients of the data set are prohibited under the AHRQ confidentiality statute and the terms of this Agreement from releasing, disclosing, publishing, or presenting any individually identifying information obtained under this Agreement. AHRQ omits from the data set all direct identifiers that are required to be excluded from limited data sets as defined by the HIPAA Privacy Rule. It may be possible in limited situations, through deliberate technical analysis, and with outside information, to ascertain from the limited data sets the identity of particular persons. Considerable harm could ensue if this were to occur. Therefore, any attempts to identify individuals are prohibited and information that could identify individuals directly or by inference must not be released or published. In addition, users of the data must not attempt to contact individuals for any purpose, including verifying information supplied in the data set. Any questions about the data must be referred exclusively to AHRQ.

*Use of Establishment Identifiers*—Section 924(c) of the Public Health Service Act (42 U.S.C. 299c-3(c)) also restricts the use of any information that permits the identification of establishments for purposes other than those for which the information was originally supplied. Permission is obtained from the HCUP data sources (state data organizations, hospital associations, and data consortia) to use the identification of hospitals (when such identification appears in the data sets) for research, analysis, and aggregate statistical reporting. This may include linking institutional information from outside data sets for these purposes. Such purpose does *not* include the use of information in the data sets concerning individual establishments for commercial or competitive purposes involving those individual establishments, or to determine the rights, benefits, or privileges of establishments. Users of the data must not identify establishments directly or by inference in disseminated material. In addition, users of the data must not contact establishments for the purpose of verifying information supplied in the data set. Any questions about the data must be referred exclusively to AHRQ. Misuse of identifiable HCUP data about hospitals would violate the AHRQ confidentiality statute and trigger its penalty provisions.

**The undersigned gives the following assurances with respect to the SID data set:**

- I will not use and will prohibit others from using or disclosing the data set (or any part), except for research, analysis, and aggregate statistical reporting, and only as permitted by this Agreement.
- I will ensure that the data are kept in a secured environment and that only authorized users will have access to the data.
- I will not release or disclose, and will prohibit others from releasing or disclosing, any data that are individually identifiable under the HIPAA Privacy Rule, or any information that identifies persons, directly or indirectly, except as permitted under this Agreement and in accordance with the above-mentioned AHRQ confidentiality statute.
- I will not release or disclose information where the number of observations (i.e., individual discharge records) in any given cell of tabulated data is less than or equal to 10.
- I will not release or disclose, and will prohibit others from releasing or disclosing, the data set (or any part) to any person who is not a member, agent, or contractor of the organization (specified below), except with the approval of AHRQ.
- I will require others employed in my organization (specified below), and any agents or contractors of my organization, who will use or will have access to the data set, to sign a copy of this Agreement (specifically acknowledging their agreement to abide by its terms) and I will submit those signed Agreements to AHRQ or its agent before granting access.
- I will not attempt to link, and will prohibit others from attempting to link, the discharge records of persons in the data set with individually identifiable records from any other source.
- I will not attempt to use and will prohibit others from using the data set to learn the identity of any person included in the data set or to contact any such person for any purpose.
- In accordance with the AHRQ confidentiality statute, I will not use and will prohibit others from using the data set concerning individual establishments (1) for commercial or competitive purposes involving those individual establishments; (2) to determine the rights, benefits, or privileges of individual establishments; or (3) to report, through any medium, data that could identify, directly or by inference, individual establishments.
- When the identities of establishments are not provided in the data sets, I will not attempt to use and will prohibit others from using the data set to learn the identity of any establishment.
- I will not contact and will prohibit others from contacting establishments or persons in the data set to question, verify, or discuss data in the HCUP databases.
- I will indemnify, defend, and hold harmless AHRQ and the data organizations that provide data to AHRQ for HCUP from any or all claims and losses accruing to any person, organization, or other legal entity as a result of violation of this Agreement. This provision applies only to the extent permitted by Federal and State law.
- I will make no statement and will prohibit others from making statements indicating or suggesting that interpretations drawn are those of the data sources or AHRQ.
- I will provide an abstract and reference for any published research material resulting from the use of these HCUP State Inpatient Databases to the HCUP Central Distributor.

- I will acknowledge in all reports based on these data that the source of the data is the specific state(s) or data organization(s) that submitted data to the HCUP (e.g., “*state name(s)*, State Inpatient Databases (SID), Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality.”

*Safeguards.* I agree to use appropriate safeguards to prevent use or disclosure of the data set other than as permitted by this Agreement.

*Permitted Access to Limited Data Set.* I shall limit the use or receipt of the data set to the individuals who require access in order to perform activities permitted by this Agreement. This Agreement must be signed by all such individuals and submitted to AHRQ or its agent before access to the data set may be granted.

*Re-disclosure.* I will not re-disclose (i.e., share) the data set (or any part), unless the individual who will receive the data has agreed in writing to be bound by the same restrictions and conditions that apply to me under this Agreement.

*The HIPAA Privacy Rule.* I agree not to use or disclose the data set in any manner that would violate the HIPAA Privacy Rule if I were a covered entity under the Privacy Rule.

*Agents and Contractors.* I shall ensure that any agents, including contractors and subcontractors to whom I provide the data set, agree in writing to be bound by the same restrictions and conditions that apply to me with respect to the limited data set.

*Reporting Violations of this Agreement.* I agree to report any violations to AHRQ within twenty-four (24) hours of becoming aware of any use or disclosure of the limited data set in violation of this Agreement or applicable law.

*Term, Breach, and Termination of this Agreement.* This Agreement shall continue in full effect until the data recipient has returned all copies of the data set to AHRQ. Any noncompliance by the data recipient with the terms of this Agreement will be grounds for immediate termination of the Agreement if, at the sole determination of AHRQ, the data recipient knew or should have known of such noncompliance and failed to immediately take reasonable steps to remedy the noncompliance.

*Reporting to the United States Department of Health and Human Services.* If the data recipient fails to remedy any breach or violation of this Agreement to the satisfaction of AHRQ, and if termination of the Agreement is not feasible, AHRQ shall report the recipient’s breach or violation to the Secretary of the United States Department of Health and Human Services, and the recipient agrees that he or she shall not have or make any claims against AHRQ with respect to such report(s).

I understand that this Agreement is requested by the United States Agency for Healthcare Research and Quality to ensure compliance with its statutory confidentiality requirement. My signature indicates my Agreement to comply with the above-stated requirements with the knowledge that any violation of the AHRQ confidentiality statute is subject to a civil penalty of up to \$10,000 under 42 U.S.C. 299c-3(d), and that deliberately making a false statement about this or any matter within the jurisdiction of any department or agency of the Federal Government violates 18 U.S.C. 1001 and is punishable by a fine of up to \$10,000 or up to five years in prison. Violators of this Agreement may also be subject to penalties under state confidentiality statutes that apply to these data for particular states.

Signed: \_\_\_\_\_ Date: \_\_\_\_\_

Print or Type Name of Data Recipient: \_\_\_\_\_

Title: \_\_\_\_\_

Organization: \_\_\_\_\_

Address: \_\_\_\_\_

City: \_\_\_\_\_ State: \_\_\_\_\_ ZIP Code: \_\_\_\_\_

Phone Number: \_\_\_\_\_ Fax: \_\_\_\_\_

E-mail: \_\_\_\_\_

The information above is maintained by AHRQ for the purpose of enforcement of this Agreement. This information may also be used by AHRQ to create an HCUP mailing list. The mailing list allows AHRQ to send users information such as notices about the release of new databases and errata when data errors are discovered.

**Please include me on the HCUP mailing list.**

Note to Purchaser: Shipment of the requested data product will only be made to the person who signs this Agreement, unless special arrangements that safeguard the data are made with AHRQ or its agent.

Agency for Healthcare Research and Quality  
HCUP Project Officer  
540 Gaither Road  
Rockville, Maryland 20850  
<http://www.hcup-us.ahrq.gov/home.jsp>