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 KIDS' INPATIENT DATABASE (KID)

2022

Please read all documentation carefully.

BEGINNING WITH DATA YEAR 2016, THE KID CONTAINS ICD-10-CM/PCS CODES.

Beginning with data year 2016, the KID includes diagnosis and procedure codes reported using the ICD-10-CM/PCS[1] coding system.

Data elements derived from AHRQ software tools are not available for ICD-10-CM/PCS data on the 2016 KID.

The KID ENHANCED CONFIDENTIALITY beginning with 2012.

Starting in data year 2012, the KID eliminated State identifiers, unencrypted hospital identifiers, and other data elements that are not uniformly available across States.

These pages provide only an introduction to the KID 2022 package.

For full documentation and notification of changes, visit the HCUP User Support (HCUP-US) Web site at
www.hcup-us.ahrq.gov.

Issued December 2024
Agency for Healthcare Research and Quality
Healthcare Cost and Utilization Project (HCUP)
Phone: (866) 290-HCUP (4287)
Email: hcup@ahrq.gov
Website: www.hcup-us.ahrq.gov

KID Data and Documentation Distributed by:
HCUP Central Distributor
Phone: (866) 290-HCUP (4287) (toll free)
Email: hcup@ahrq.gov

Table of Contents

HCUP KIDS' INPATIENT DATABASE (KID)
SUMMARY OF DATA USE RESTRICTIONS

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

All users of the KID must take the online HCUP 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, 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 not use data elements from the proprietary severity adjustment software packages (e.g., 3M™ APR-DRGs) for any commercial purpose or to disassemble, decompile, or otherwise reverse engineer the proprietary software.
- 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 Nationwide 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 online DUA training course is available at: www.hcup-us.ahrq.gov/tech_assist/dua.jsp.

The **HCUP Data Use Agreement for the KID** is available on the HCUP User Support (HCUP-US) website at: https://www.hcup-us.ahrq.gov/team/NationwideDUA.jsp.

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: 866-290-4287 (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.

WHAT'S NEW IN THE 2022 KIDS' INPATIENT DATABASE (KID)?

- The number of discharges in the 2022 KID decreased by 2.6 percent from the number of discharges in the 2019 KID.
- In data year 2022, Nevada data were not available for inclusion in the KID.
- There are four additional data elements based on the following new categories for the Clinical Classifications Software Refined (CCSR) for ICD-10-CM/PCS:
  - DXCCSR_DEN001 *Any dental condition including traumatic injury*
  - DXCCSR_DEN001 *Nontraumatic dental conditions*
  - DXCCSR_DEN001 *Caries, periodontitis, and other preventable dental conditions*
  - PRCCSR_ADM022 *COVID-19 vaccinations.**

* The data element DXCCSR_INF012 *Coronavirus disease – 2019 (COVID-19)* was included in the 2019 KID, but no records were flagged by this data element.

UNDERSTANDING THE KID

- This document, *Introduction to the KID, 2022* summarizes the content of the KID and describes the development of the KID sample and weights.
- In addition, the HCUP-US website has a section on ICD-10-CM/PCS Resources that summarizes key issues for researchers using HCUP and other administrative databases that include ICD-10-CM/PCS coding. The web page provides general guidance and forewarning to users analyzing outcomes that may be affected by the transition to the ICD-10-CM/PCS coding system and lists other related web resources.
- Important considerations for data analysis are provided along with references to detailed reports.
- In-depth documentation for the KID is available on the HCUP User Support (HCUP-US) website (https://www.hcup-us.ahrq.gov/db/nation/kid/kiddbdocumentation.jsp). Please refer to detailed documentation before using the data.

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 Kids' Inpatient Database (KID).

HCUP Kids' Inpatient Database (KID)

ABSTRACT

The Kids' Inpatient Database (KID) is part of the Healthcare Cost and Utilization Project (HCUP), sponsored by the Agency for Healthcare Research and Quality (AHRQ).

The KID is the largest publicly available all-payer pediatric inpatient care database in the United States, yielding national estimates of hospital inpatient stays by children. The KID is a sample of pediatric discharges from all community, non-rehabilitation hospitals in States participating in HCUP. The target universe includes pediatric discharges from community, non-rehabilitation hospitals in the United States. Pediatric discharges are defined as all discharges where the patient was age 20 or less at admission. See Table 1 in Appendix I for a list of the statewide data organizations participating in the KID. The number of sample hospitals and discharges by State and year are available in Table 2 in Appendix I.

Inpatient stay records in the KID include clinical and resource use information typically available from discharge abstracts created by hospitals for billing. The KID contains charge information on all patients, regardless of the expected payer for the hospital stay. The KID's large sample size enables analyses of rare conditions, such as congenital anomalies and uncommon treatments, such as cardiac surgery. It can be used to study a wide range of topics including the economic burden of pediatric conditions, access to services, quality of care and patient safety, and the impact of health policy changes. Discharge weights are provided for calculating national estimates.

Key features of the most recent KID database year (2022) include:

- A sample of 10 percent of normal newborns and 80 percent of other pediatric discharges (age 20 or less at admission) from all U.S. community hospitals (defined as short-term, non-Federal, general and specialty hospitals, excluding hospital units of other institutions), excluding rehabilitation hospitals, in States participating in HCUP
- A large sample size, which enables analyses of rare conditions (e.g. congenital anomalies) as well as, uncommon treatments (e.g. cardiac surgery)
- Pediatric discharges for all types of expected payers, including Medicaid, private insurance, as well as self-pay and no charge

Beginning with 2012 data, the Kids' Inpatient Database (KID) incorporated the following changes to enhance confidentiality:

- The 2012 KID eliminated State identifiers, unencrypted hospital identifiers, and other data elements that were not uniformly available across States.

- HOSP_KID is now the encrypted hospital number. Prior to the 2012 KID, the HCUP hospital identifier (HOSPID) provided the linkage between the KID Inpatient Core files and the Hospital Weights file. Beginning with the 2012 KID, the KID hospital number (HOSP_KID) provides the linkage between the KID Inpatient Core files and the Hospital Weights file. The HOSP_KID values are reassigned each year, so they cannot be used to link hospitals across years.

- The 2012 KID no longer includes hospital identifiers that can be linked to other hospital-level data (such as the AHA Annual Survey).

- KID_STRATUM=9999 (or 9998 in 2003) is now the way to identify children's hospitals. Prior to 2012, children's hospitals could be selected using the NACHTYPE data element in the KID. Beginning with 2012 data, NACHTYPE is no longer available, but discharges from freestanding children's hospitals are stratified as children's hospitals (KID_STRATUM=9999 [or 9998 in 2003]).

The KID is available every three years beginning with 1997. Periodically, new data elements are added to the KID and some are dropped; see Appendix III for a summary of data elements and when they are effective.

Hospital discharge data for 2015 contained a mix of ICD-9 and ICD-10 data; the first three quarters of 2015 contained ICD-9-CM data and the last quarter contained ICD-10-CM/PCS. Because of the complexities of analyzing a mixed data year, the KID was not released in 2015 and instead released in 2016. Beginning with 2016 data, the KID includes ICD-10-CM/PCS data only.

Access to the KID is open to users who sign Data Use Agreements. Uses are limited to research and aggregate statistical reporting.

For more information on the KID, visit the AHRQ-sponsored HCUP User Support (HCUP-US) Web site at www.hcup-us.ahrq.gov.

INTRODUCTION TO THE HCUP KIDS' INPATIENT DATABASE (KID)

Overview of KID Data

The Kids' Inpatient Database (KID) is part of the Healthcare Cost and Utilization Project (HCUP), sponsored by the Agency for Healthcare Research and Quality (AHRQ).

The KID is the largest publicly available all-payer pediatric inpatient care database in the United States, yielding national estimates of hospital inpatient stays by children. The KID is a sample of pediatric discharges from all community, non-rehabilitation hospitals in States participating in HCUP. See Table 1 in Appendix I for a list of the statewide data organizations participating in the KID.

Inpatient stay records in the KID include clinical and resource use information typically available from discharge abstracts created by hospitals for billing. The KID contains charge information on all patients, regardless of the expected payer for the hospital stay. The KID's large sample size enables analyses of rare conditions, such as congenital anomalies and uncommon treatments, such as cardiac surgery. It can be used to study a wide range of topics including the economic burden of pediatric conditions, access to services, quality of care and patient safety, and the impact of health policy changes. Discharge weights are provided for calculating national estimates.

The KID target universe includes pediatric discharges from community, non-rehabilitation hospitals in the United States.[1] Pediatric discharges are defined as all discharges where a patient was 20 years or less at admission. Discharges with missing, invalid, or inconsistent ages are excluded. Pediatric discharges are identified as one of three types of records:

- Normal newborns (HOSPBRTH = 1 and UNCBRTH = 1)
- Other newborns (HOSPBRTH = 1 and UNCBRTH = 0)
- All other pediatric cases (HOSPBRTH = 0).

In-hospital births (HOSPBRTH = 1) are identified by any ICD-10-CM principal or secondary diagnosis code of 'Z3800', 'Z3801', 'Z382', 'Z3830', 'Z3831', 'Z385', 'Z3861' 'Z3862', 'Z3863', 'Z3864', 'Z3865', 'Z3866', 'Z3868', 'Z3869', or 'Z388' (previously by any ICD-9-CM principal or secondary diagnosis in the range of V3000 to V3901 with the last two digits of "00" or "01") and the patient is not transferred from another acute care hospital or health care facility. Normal newborns (UNCBRTH = 1) have a Diagnosis Related Group (DRG) indicating "Normal Newborn" (391 prior to 2009, or 795 beginning in 2009).

The KID includes a sample of pediatric discharges from all HCUP hospitals in the sampling frame – the State Inpatient Databases (SID) that agreed to participate in the KID. For sampling,

[1] Community hospitals, as defined by the American Hospital Association (AHA), include "all non-Federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions." Included among community hospitals are specialty hospitals such as obstetrics-gynecology, ear-nose-throat, short-term rehabilitation, orthopedic, and pediatric institutions. Also included are public hospitals and academic medical centers. Starting in 2005, the AHA included long term acute care facilities in the definition of community hospitals. These facilities provide acute care services to patients who need long term hospitalization (stays of more than 25 days). Excluded from the KID are short-term rehabilitation hospitals (beginning with 2000 data), long-term non-acute care hospitals, psychiatric hospitals, and alcoholism/chemical dependency treatment facilities.

pediatric discharges are stratified by normal newborns, other newborns, and all other pediatric cases. To further ensure an accurate representation of each hospital's pediatric case-mix, the discharges are sorted by hospital, DRG, and a random number within each DRG. Systematic random sampling is used to select 10% of normal newborns and 80% of other newborns and pediatric cases from each frame hospital.

To obtain national estimates, discharge weights are developed using the American Hospital Association (AHA) universe of community, non-rehabilitation hospitals as the standard. For the weights, hospitals are post-stratified on six characteristics contained in the AHA hospital files—ownership/control, bed size, teaching status, rural/urban location, U.S. region, and a stratum for freestanding children's hospitals. To create weights, if there were fewer than two frame hospitals, 30 normal newborns, 30 other newborns, and 30 non-birth pediatric discharges sampled in a stratum, that stratum is combined with an "adjacent" stratum containing hospitals with similar characteristics. Discharge weights are created by stratum in proportion to the number of AHA newborns for newborn discharges and in proportion to the total number of (non-newborn) AHA discharges for non-newborn discharges.

Detailed information on the design of the KID prior to 2006 is available in the year-specific special reports on *Design of the Kids' Inpatient Database* found on the HCUP-US Web site (www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp). Starting with the 2006 KID, the information on the design of the KID was incorporated into this report, which describes the KID sample and weights, summarizes the contents of the KID, and discusses data analysis issues. This document highlights cumulative information for all previous KID releases to provide a longitudinal view of the database. Over time, we have enhanced the nationwide representation of the sample by incorporating data from additional HCUP State Partners.

KID data sets are currently available for multiple years. Each release of the KID includes:

- Data in fixed-width ASCII format
- Approximately 3 million pediatric inpatient records per year
- 2,500 to 4,200 hospitals per year (all SID hospitals with pediatric discharges)
- Discharge-level weights to calculate national estimates
- Hospital File with hospital characteristics
- KID Documentation and tools – including file specifications, programming source code for loading ASCII data into SAS and SPSS, and value labels. Beginning in 2006, code is also provided for loading the KID ASCII file into Stata.

ICD-10-CM/PCS Started October 1, 2015 at the Beginning of Fiscal Year 2016

On October 1, 2015, the United States transitioned from using ICD-9-CM to ICD-10-CM/PCS code sets for reporting medical diagnoses and inpatient procedures.[2] ICD-10-CM/PCS consists of two parts:

- ICD-10-CM: diagnosis coding on inpatient and outpatient data
- ICD-10-PCS: procedure coding on inpatient data.

[2] ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification; ICD-10-CM/PCS: International Classification of Diseases, Tenth Revision, Clinical Modification/Procedure Coding System

The HCUP-US Web site has a section on ICD-10-CM/PCS Resources that summarizes key issues for researchers using HCUP and other administrative databases that include ICD-10-CM/PCS coding. The Web page provides general guidance and forewarning to users analyzing outcomes that may be affected by the transition to the ICD-10-CM/PCS coding system and lists other related Web resources.

KID Data Sources, Hospitals, and Inpatient Stays

Table 2 in Appendix I contains a summary of the data sources, number of hospitals, and number of inpatient stays in each KID database. It also lists the differences in types of hospitals and age inclusion for pediatric cases.

State-Specific Restrictions

Some HCUP Partners that contributed data to the KID imposed restrictions on the release of certain data elements or on the number and types of hospitals that could be included in the database. Because of confidentiality laws, some data sources were prohibited from providing HCUP with discharge records that indicated specific medical conditions, such as HIV/AIDS or behavioral health. Detailed information on these State-specific restrictions is available in Appendix II.

Contents of the KID

The KID product is downloaded in a single zipped file for each year which contains several data-related compressed files and accompanying documentation. There are three discharge-level files and one hospital-level file:

Discharge-level files

- **Core File** is a single discharge-level file containing commonly used data elements (e.g., age, expected primary payer, discharge status, ICD-10-CM/PCS codes, total charges).
    - This file is available in all years of the KID.
    - The file is sorted by the KID hospital identifier (HOSP_KID) and the HCUP unique record identifier (RECNUM). These two variables provide the linkage between the different discharge-level files.
- **Severity File** is a single discharge-level file containing additional data elements to aid in identifying the severity of the condition for a specific discharge.
    - This file is available beginning with the 2003 KID.
    - The file is sorted by the KID hospital identifier (HOSP_KID) and the HCUP unique record identifier (RECNUM).
- **Diagnosis and Procedure Groups File** is a single discharge-level file containing additional information based on the ICD-10-CM diagnoses and ICD-10-PCS procedures that is created by the Agency for Healthcare Research and Quality (AHRQ) software tools.
    - This file is available beginning with the 2006 KID.

- For data years 2016, this file was not available in the KID. Data elements derived from the ICD-10-CM/PCS AHRQ software tools were not included in the KID because they were still in development and testing.
- Beginning with data year 2019, this file is available in the KID and includes data elements derived from the Clinical Classifications Software Refined (CCSR) for ICD-10-CM diagnoses, the CCSR for ICD-10-PCS procedures, and Procedure Classes Refined for ICD-10-CM.
- The file is sorted by the KID hospital identifier (HOSP_KID) and the HCUP unique record identifier (RECNUM). These two variables provide the linkage between the different discharge-level files.

Hospital-level files

- **Hospital File** is a single hospital-level file containing information on hospital characteristics.
  - This file is available in all years of the KID.
  - Linkage between the Inpatient Core File and the Hospital File
    - Prior to the 2012 KID, the HCUP hospital identifier (HOSPID) provided the linkage between the KID Inpatient Core File and the Hospital File.
    - Beginning with the 2012 KID, the KID hospital number (HOSP_KID) provides the linkage between the KID Inpatient Core File and the Hospital File. The HOSP_KID values are reassigned each year, so they cannot be used to link hospitals across years.

On the HCUP-US website, KID users can access complete file documentation, including data element notes, file layouts, summary statistics, and related technical reports. Similarly, users can also download SAS, SPSS, and Stata load programs from this website. Available online documentation and supporting files are detailed in Appendix I, Table 3.

KID Data Elements

All releases of the KID contain two types of data: inpatient stay records and hospital information with weights to calculate national estimates. Appendix III identifies the data elements in each KID file:

- Table 1 for the Core File (record = inpatient stay)
- Table 2 for the Hospital File (record = hospital)
- Table 3 for the Severity File (record = inpatient stay)
- Table 4 for the Diagnosis and Procedure Groups File (record = inpatient stay).

Not all data elements in the KID are uniformly coded or available across all States. The tables in Appendix III are not complete documentation for the data. Please refer to the KID documentation located on the HCUP-US website (http://www.hcup-us.ahrq.gov) for comprehensive information about data elements and the files.

Users interested in applying AHRQ software tools to the KID for data years including ICD-10-CM/PCS-coded data to produce data elements currently unavailable in the database files may do so by downloading the respective tool(s) from the Research Tools section of the HCUP User Support (HCUP-US) website. Additionally, users may wish to review the HCUP Software Tools Tutorial, which provides instructions on how to apply the AHRQ software tools to HCUP or other

administrative databases.

Getting Started

Computer Specifications Required for Using the KID

To load and analyze the KID data on a computer, users will need the following:

- A hard drive with *at least* 35 gigabytes of space available
- A third-party zip utility such as ZIP Reader, 7-Zip© , SecureZIP®, WinZip®, or Stuffit Expander®
- SAS®, SPSS®, Stata® or similar analysis software.

Decompressing the KID Files

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

1) Create a directory for the KID on your hard drive.
2) Unzip the compressed KID product file into the new directory using a third-party zip utility. This will create four 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 same password and third-party zip utility method. This will place the data-related ASCII files in this same directory by default.

Downloading and Running the Load Programs

Programs to load the data into SAS, SPSS, or Stata, are available on the HCUP User Support website (HCUP-US). The SAS and SPSS programs are available beginning with 2000. The Stata programs begin with 2006. To download and run the load programs, follow these steps:

1) Go to the KID Database Documentation page on HCUP-US at https://www.hcup-us.ahrq.gov/db/nation/kid/kiddbdocumentation.jsp.

2) Go to the “File Specifications and Load Programs” section on this page.
3) Click on “Nationwide SAS Load Programs”, “Nationwide SPSS Load Programs”, or “Nationwide Stata Load Programs” to go to the corresponding Load Programs page.
4) Select the data year and the database (“KID”) from the drop-down lists on this page. Or you may select “KID Load All Years” to obtain a zipped file with all load programs for multiple years at once.
5) Select and save the load programs you need. **The load programs are specific to the data year.** For example, the load program for the 2022 KID Core File is found under the link “SAS KID 2022 Core File” in the list generated by selecting “2022” and “KID.” Save the load programs into the same directory as the KID ASCII files on your computer.
6) Edit and run the load programs as appropriate for your computing environment to create the analysis files. For example, modify the directory paths to point to the location of your input and output files.

KID Documentation

KID documentation files on the HCUP-US Web site (www.hcup-us.ahrq.gov) provide important resources for the user. Refer to these resources to understand the structure and content of the KID and to aid in using the database.

- To locate the KID documentation on HCUP-US, choose “Databases” from the home page (www.hcup-us.ahrq.gov). The section labeled “KIDS’ Inpatient Database (KID) is specific to the KID.

Table 3 in Appendix I details both the KID related reports and the comprehensive KID documentation available on HCUP-US.

HCUP Online Tutorials

For additional assistance, AHRQ has created the HCUP Online Tutorial Series, a series of free, interactive courses which provide training on technical methods for conducting research with HCUP data. Topics include an HCUP Overview Course and these tutorials:

- The Load and Check HCUP Data tutorial provides instructions on how to unzip (decompress) HCUP data, save it on your computer, and load the data into a standard statistical software package. This tutorial also describes how to verify that the data has loaded correctly.

- The HCUP Sampling Design tutorial is designed to help users learn how to account for sample design in their work with HCUP national (nationwide) databases.

- The Producing National HCUP Estimates tutorial is designed to help users understand how the three national (nationwide) databases – the NIS, Nationwide Emergency Department Sample (NEDS), and Kids' Inpatient Database (KID) – can be used to produce national and regional estimates.

- The Calculating Standard Errors tutorial shows how to accurately determine the precision of the estimates produced from the HCUP nationwide databases. Users will learn two

methods for calculating standard errors for estimates produced from the HCUP national (nationwide) databases.

- The HCUP Multi-year Analysis tutorial presents solutions that may be necessary when conducting analyses that span multiple years of HCUP data.

- The HCUP Software Tools Tutorial provides instructions on how to apply the AHRQ software tools to HCUP or other administrative databases.

New tutorials are added periodically, and existing tutorials are updated when necessary. The Online Tutorial Series is located on the HCUP-US website at www.hcup-us.ahrq.gov/tech_assist/tutorials.jsp.

HOW TO USE THE KID FOR DATA ANALYSIS

This section provides a brief synopsis of special considerations when using the KID. Before reporting findings using the KID, you should refer to the *Checklist for Working with the KID* (www.hcup-us.ahrq.gov/db/nation/kid/kidchecklist.jsp) to verify adherence to data use, methodology, and reporting requirements.

- If anyone other than the original purchaser uses the KID data, be sure to have them read and sign a Data Use Agreement, after viewing the online Data Use Agreement Training Tool available on the HCUP-US Web site (www.hcup-us.ahrq.gov). A copy of the signed Data Use Agreements must be sent to AHRQ. See page 2 for the mailing address.

- The KID contains discharge-level records, not patient-level records. This means that individual patients who are hospitalized multiple times in one year may be present in the KID multiple times. There is no uniform patient identifier available that allows a patient-level analysis with the KID. This will be especially important to remember for certain conditions for which patients may be hospitalized multiple times in a single year.

Choosing Data Elements for Analysis

- For all data elements you plan to use in your analysis, first perform descriptive statistics and examine the range of values, including the number of missing cases. Summary statistics for the entire KID are provided on the Summary Statistics page of the HCUP-US website. Performing descriptive statistics by hospital can be helpful in detecting hospital-specific data anomalies.

- Not all data elements in the KID are provided by each hospital. These data elements are provided on the KID because they can be valuable for research purposes, but they should be used cautiously. For example, RACE is missing for some hospitals; thus, national estimates using RACE should be interpreted and reported with caveats.

- Differences exist across the State data sources in the collection of information that could not be accounted for during HCUP processing to make the data uniform. For example, the most reliable way to identify ED admissions in the HCUP databases is to use the data element HCUP_ED, which considers all possible evidence of ED services. Unfortunately, this information is not always complete.

- ICD-9-CM and ICD-10-CM diagnosis and ICD-10-PCS procedure codes provide valuable insights into the reasons for hospitalization and what procedures patients receive, but these codes need to be carefully used and interpreted. ICD-9-CM and ICD-10-CM/PCS codes change every October as new codes are introduced, and some codes are retired. It is critical to check all ICD-9-CM and/or ICD-10-CM/PCS codes used for analysis to ensure the codes are in effect during the time period studied.

- Although the KID contains up to 40 diagnoses (30 prior to the 2019 KID; 25 prior to the 2016 KID; 15 prior to the 2009 KID) and 25 procedures (15 prior to the 2019 KID), the number of diagnoses and procedures varies across hospitals. Some hospitals provide as many as 50 diagnoses and procedures or more, while others provide as few as 9 diagnoses and 6 procedures. Because very few cases have more than 40 diagnoses or more than 25 procedures, the diagnosis and procedure vectors were truncated when necessary to save space in the KID data files. See the notes on diagnoses and procedures on the HCUP-US website.

- The collection and reporting of external cause of injury (E codes under ICD-9-CM) and external cause of morbidity (V, W, X, and Y codes under ICD-10-CM) also varies across hospitals depending on the presence of State laws or mandates for the collection of these codes. Some States do not require hospitals to report codes for "misadventures to patients during surgical and medical care" (codes E870-E879 under ICD-9-CM) which means that these occurrences will be underreported. Beginning with 2017 data, separate external cause code data elements are discontinued (formerly HCUP I10_ECAUSEn data elements). External cause codes are now at the end of the ICD-10-CM diagnosis array (I10_DXn).

Diagnosis-Related Groups

- The Diagnosis-Related Group (DRG) appropriate for the date of discharge is assigned by the Medicare DRG Grouper algorithm during HCUP processing and stored in the HCUP data element DRG. The software for DRG assignment changes on the fiscal year, so the KID includes DRGs assigned with one version of the software from January to September and another version for discharges in October to December. The version of the DRG software is indicated in the HCUP data element DRGVER.

- Coding for DRGs has changed over time:
    - Effective October 2007 (Version 25 of the DRGs), CMS revised the DRG scheme into Medicare-Severity DRGs (MS-DRGs). This expanded the number of DRGs from 540 to over 700. First, DRGs were consolidated into 335 base MS-DRGs. Of these, 106 were split into two subgroups, and 152 were split into three subgroups, to arrive at 745 total MS-DRGs. The subgroups were based on the presence of complications or comorbidities (CCs) or major CCs (MCCs).
    - Effective October 2015 (Version 33 of the DRGs), the MS-DRG algorithm uses ICD-10-CM/PCS codes. Differences in the underlying coding schemes can cause differences in the DRG assignment. An analysis of trends by DRG under ICD-9-CM and ICD-10-CM/PCS is available on the ICD-10-CM/PCS Resources page of the HCUP-US website.

Missing Values

- Missing data values can compromise the quality of estimates. If the outcome for discharges with missing values is different from the outcome for discharges with valid values, then

sample estimates for that outcome will be biased and inaccurately represent the discharge population. For example, in the 1997 KID, RACE was missing on over 22 percent of records because some hospitals and HCUP Partners did not supply it. As a result, race and ethnicity-specific estimates, especially estimates of discharge totals by race and ethnicity, could have been biased. However, RACE coding has improved in recent years. RACE is only missing on about 8.4 percent of discharges in the 2022 KID.

- There are several techniques available to help assess and overcome this missing data bias.[3] Descriptions of such data preparation and adjustment are outside the scope of this report; however, it is recommended that researchers evaluate and adjust for missing data, if necessary. For details, see the report, *Missing Data Methods for the NIS and the SID*, available on the HCUP-US website.

Hospital-Level Data Elements

- Beginning with the 2012 KID, specific hospital identifiers (e.g. the AHA identifier) are no longer available. This means that you will not be able to link the KID to outside data sources that require hospital-specific identifiers. However, there are hospital-level data elements for all hospitals in the KID that allow you to study certain hospital characteristics including ownership/control, teaching status, rural/urban location, bed size, the four census regions, and a data element that allows you to identify whether a hospital is a freestanding children's hospital (KID_STRATUM=9999 [or 9998 in 2003]). A detailed description of the data elements is available on the HCUP-US website.

Hospital-Level Analyses

- The KID contains a systematic random sample of 10% of normal newborns born in the hospital and 80% of other newborns and pediatric cases from each sampled hospital. Therefore, researchers could estimate the number of discharges for each hospital as: sampled discharges / the sampling rate (.8 for cases other than normal newborns). However, because the KID includes less than 100% of pediatric discharges for each hospital, researchers will not be able to calculate hospital-specific outcomes with certainty.

Longitudinal Hospital Analyses

- The KID includes a sample of pediatric discharges from all HCUP hospitals in the sampling frame. However, beginning with 2012 data the KID hospital number (HOSP_KID) values are reassigned each year, so they cannot be used to link hospitals across years. Thus, the KID cannot support longitudinal analyses of specific hospitals.

[3] See, for example, van Buuren, S. (2012). *Flexible Imputation of Missing Data.* CRC Press, Boca Raton, FL.

Calculating National Estimates

- **In order to produce national estimates, you MUST use discharge weights.** Use the discharge weight (DISCWT[4]) to project discharges in the KID Core files to the discharges from all U.S. community, non-rehabilitation hospitals. (For trends analysis using 1997 KID data, see the next section of this report regarding trend weights.)

- Because the KID is a stratified sample, proper statistical techniques must be used to calculate standard errors and confidence intervals when using the KID. For detailed instructions, refer to the special report *Calculating Kids' Inpatient Database (KID) Variances*,[5] available on www.hcup-us.ahrq.gov.

- When creating national estimates, it is a good idea to check your estimates against other data sources, if available.

- To ensure that you are using the weights appropriately and calculating estimates and variances accurately, check your estimates against HCUPnet, the free online query system . 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.

Studying Trends

- For analysis of diagnoses and procedures, beginning with 2015 data, the introduction of ICD-10-CM/PCS means that trends that rely on diagnosis and procedures may be interrupted. Analyses that do not rely on diagnosis and procedure coding should not be affected. HCUP has developed recommendations for reporting trends that span the October 1, 2015 transition date (before and after the introduction of ICD10-CM/PCS). In addition, the HCUP-US website has a section on ICD-10-CM/PCS Resources that summarizes key issues for researchers using HCUP and other administrative databases that include ICD-9-CM and ICD-10-CM/PCS coding.

- KID data are available every three years from 1997 through 2012; then every three years beginning with 2016, which allows researchers to analyze trends over time. When studying trends using the KID, be aware that the sampling frame for the KID changes over time (i.e., more States have been added). Estimates from earlier years of the KID may be subject to more sampling bias than later years of the KID.

- Short-term rehabilitation hospitals are included in the 1997 KID but are excluded from later

[4] Prior to 2000, the discharge weight was named DISCWT_U. For 2000 only, use DISCWT to create national estimates for all analyses except those that involve total charges; and use DISCWTCHARGE to create national estimates of total charges.
[5] This report has not been updated for the 2012 KID data element changes. However, the same statistical techniques should be used to calculate standard errors and confidence intervals. There is one change in example programs: HOSPID (the encrypted hospital identifier) should be replaced by HOSP_KID.

years of the KID. Patients treated in short-term rehabilitation hospitals tend to have lower mortality rates and longer lengths of stay than patients in other community hospitals. The elimination of rehabilitation hospitals may affect trends, but the effect is likely small since only about 3% of community hospitals are short-term rehabilitation hospitals and not all State data sources included short term rehabilitation hospitals.

- To facilitate analysis of trends including the 1997 KID, another set of KID discharge weights for the 1997 HCUP KID were developed. These trends weights were calculated in the same way as the weights for the 2000 and later years of the KID. (Trend analyses for 2000 and later data do not need the KID trends weights.) The report, *Using the Kids' Inpatient Database (KID) to Estimate Trends*, includes details regarding the KID trends weights and other recommendations for trends analysis. The KID trends report is available on the HCUP-US Web site at www.hcup-us.ahrq.gov/reports/methods/2007_02.pdf, and the KID trends weights are available on the HCUP-US Web site at www.hcup-us.ahrq.gov/db/nation/kid/kidtrends.jsp.

Studying Readmissions

- The KID contains discharge-level records, not patient-level records. This means that individual patients who are hospitalized multiple times in one year may be present in the KID multiple times. There is no uniform patient identifier available that allows a patient-level analysis with the KID. This will be especially important to remember for certain conditions for which patients may be hospitalized multiple times in a single year. Researchers wishing to examine readmissions should use either the Nationwide Readmissions Database (NRD), or the State Inpatient Databases (SID) and accompanying Supplemental Variable for Revisit Analysis which allow identification of readmissions for individual patients. See the Databases documentation on the HCUP-US website for more information.

Variance Calculations

It may be important for researchers to calculate a measure of precision for some estimates based on the KID sample data. Variance estimates must take into account both the sampling design and the form of the statistic. If hospitals inside the frame are similar to hospitals outside the frame, the sample hospitals can be treated as if they were randomly selected from the entire universe of hospitals within each stratum. Discharges were randomly selected from within each hospital. Standard formulas for stratified, two-stage cluster samples without replacement may be used to calculate statistics and their variances in most applications. **To accurately calculate variances from the KID, you must use appropriate statistical software and techniques.** For details, see the special report, *Calculating Kids' Inpatient Database (KID) Variances*.[6] This report is available on the HCUP-US Web site at www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp.

A multitude of statistics can be estimated from the KID data. Several computer programs that calculate statistics and their variances from sample survey data are listed in the section below. Some of these programs use general methods of variance calculations (e.g., the jackknife and balanced half-sample replications) that take into account the sampling design. However, it may be desirable to calculate variances using formulas specifically developed for some statistics.

[6] This report has not been updated for the 2012 KID data element changes. However, the same statistical techniques should be used to calculate standard errors and confidence intervals. There is one change in example programs: HOSPID (the encrypted hospital identifier) should be replaced by HOSP_KID.

These variance calculations are based on finite-sample theory, which is an appropriate method for obtaining cross-sectional, nationwide estimates of outcomes. According to finite-sample theory, the intent of the estimation process is to obtain estimates that are precise representations of the nationwide population at a specific point in time. In the context of the KID, any estimates that attempt to accurately describe characteristics (such as expenditure and utilization patterns or hospital market factors) and interrelationships among characteristics of hospitals and discharges during a specific year should be governed by finite-sample theory.

Alternatively, in the study of hypothetical population outcomes not limited to a specific point in time, the concept of a "superpopulation" may be useful. Analysts may be less interested in specific characteristics from the finite population (and time period) from which the *sample* was drawn than they are in hypothetical characteristics of a conceptual superpopulation from which any particular finite *population* in a given year might have been drawn. According to this superpopulation model, the nationwide population in a given year is only a snapshot in time of the possible interrelationships among hospital, market, and discharge characteristics. In a given year, all possible interactions between such characteristics may not have been observed, but analysts may wish to predict or simulate interrelationships that may occur in the future.

Under the finite-population model, the variances of estimates approach zero as the sampling fraction approaches one. This is the case because the population is defined at that point in time, and because the estimate is for a characteristic as it existed when sampled. This contrasts with the superpopulation model, which adopts a stochastic viewpoint rather than a deterministic viewpoint. That is, the nationwide population in a particular year is viewed as a random sample of some underlying superpopulation over time. Different methods are used for calculating variances under the two sample theories. The choice of an appropriate method for calculating variances for nationwide estimates depends on the type of measure and the intent of the estimation process.

Computer Software for Variance Calculations

The discharge weights would be used to weight the sample data in estimating population statistics. In most cases, computer programs are readily available to perform these calculations. Several statistical programming packages allow weighted analyses.[7] For example, nearly all SAS procedures incorporate weights. In addition, several statistical analysis programs have been developed to specifically calculate statistics and their standard errors from survey data. Version eight or later of SAS contains procedures (PROC SURVEYMEANS and PROC SURVEYREG) for calculating statistics based on specific sampling designs. STATA and SUDAAN are two other common statistical software packages that perform calculations for numerous statistics arising from the stratified, single-stage cluster sampling design. Examples of the use of SAS, SUDAAN, and STATA to calculate KID variances are presented in the special report: *Calculating Kids' Inpatient Database (KID) Variances*. This report is available on the HCUP-US Web site at www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp. For an excellent review of programs to calculate statistics from survey data, visit the following Web site: www.hcp.med.harvard.edu/statistics/survey-soft/.

The KID database includes a Hospital file with data elements required to calculate finite

[7] Carlson BL, Johnson AE, Cohen SB. "An Evaluation of the Use of Personal Computers for Variance Estimation with Complex Survey Data." Journal of Official Statistics, vol. 9, no. 4, 1993: 795-814.

population statistics. The file includes hospital identifiers (Primary Sampling Units or PSUs), stratification data elements, and stratum-specific totals for the numbers of discharges and hospitals so that finite-population corrections can be applied to variance estimates.

In addition to these subroutines, standard errors can be estimated by validation and cross-validation techniques. Given that a very large number of observations will be available for most analyses, it may be feasible to set aside a part of the data for validation purposes. Standard errors and confidence intervals can then be calculated from the validation data.

If the analytical file is too small to set aside a large validation sample, cross-validation techniques may be used. For example, tenfold cross-validation would split the data into ten equal-sized subsets. The estimation would take place in ten iterations. In each iteration, the outcome of interest is predicted for one-tenth of the observations by an estimate based on a model fit to the other nine-tenths of the observations. Unbiased estimates of error variance are then obtained by comparing the actual values to the predicted values obtained in this manner.

Finally, it should be noted that a large array of hospital-level data elements are available for the entire universe of hospitals, including those outside the sampling frame. For instance, the data elements from the AHA surveys and from the Medicare Cost Reports are available for nearly all hospitals in the U.S, although hospital identifiers are suppressed in the KID for several States. For these States it will not be possible to link to outside hospital-level data sources. To the extent that hospital-level outcomes correlate with these data elements, they may be used to sharpen regional and nationwide estimates.

SAMPLING OF DISCHARGES

Sampling of Discharges Included in the KID

Unlike the HCUP Nationwide Inpatient Sample (NIS) prior to 2012, the KID has never involved sampling hospitals. Instead, the KID includes a sample of pediatric discharges from all hospitals in the sampling frame[8]. For the sampling, pediatric discharges in all participating States are stratified by normal newborns, other newborns, and all other pediatric cases. To further ensure an accurate representation of each hospital's pediatric case-mix, the discharges are sorted by State, hospital, DRG, and a random number within each DRG. Systematic random sampling is used to select 10 percent of normal newborns born in the hospital and 80 percent of other newborns and pediatric cases from each frame hospital.

To obtain national estimates, discharge weights are developed using the AHA universe as the standard. For the weights, hospitals are post-stratified on six characteristics contained in the AHA hospital files—ownership/control, bed size, teaching status, rural/urban location, U.S. region, and a stratum for freestanding children's hospitals. If there were fewer than two frame hospitals, 30 normal newborns, 30 other newborns, and 30 non-birth pediatric discharges sampled in a stratum, that stratum is combined with an "adjacent" stratum containing hospitals with similar characteristics. Discharge weights are created by stratum in proportion to the total number of AHA newborns for in-hospital births and in proportion to the total number of AHA non-newborn admissions for non-birth pediatric discharges.

[8] As of 2012, the sampling strategy for National Inpatient Sample (NIS) was redesigned as a sample of discharges from all HCUP-participating hospitals. For more information on the new design for the NIS, see the NIS Overview, available on HCUP-US website at www.hcup-us.ahrq.gov/nisoverview.jsp.

The KID Hospital Universe

The hospital universe is defined as all hospitals located in the U.S. that were open during any part of the calendar year and that were designated as community hospitals in the AHA Annual Survey Database. The AHA defines community hospitals as follows: "All non-Federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions." Starting in 2005, the AHA included long term acute care facilities in the definition of community hospitals. These facilities provide acute care services to patients who need long term hospitalization (more than 25 days stays). Consequently, Veterans Hospitals and other Federal facilities (Department of Defense and Indian Health Service) are excluded. Beginning with the 2000 KID, short-term rehabilitation hospitals were excluded from the universe, because the type of care provided and the characteristics of the discharges from these facilities were markedly different from other short-term hospitals. (The 1997 KID includes short-term rehabilitation hospitals. The KID Trend Weights, described earlier in this report, remove these hospitals, and adjust for other design changes in the 2000 KID.) Table 2 in Appendix I displays the number of hospitals in the universe for each year, based on the corresponding AHA Annual Survey Database.

For more information on how hospitals in the data set were mapped to hospitals as defined by the AHA, refer to the special report, *HCUP Hospital Identifiers*. For a list of all data sources, refer to Table 1 in Appendix I. Detailed information on the design of the KID prior to 2006 is available in the year-specific special reports on *Design of the Kids' Inpatient Database* found on the HHCUP-US Web site at www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp. Starting with the 2006 KID, the design information was incorporated into this report.

Hospital Merges, Splits, and Closures

All U.S. hospital entities that were designated community hospitals in the AHA hospital file, except short-term rehabilitation hospitals, were included in the hospital universe. Therefore, when two or more community hospitals merged to create a new community hospital, the original hospitals and the newly-formed hospital were all considered separate hospital entities in the universe during the year they merged. Similarly, if a community hospital split, the original hospital and all newly-created community hospitals were treated as separate entities in the universe during the year this occurred. Finally, community hospitals that closed during a given year were included in the hospital universe, if they were in operation during some part of the calendar year.

Stratification Data Elements

To calculate discharge weights, we post-stratified hospitals on six characteristics contained in the AHA hospital files—ownership/control, bed size, teaching status, rural/urban location, and U.S. region, with the addition of a stratum for freestanding children's hospitals. The definitions of some of the strata were revised beginning with the 2000 KID. (A description of the strata used for the 1997 KID can be found in the *Kids' Inpatient Database (KID) Design Report, 1997*. This report is available on the HCUP-US Web site at www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp.)

Beginning with the 2000 KID, the stratification data elements were defined as follows:

- *Geographic Region – Northeast, Midwest, West, and South.* This is an important stratification data element because practice patterns have been shown to vary substantially by region. For example, lengths of stay tend to be longer in East Coast hospitals than in West Coast hospitals. Figure 1 in Appendix I shows the KID States by region.

- *Control – government non-Federal (public), private not-for-profit (voluntary), and private investor-owned (proprietary)*. These types of hospitals tend to have different missions and different responses to government regulations and policies. When there were enough hospitals of each type to allow it, hospitals were stratified as public, voluntary, and proprietary. This stratification was used for Southern rural, Southern urban non-teaching, and Western urban non-teaching hospitals. For smaller strata – the Midwestern rural and Western rural hospitals – a collapsed stratification of public versus private was used, with the voluntary and proprietary hospitals combined to form a single "private" category. For all other combinations of region, location, and teaching status, no stratification based on control was advisable, given the number of hospitals in these cells.

- *Location – urban or rural. Government payment policies for hospital services often differ according to this designation. Also, rural hospital*s are generally smaller and offer fewer services than urban hospitals. We categorized hospitals with a CBSA type of *Metropolitan* or *Division* as urban, while we designated hospitals with a CBSA type of *Micropolitan* or *Rural* as rural.

- *Teaching Status – teaching or non-teaching*. The missions of teaching hospitals differ from non-teaching hospitals. In addition, financial considerations differ between these two hospital groups. Currently, the Medicare DRG payments are uniformly higher to teaching hospitals. Prior to 2006, the teaching status of hospitals identified as children's hospitals by the Children's Hospital Association (CHA, formerly National Association of Children's Hospitals and Related Institutions, NACHRI) was based on an indicator provided by CHA. For 2006, and for 2012 and later data, the CHA teaching status indicator was not available, so teaching status was determined using only information from the AHA Annual Survey Database for all hospitals. For 2009, both CHA and AHA information was used to define the teaching status of children's hospitals.

  In the 1997 KID, we considered other hospitals to be teaching hospitals if they had any residents or interns and met one of the following two criteria:

  - Residency training approval by the Accreditation Council for Graduate Medical Education (ACGME)
  - Membership in the Council of Teaching Hospitals (COTH).

  Beginning with the 2000 KID, we considered other hospitals to be teaching hospitals if they met any one of the following three criteria: (See Appendix IV for details.)

  - Residency training approval by the Accreditation Council for Graduate Medical Education (ACGME)
  - Membership in the Council of Teaching Hospitals (COTH)
  - A ratio of full-time equivalent interns and residents to beds of .25 or greater.

- *Bed Size – small, medium, and large.* Bed size categories are based on hospital beds and are specific to the hospital's region, location, and teaching status, as illustrated in Table 5 of Appendix I. The bed size cutoff points were chosen so that approximately one-third of the hospitals in a given region, location, and teaching status combination would fall within each bed size category (small, medium, or large). Different cutoff points for rural, urban non-teaching, and urban teaching hospitals were used because hospitals in those categories tend to be small, medium, and large, respectively. For example, a medium-sized teaching hospital would be considered a rather large rural hospital. Further, the size distribution is different among regions for each of the urban/teaching categories. For example, teaching hospitals tend to be smaller in the West than they are in the South. Using differing cutoff points in this manner avoids strata containing small numbers of hospitals.

  Rural hospitals were not split according to teaching status, because rural teaching hospitals were rare. For example, rural teaching hospitals generally comprise about 3% or less of the total hospital universe. The bed size categories were defined within location and teaching status because they would otherwise have been redundant. Rural hospitals tend to be small; urban non-teaching hospitals tend to be medium-sized; and urban teaching hospitals tend to be large. Yet it was important to recognize gradations of size within these types of hospitals. For example, in serving rural discharges, the role of "large" rural hospitals (particularly rural referral centers) often differs from the role of "small" rural hospitals.

- *Hospital Type – freestanding children's or other hospital*. Children's hospitals restrict admissions to children, while other hospitals admit both adults and children. There may be significant differences in practice patterns, casemix, severity of illness, and available services between children's hospitals and other hospitals. Freestanding children's hospitals are assigned to a separate stratum and can be identified using KID_STRATUM=9999 (or 9998 in 2003). Children's units in general hospitals were not stratified as children's hospitals. Prior to the 2016 KID, data from the Children's Hospital Association (CHA) were used to help verify the AHA list of children's hospitals. Beginning with 2016, children's hospitals were identified using only AHA data.

KID Sampling Frame

The *universe* of hospitals was established as all community hospitals located in the U.S. with the exception, beginning in 2000, of short-term rehabilitation hospitals. However, some hospitals do not supply data to HCUP. Therefore, we constructed the KID *sampling frame* from the subset of universe hospitals that released their discharge data to AHRQ for research use. The number of State Partners and hospitals contributing data to the KID has expanded over the years, as shown in Table 2 of Appendix I.

The list of the entire frame of hospitals was composed of all AHA community, non-rehabilitation hospitals in each of the frame States *that could be matched to the discharge data provided to HCUP*. If an AHA hospital could not be matched to the discharge data provided by the data source, it was eliminated from the sampling frame (but not from the target universe).

Table 5 of Appendix I shows the number of AHA, HCUP SID, and KID hospitals by Region. In most cases, the difference between the universe and the frame represents the difference between the number of community, non-rehabilitation hospitals in the 2016 AHA Annual Survey Database and the number of hospitals with children's discharges that were supplied to HCUP that could be matched to the AHA data.

Beginning with the 2000 KID, pediatric discharges were defined as having an age at admission of 20 or less. This differs from the 1997 KID, which included discharges with an admission age of 18 or less. Discharges with missing, invalid, or inconsistent ages were excluded.

KID Sample Design

Design Considerations

The overall design objective was to select a sample of pediatric discharges that accurately represents the target universe of U.S. community, non-rehabilitation hospitals. Moreover, this sample was to be geographically dispersed, yet drawn exclusively from hospitals in States that participate in HCUP.

It should be possible, for example, to estimate DRG-specific average lengths of stay across all U.S. hospitals using weighted average lengths of stay, based on averages or regression coefficients calculated from the KID. Ideally, relationships among outcomes and their correlates estimated from the KID should accurately represent all U.S. hospitals. It is advisable to verify your estimates against other data sources, especially for specific patient populations (e.g. organ transplant recipients).

In order to sample and project births up to the number of births reported by the AHA, which reports in-hospital births, the KID development team identified all in-hospital births in the KID data. We further separated the in-hospital births in HCUP data into normal newborns and other newborns. We sampled normal newborns at a lower rate because they have little variation in their outcomes.

In-hospital births (HOSPBRTH = 1) were identified by any ICD-10-CM principal or secondary diagnosis code of 'Z3800', 'Z3801', 'Z382', 'Z3830', 'Z3831', 'Z385', 'Z3861' 'Z3862', 'Z3863', 'Z3864', 'Z3865', 'Z3866', 'Z3868', 'Z3869', or 'Z388' (previously by any ICD-9-CM principal or secondary diagnosis in the range of V3000 to V3901 with the last two digits of "00" or "01") and the patient was not transferred from another acute care hospital or health care facility.

We classified neonates transferred from other facilities as pediatric non-births because they are not included in births reported by the AHA. An age of zero days was not a reliable in-hospital birth indicator because neonates transferred from another hospital or born before admission to the hospital could also have an age of zero days. There were also some cases with birth diagnoses, but with ages of a few days. Because the HCUP data are already edited for neonatal diagnoses inconsistent with age, we did not include any age criteria in the in-hospital birth screen.

Normal, in-hospital births are identified as cases that meet the above screen and have a Diagnosis Related Group (DRG) indicating "Normal Newborn" (391 prior to 2009, or 795 beginning in 2009). In the KID, a small percentage of the cases with a DRG of "Normal Newborn" do not meet the in-hospital birth screen. These cases have diagnoses that imply a newborn, but do not specifically indicate an in-hospital birth. It is possible that some of these may have been born in the hospital but lacked the proper diagnosis code. Others may be readmissions or may have been born before admission to the hospital. Some of these cases have an admission type of newborn (ATYPE = 4).

Changes to Sampling and Weighting Strategy Beginning with the 2000 KID

We revised some of the hospital universe and strata definitions beginning with the 2000 KID. These changes included:

- Revising definitions of the strata data elements
- Excluding rehabilitation hospitals from the hospital universe
- Changing the calculation of hospital universe discharges for the weights.

Sampling Procedure

The KID includes a sample of pediatric discharges from all hospitals in the sampling frame. For the sampling, we stratified the pediatric discharges by normal newborns, other newborns, and non-newborn pediatric discharges. To further ensure an accurate representation of each hospital's pediatric case-mix, we also sorted the discharges by State, hospital, DRG, and a random number within each DRG. We then used systematic random sampling to select 10 percent of "normal newborns" born in the hospital and 80 percent of other newborns and pediatric cases from each frame hospital.

It should be observed that the KID includes fewer than 100 percent of the pediatric discharges for each hospital in the database. Therefore, researchers will not be able to calculate hospital-specific outcomes with certainty.

SAMPLE WEIGHTS

To obtain national estimates, we developed discharge weights using the AHA universe as the standard. For the weights, hospitals are post-stratified on six characteristics contained in the AHA hospital files—ownership/control, bed size, teaching status, rural/urban location, and U.S. region, with the addition of a stratum for freestanding children's hospitals. We also stratified the KID discharges according to whether the discharge was a normal newborn, other newborn, or a non-newborn pediatric discharge. If there were fewer than two frame hospitals, 30 normal newborns, 30 other newborns, and 30 non-birth pediatric discharges sampled in a stratum, we merged that stratum with an "adjacent" stratum containing hospitals with similar characteristics.

Prior to the 2016 KID, we used Children's Hospital Association (CHA) data to help verify the AHA list of children's hospitals in the target universe. Beginning with 2016, CHA data was not available. Data analysts may find it useful to identify discharges from children's hospitals. Prior to 2012, children's hospitals within general hospitals were not stratified as children's hospitals, but they could be selected using the NACHTYPE data element in the KID. Beginning with 2012 data, NACHTYPE is no longer available, but discharges from freestanding children's hospitals are stratified as children's hospitals (KID_STRATUM=9999 [or 9998 in 2003]).

Discharge Weights

The discharge weights usually are constant for all discharges of the same type (normal newborns, and other pediatric discharges) within a stratum. The only exceptions are for strata with sample hospitals that, according to the AHA files, were open for the entire year but contributed less than their full year of data to the KID. For those hospitals, we *adjusted* the number of observed discharges by a factor of 4 ÷ Q, where Q was the number of calendar quarters that the hospital contributed discharges to the KID. For example, when a sample hospital contributed only two quarters of discharge data to the KID, the *adjusted* number of

discharges was double the observed number.

With that minor adjustment, each discharge weight is essentially equal to the number of AHA universe discharges that each sampled discharge represents in its stratum. This calculation was possible because the numbers of total discharges and births were available for every hospital in the universe from the AHA files.

Discharge weights to the universe were calculated by post-stratification. Hospitals were stratified on geographic region, urban/rural location, teaching status, bed size, control, and hospital type. In some instances, strata were collapsed for sample weight calculations. Within stratum *k*, for hospital *i*, each KID sample discharge's universe weight was calculated as:

$$W_{ik} = [T_k / (R_k * A_k)] * (4 \div Q_i)$$

In the birth strata (both normal newborns and other newborns):

- T_k is the total number of births reported in the AHA survey.
- A_k is the total number of adjusted births in the restricted sampling frame.
- In the normal newborns strata, R_k is the frame sampling rate for normal newborns calculated as the sum of the adjusted number of normal newborns sampled divided by the sum of the adjusted number of normal newborns born in the hospital in the restricted frame.
- In the other newborn strata, R_k is the frame sampling rate for other newborns born in the hospital.

In the non-newborn strata:

- T_k is the total number of non-newborns reported in the AHA survey.
- A_k is the total number of adjusted non-newborn discharges in the sampling frame.
- R_k is the frame sampling rate for non-newborns from all non-newborn discharges in the sampling frame.
- Qi is the number of quarters of discharge data contributed by hospital i to the KID (usually $Q_i = 4$).
- T_k / A_k estimates the number of discharges in the population that is represented by each discharge in the sampling frame. R_k adjusts for the fact that we are taking a sample of the frame in each stratum.

Normal newborns were sampled at a lower rate than other discharges because the variation in hospital outcomes for normal newborns is considerably less than that for other pediatric cases and because we expect research to focus much more on other pediatric patients. We sampled normal newborns at the nominal rate of 10 percent and sampled other pediatric discharges (other newborns and other pediatric cases) at the nominal rate of 80 percent from the discharges available in the (restricted) frame. To avoid rounding errors in the weights calculation, the actual sampling rate for a discharge type (normal newborn, other newborn, or non-newborn pediatric discharge) in stratum *k*, R_k, was calculated as follows:

$R_k = S_k / H_k$

- S_k is the number of adjusted discharges ***sampled*** for the discharge type in stratum *k*.
- H_k is the number of adjusted discharges in the sampling frame for the discharge type in stratum *k*.

The AHA birth counts include both normal newborns and other newborns born in the hospital. Therefore, the weights in the normal newborn strata implicitly assume that the proportion of normal newborns in the frame is representative of the proportion of normal newborns in the population for each stratum. A similar assumption is made for other newborns.

Similarly, the non-birth AHA counts include all non-birth admissions, not just non-birth pediatric counts. Consequently, the weights in the non-birth strata implicitly assume that the proportion of non-birth discharges that are pediatric across the HCUP SID hospitals is the same as the proportion of non-birth admissions that are pediatric across the universe of AHA hospitals, in the aggregate within each hospital stratum.

Weight Data Elements

To produce nationwide estimates, use the discharge weights to project sampled discharges in the Core file to the discharges from all U.S. community, non-rehabilitation hospitals. Beginning with the 2003 KID, use DISCWT to calculate nationwide estimates for all analyses. For the 2000 KID, use DISCWT to create nationwide estimates for all analyses except those that involve total charges, and use DISCWTCHARGE to create nationwide estimates of total charges. For the 1997 KID, use DISCWT_U for all analyses. (For trends analysis using 1997 KID data, see the previous section of this report regarding "Studying Trends.")

THE FINAL KID SAMPLE

In Appendix I, we present tables and figures that summarize the final KID sample.

Table 2 summarizes information across all years of the KID, including the KID States, data sources, sample hospitals, and sample discharges.

Table 6 shows the number of hospitals and discharges for children's hospitals and other hospitals. For each hospital type, the table shows the number of:

- AHA universe hospitals and total admissions plus births
- Non-rehabilitation community hospitals in the SID and associated pediatric discharges
- Hospitals and pediatric discharges included in the KID.

Table 7 displays the unweighted and weighted number of uncomplicated births, complicated births, and pediatric non-births by hospital type in the KID.

Figure 2 displays the KID hospitals by geographic region. For each region, the chart presents:

- The number of hospitals in the AHA universe

- The number of SID community hospitals with pediatric discharges
- The number of hospitals in the KID (and the percentage of AHA universe hospitals).

Although pediatric discharges from hospitals in each region are selected for the KID, the comprehensiveness of the sampling frame varies by region, as shown in Figure 2.

Figure 3 summarizes the estimated U.S. population by geographic region on July 1, 2022. For each region, the figure reveals:

- The estimated U.S. population
- The estimated population of States in the KID
- The percentage of estimated U.S. population included in KID States.

This figure shows that the sampling frame for the KID includes states that comprise 97 percent of the U.S. population.

Appendix I: Tables and Figures

Table 1. Data Sources for the 2022 KID

State	Data Organization
AK	Alaska Department of Health
AR	Arkansas Department of Health
AZ	Arizona Department of Health Services
CA	Department of Health Care Access and Information
CO	Colorado Hospital Association
CT	Connecticut Hospital Association
DC	District of Columbia Hospital Association
DE	Delaware Division of Public Health
FL	Florida Agency for Health Care Administration
GA	Georgia Hospital Association
HI	Hawaii Laulima Data Alliance
IA	Iowa Hospital Association
IL	Illinois Department of Public Health
IN	Indiana Hospital Association
KS	Kansas Hospital Association
KY	Kentucky Cabinet for Health and Family Services
LA	Louisiana Department of Health and Hospitals
MA	Massachusetts Center for Health Information and Analysis
MD	Health Services Cost Review Commission
ME	Maine Health Data Organization
MI	Michigan Health & Hospital Association
MN	Minnesota Hospital Association
MO	Hospital Industry Data Institute
MS	Mississippi Department of Health
MT	Montana Hospital Association
NC	North Carolina Department of Health and Human Services

State	Data Organization
ND	North Dakota (data provided by the Minnesota Hospital Association)
NE	Nebraska Hospital Association
NH	New Hampshire Department of Health & Human Services
NJ	New Jersey Department of Health
NM	New Mexico Department of Health
NY	New York State Department of Health
OH	Ohio Hospital Association
OK	Oklahoma State Department of Health
OR	Oregon Association of Hospitals and Health Systems
PA	Pennsylvania Health Care Cost Containment Council
RI	Rhode Island Department of Health
SC	South Carolina Revenue and Fiscal Affairs Office
SD	South Dakota Association of Healthcare Organizations
TN	Tennessee Hospital Association
TX	Texas Department of State Health Services
UT	Utah Department of Health
VA	Virginia Health Information
VT	Vermont Association of Hospitals and Health Systems
WA	Washington State Department of Health
WI	Wisconsin Department of Health Services
WV	West Virginia Health Care Authority
WY	Wyoming Hospital Association

Return to Introduction

Table 2. Summary of KID States, Hospitals, and Inpatient Stays, 1997-2022

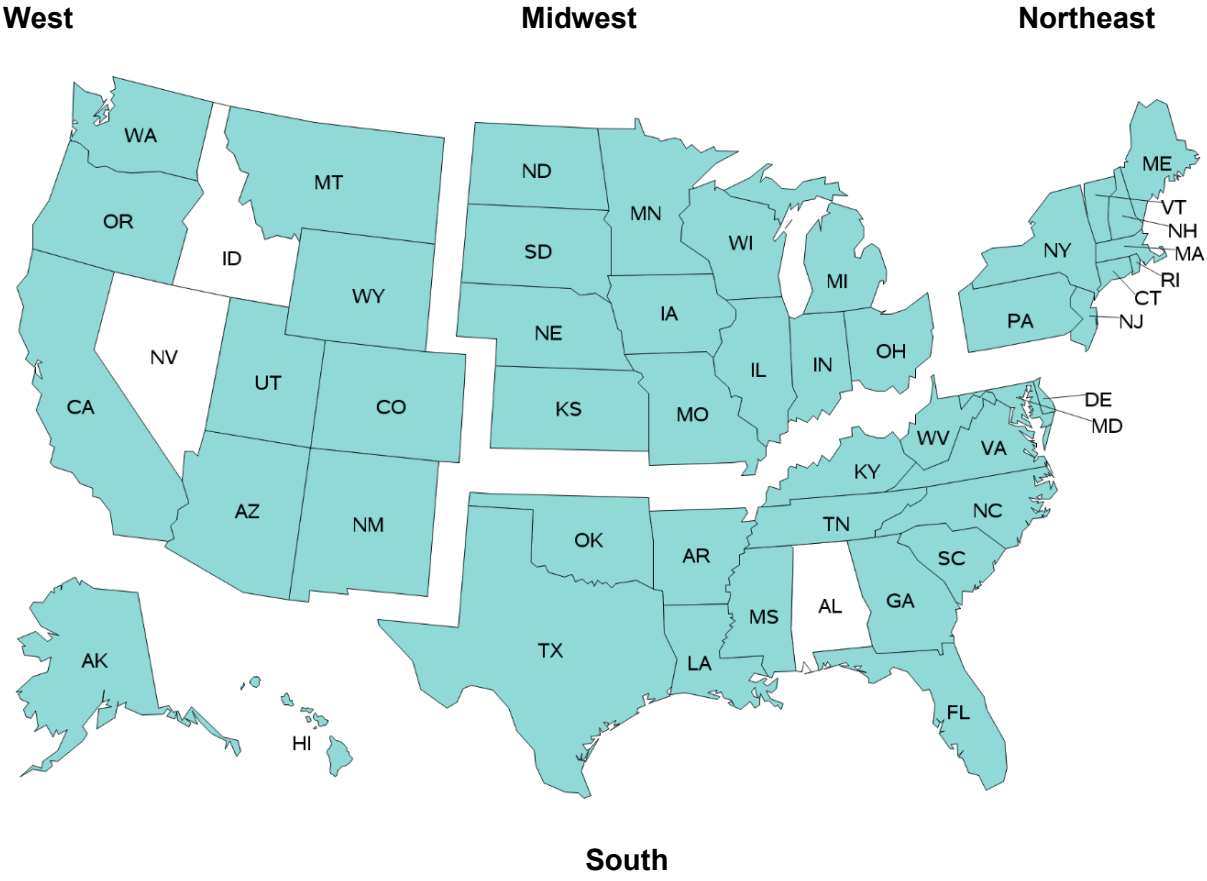
Year	Number of States	Data sources	AHA Hospital universe[1]	Number of SID hospitals with pediatric discharges	Number of pediatric discharges (unweighted)[2]	Number of pediatric discharges (weighted)[2]
2022	48	AK AR AZ CA CO CT DC DE FL GA HI IA IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY (NV not included)	4,795	3,811	3,009,812	5,660,863
2019	49	AK AR AZ CA CO CT DC DE FL GA HI IA IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY (Added DE and NH)	4,852	3,998	3,089,283	5,902,538
2016	47	AK AR AZ CA CO CT DC FL GA HI IA IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY (Added DC and MS. NH is not included)	5,001	4,200	3,117,413	6,266,285
2012	44	AK AR AZ CA CO CT FL GA HI IA IL IN KS KY LA MA MD MI MN MO MT NC ND NE NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY (Added AK, ND. ME and NH are not included)	5,118	4,179	3,195,782	6,675,222
2009	44	AR AZ CA CO CT FL GA HI IA IL IN KS KY LA MA MD ME MI MN MO MT NC NE NH NM NJ NV NY OH OK OR PA RI SC	5,128	4,121	3,407,146	7,370,203

Year	Number of States	Data sources	AHA Hospital universe[1]	Number of SID hospitals with pediatric discharges	Number of pediatric discharges (unweighted)[2]	Number of pediatric discharges (weighted)[2]
		SD TN TX UT VA VT WA WI WV WY (Added LA, ME, MT, NM, PA and WY)				
2006	38	AR AZ CA CO CT FL GA HI IA IL IN KS KY MA MD MI MN MO NC NE NH NJ NV NY OH OK OR RI SC SD TN TX UT VA VT WA WI WV (Added AR and OK. ME and PA are not included)	5,124	3,739	3,131,324	7,558,812
2003	36	AZ CA CO CT FL GA HI IA IL IN KS KY MA MD MI MN MO NC NE NH NJ NV NY OH OR RI SC SD TN TX UT VA VT WA WI WV (Added IL, IN, MI, MN, NE, NH, NV, OH, RI, SD, VT. ME and PA are not included)	4,836	3,438	2,984,129	7,409,162
2000	27	AZ CA CO CT FL GA HI IA KS KY MA MD ME MO NC NJ NY OR PA SC TN TX UT VA WA WI WV (Added KY, ME, NC, TX, VA, WV. IL is not included)	4,839	2,784	2,516,833	7,291,032
1997	22	AZ CA CO CT FL GA HI IL IA KS MD MA MO NJ NY OR PA SC TN UT WA WI	5,113	2,521	1,905,797	6,657,326

[1] The numbers of hospitals for the KID are based on the AHA Annual Survey files. Many AHA survey responses from hospitals cover a fiscal year other than a January-to-December calendar year.
For 1997, the hospital universe included community hospitals, including rehabilitation hospitals.
Beginning with 2000, the hospital universe includes community, non-rehabilitation hospitals.
[2] For 1997, discharges with age at admission of 18 years of less were included.
Beginning with 2000, discharges with age at admission of 20 years or less were included.

Return to Introduction

Figure 1. KID States, by Census Region, 2022



All States, by Region[9]

Region	States
Northeast	Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont.
Midwest	Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin.
South	*Alabama*, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia.
West	Alaska, Arizona, California, Colorado, Hawaii, *Idaho*, Montana, *Nevada*, New Mexico, Oregon, Utah, Washington, Wyoming.

Return to Introduction

[9] States and areas in italics do not participate in HCUP.

Table 3. KID Related Reports and Database Documentation Available on HCUP-US

Description of the KID Database

- KID Overview
  - HCUP Partners in the KID
- Introduction to the KID, 2022 – *this document*
- KID Related Reports
  Links to HCUP-US page with various KID related reports such as the following:
  - Design of the Kids' Inpatient Database for 1997 to 2003.
  - Changes in NIS Sampling and Weighting Strategy for 1998 (*for information on the 1998 NIS universe and strata which the KID uses*)
  - Calculating KID Variance Reports
  - KID Trends Report
  - KID Comparison Reports (available for years in which the KID sample changed)
  - HCUP E-Code Evaluation Report

Restrictions on Use

- Data Use Agreement Training
- Data Use Agreement for the KID
- Requirements for Publishing with HCUP data

File Specifications and Load Programs

- KID File Specifications
- Nationwide SAS Load Programs
- Nationwide SPSS Load Programs
- Nationwide Stata Load Programs

Data Elements

- KID Description of Data Elements – details uniform coding and State-specific idiosyncrasies
- KID Summary Statistics – lists means and frequencies on nearly all data elements
- Frequencies by Diagnosis and Procedure Codes – Excel files with discharge counts by clinical categories

ICD-10-CM/PCS Data Included in the KID Starting With 2016

- KID Changes Beginning Data Year 2016
- Additional ICD-10-CM/PCS Resources
- HCUP Software Tools Tutorial

Information on Change to KID Design in 2000

- Trend Weights for the 1997 KID for Consistent Estimates with the KID after 1997

Known Data Issues

- Information on corrections to the KID data sets

KID Supplemental Files

- Cost-to-Charge Ratio Files
- Hospital Market Structure (HMS) Files

HCUP Tools: Labels and Formats

- Format Programs
  - Labels file for multiple versions of Diagnosis Related Groups (DRGs) and Major Diagnostic Categories (MDCs)
  - SAS format library program to create value labels for KID data elements
  - HCUP Diagnosis and Procedure Groups Formats Program - formats to label DX_PR_Groups including CCS data elements
  - ICD-9-CM formats to label ICD-9-CM diagnoses and procedures
  - ICD-10-CM formats to label ICD-10-CM diagnoses and procedures
  - Severity formats to label severity data elements

Additional Resources for Data Elements		Obtaining HCUP Data
• KID Severity Measures – provides detailed documentation on the different types of measures • 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		• Purchase HCUP data from the HCUP Central Distributor

Return to Introduction

Table 4. Bed Size Categories, by Region

Location and Teaching Status	Hospital Bed Size		
	Small	Medium	Large
NORTHEAST			
Rural	1-49	50-99	100+
Urban, non-teaching	1-124	125-199	200+
Urban, teaching	1-249	250-424	425+
MIDWEST			
Rural	1-29	30-49	50+
Urban, non-teaching	1-74	75-174	175+
Urban, teaching	1-249	250-374	375+
SOUTH			
Rural	1-39	40-74	75+
Urban, non-teaching	1-99	100-199	200+
Urban, teaching	1-249	250-449	450+
WEST			
Rural	1-24	25-44	45+
Urban, non-teaching	1-99	100-174	175+
Urban, teaching	1-199	200-324	325+

Reurn to Introduction

Table 5: Number of AHA, HCUP SID, and KID Hospitals, by Region, 2022*

Census Region	AHA Universe Hospitals	SID Community Non-Rehab Hospitals	SID Community Non-Rehab Hospitals With Peds Discharges	KID Sample Hospitals
Total	4,795	4,395	3,865	3,811
Census Region				
1: Northeast	570	555	522	519
2: Midwest	1,409	1,321	1,127	1,101
3: South	1,860	1,681	1,443	1,426
4: West	956	838	773	765

*The columns in the table are defined as follows:

- "AHA Universe Hospitals" lists all community, non-rehabilitation hospitals in the AHA Survey data.
- "SID Community, Non-Rehabilitation Hospitals" lists potential KID sampling-frame hospitals before applying restrictions to the frame and before excluding hospitals without any pediatric discharges.
- "SID Community, Non-Rehabilitation Hospitals with Pediatric Discharges" lists potential KID sampling-frame hospitals with pediatric discharges after applying restrictions to the frame.
- "KID Sample Hospitals" lists the hospitals selected for the KID. Some hospitals may not be included in the KID because they had so few pediatric discharges that none were randomly sampled.

Return to Introduction

Table 6. Number of Hospitals and Discharges in the AHA Universe, SID, and KID, by Hospital Type, 2022

	AHA Universe		SID		KID	
Hospital Type	**Hospitals**	**Admissions Plus Births**	**Hospitals with Pediatric Discharges**	**Pediatric Discharges**	**Hospitals**	**Pediatric Discharges**
Children's Hospital	77	633,992	70	546,992	70	432,663
Not a Children's Hospital	4,718	33,936,539	3,795	4,761,058	3,741	2,577,149
Total	4,795	34,570,531	3,865	5,308,050	3,811	3,009,812

Return to Introduction

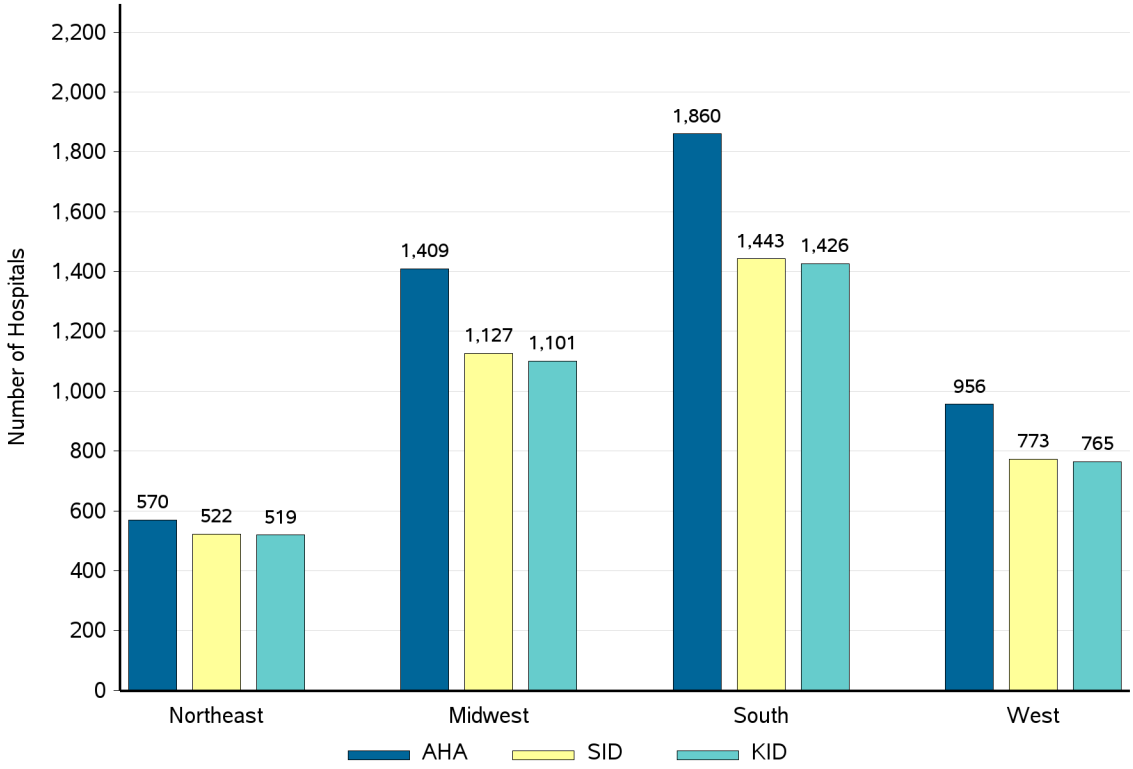
Table 7. 2022 KID Discharges, by Hospital Type*

Hospital Type	Normal Newborns	Other Newborns	Other Pediatric Discharges	Total Pediatric Discharges
Unweighted:				
Children's Hospital	705	8,973	422,985	432,663
Not a Children's Hospital	175,949	1,251,293	1,149,907	2,577,149
Total	176,654	1,260,266	1,572,892	3,009,812
Weighted:				
Children's Hospital	7,839	12,481	581,796	602,116
Not a Children's Hospital	1,854,177	1,636,212	1,568,359	5,058,748
Total	1,862,016	1,648,693	2,150,154	5,660,863

*Prior to the 2016 KID, data from the Children's Hospital Association (CHA) were used to help verify the AHA list of children's hospitals. Beginning with 2016, children's hospitals were identified using only AHA data.

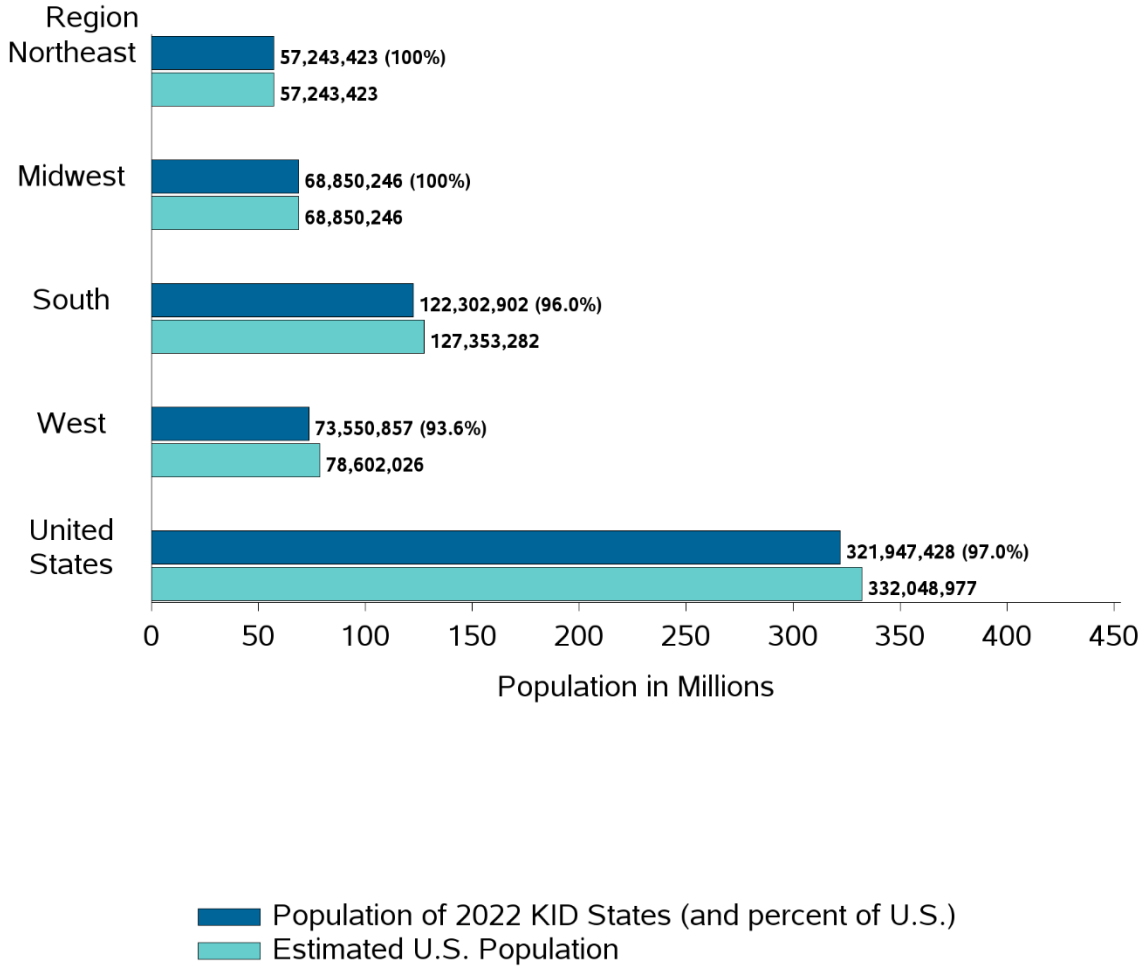
Return to Introduction

Figure 2. Number of Hospitals in the 2022 AHA Universe, SID, and KID, by Region



Return to Introduction

Figure 3. Percentage of U.S. Population in 2022 KID States, by Region
Calculated using the estimated U.S. population on July 1, 2022.[10]



Return to Introduction

[10] Source: Table 1. Annual Estimates of the Population for the United States, Regions, States, and Puerto Rico: April 1, 2020 to July 1, 2023 (NST-EST2023-POP). Source: U.S. Census Bureau, Population Division. Release Date: December 2020.

Appendix II: State-Specific Restrictions

The table below enumerates the types of restrictions applied to the KID. Restrictions include the following types:

- Confidentiality of records
    - Restricted release of age in years
    - Restricted release of medical misadventure or adverse reaction cause codes
- Missing discharges.

Confidentiality of Records - Restricted Release of Age in Years
• At least one Partner required ages in years (AGE) to be set to the midpoints of age ranges. • At least one Partner prohibits the release of medical misadventure or adverse reaction cause codes.

Missing Discharges
• At least one Partner prohibits the release of abortion discharges. • At least one Partner prohibits the release of discharge records for patients with HIV diagnoses. • At least one Partner prohibits the release of Alternate Level of Care (SNF / Swing Bed Skilled) discharges.

Return to Introduction

Appendix III: Data Elements

Table 1. Data Elements in the 2022 KID Inpatient Core File

For prior years, refer to the KID Description of Data Elements page on the HCUP-US Web site or to previous versions of the KID Introduction.

Type of Data Element	HCUP Name	Coding Notes
Admission day of week or weekend	AWEEKEND	Admission on weekend: (0) admission on Monday-Friday, (1) admission on Saturday-Sunday
Admission month	AMONTH	Admission month coded from (1) January to (12) December
Transferred into hospital	TRAN_IN	Transfer In Indicator: (0) not a transfer, (1) transferred in from a different acute care hospital [ATYPE NE 4 & (ASOURCE=2 or POO=4)], (2) transferred in from another type of health facility [ATYPE NE 4 & (ASOURCE=3 or POO=5, 6)]
Admission type	ELECTIVE	Indicates elective admission: (1) elective, (0) non-elective admission
Age at admission	AGE	Age in years coded 0-124 years
Neonatal age	AGE_NEONATE	Neonatal age (first 28 days after birth) indicator
Diagnosis information	I10_DX1 – I10_DX40	ICD-10-CM diagnoses, principal and secondary, with external cause of morbidity codes at the end of the array
	I10_HOSPBRTH	Birth diagnosis, in this hospital
	I10_NDX	Number of diagnoses coded on the original record
	I10_UNCBRTH	Normal, uncomplicated birth in hospital
Diagnosis Related Group (DRG)	DRG	DRG in use on discharge date
	DRG_NoPOA	DRG in use on discharge date, calculated without Present On Admission (POA) indicators
	DRGVER	Grouper version in use on discharge date
Discharge quarter	DQTR	Coded: (1) Jan - Mar, (2) Apr - Jun, (3) Jul - Sep, (4) Oct - Dec
Discharge weights	DISCWT	Weight to discharges in AHA universe for national estimates. In 2000, the discharge weight DISCWTCHARGE should be used for estimates of total charges.
Discharge year	YEAR	Calendar year
Disposition of patient	DIED	Indicates in-hospital death: (0) did not die during hospitalization, (1) died during hospitalization

Type of Data Element	HCUP Name	Coding Notes
(discharge status)	DISPUNIFORM	Disposition of patient, uniform coding used beginning in 1998: (1) routine, (2) transfer to short term hospital, (5) other transfers, including skilled nursing facility, intermediate care, and another type of facility, (6) home health care, (7) against medical advice, (20) died in hospital, (99) discharged alive, destination unknown
	TRAN_OUT	Transfer Out Indicator: (0) not a transfer, (1) transferred out to a different acute care hospital, (2) transferred out to another type of health facility
Sex of patient	FEMALE	Indicates sex for KID beginning in 1998: (0) male, (1) female
Hospital information	HOSP_REGION	Region of hospital: (1) Northeast, (2) Midwest, (3) South, (4) West Prior to 2012, region of hospital is only available in the KID Hospital File.
	KID_STRATUM	Hospital stratum used for weights.
	HOSP_KID	KID hospital number (encrypted)
Indicates Emergency Department service	HCUP_ED	Indicator that discharge record includes evidence of emergency department (ED) services: (0) Record does not meet any HCUP Emergency Department criteria, (1) Emergency Department revenue code on record, (2) Positive Emergency Department charge (when revenue center codes are not available), (3) Emergency Department CPT procedure code on record, (4) Admission source of ED, (5) State-defined ED record; no ED charges available
Indicators	I10_BIRTH	ICD-10-CM birth indicator
	I10_DELIVERY	ICD-10-CM delivery indicator
	I10_INJURY	Injury ICD-10-CM diagnosis reported on record
	I10_MULTINJURY	Multiple ICD-10-CM injuries reported on record
	I10_SERVICELINE	ICD-10-CM/PCS hospital service line indicator
	PCLASS_ORPROC	ICD-10-PCS major operating room procedure indicator
Length of Stay	LOS	Length of stay, edited
Location of the patient	PL_NCHS	Urban–rural designation for patient's county of residence: (1) "Central" counties of metro areas >= 1 million population, (2) "Fringe" counties of metro areas >= 1 million population, (3) Counties in metro areas of 250,000 - 999,999 population, (4) Counties in metro areas of 50,000 - 249,999 population, (5) micropolitan counties, (6) non-core counties

Type of Data Element	HCUP Name	Coding Notes
Major Diagnosis Category (MDC)	MDC	MDC in use on discharge date
	MDC_NoPOA	MDC in use on discharge date, calculated without Present on Admission (POA) indicators
Median household income for patient's ZIP Code	ZIPINC_QRTL	Median household income quartiles for patient's ZIP Code: (1) quartile 1 [lowest income], (2) quartile 2, (3) quartile 3, (4) quartile 4 [highest income]. • For 2022, the median income quartiles are defined as: (1) $1–$55,999; (2) $56,000–$70,999; (3) $71,000–$93,999; and (4) $94,000 or more.
Payer information	PAY1	Expected primary payer, uniform: (1) Medicare, (2) Medicaid, (3) private including HMO, (4) self-pay, (5) no charge, (6) other
Procedure information	I10_PR1 – I10_PR25	Procedures, principal and secondary (ICD-10-CM/PCS)
	I10_NPR	Number of procedures coded on the original record
	I10_ORPROC	Major operating room procedure indicator: (0) no major operating room procedure, (1) major operating room procedure
	PRDAY1	Number of days from admission to principal procedure.
	PRDAY2 – PRDAY25	Number of days from admission to secondary procedures.
Race and ethnicity of Patient	RACE[11]	Race and ethnicity, uniform coding: (1) white, (2) black, (3) Hispanic, (4) Asian or Pacific Islander, (5) Native American, (6) other
Record identifier, synthetic	RECNUM	HCUP unique record number
Total Charges	TOTCHG	Total charges, edited

Return to Introduction

[11] RACE contains missing values on more than 7% of the records.

Table 2. Data Elements in the 2022 KID Hospital File

For prior years, refer to the KID Description of Data Elements page on the HCUP-US Web site or to previous versions of the KID Introduction.

Type of Data Element	HCUP Name	Coding Notes
Universe Counts	N_DISC_U	Number of universe discharges in the KID_STRATUM
	N_BRTH_U	Number of universe births in KID_STRATUM
	N_HOSP_U	Number of universe hospitals in KID_STRATUM
Sample Counts	S_DISC_U	Number of sampled discharges in the sampling stratum (KID_STRATUM or STRATUM)
	S_BRTH_U	Number of sample births in KID_STRATUM
	S_CHLD_U	Number of sample pediatric non-births in KID_STRATUM
	S_CMPB_U	Number of sample other newborns in KID_STRATUM
	S_UNCB_U	Number of sample normal newborns in KID_STRATUM
	S_HOSP_U	Number of sample hospitals in KID_STRATUM
Hospital Characteristics	KID_STRATUM	Hospital stratum used for weights
	HOSP_BEDSIZE	Bed size of hospital (STRATA): (1) small, (2) medium, (3) large
	H_CONTRL	Control/ownership of hospital (STRATA): (1) government, nonfederal, (2) private, non-profit, (3) private, invest-own
	HOSP_LOCTEACH	Location/teaching status of hospital (STRATA): (1) rural, (2) urban non-teaching, (3) urban teaching
	HOSP_REGION	Region of hospital (STRATA): (1) Northeast, (2) Midwest, (3) South, (4) West
	HOSP_KID	KID hospital number (encrypted)
Discharge Year	YEAR	Calendar year

Return to Introduction

Table 3. Data Elements in the 2022 KID Disease Severity Measures Files

For prior years, refer to the KID Description of Data Elements page on the HCUP-US Web site or to previous versions of the KID Introduction.

Type of Data Element	HCUP Name	Coding Notes
All Patient Refined DRG (3M)	APRDRG	All Patient Refined DRG
	APRDRG_Risk_ Mortality	All Patient Refined DRG: Risk of Mortality Subclass
	APRDRG_Severity	All Patient Refined DRG: Severity of Illness Subclass
Linkage Variables	HOSP_KID	KID hospital number (links to Hospital Weights file; does not link to previous years)
	RECNUM	HCUP record identifier (links to KID discharge level files; does not link to previous years)

Return to Introduction

Table 4: Data Elements in the 2022 KID Diagnosis and Procedure Groups Files

The Diagnosis and Procedure Groups file is available from 2006 to 2012; and is available again beginning with 2019 data, when data elements derived from the Clinical Classifications Software Refined (CCSR) for ICD-10-CM diagnoses, the CCSR for ICD-10-PCS procedures, and Procedure Classes Refined for ICD-10-CM are available in this file. This file is not available from 2016 because the ICD-10-CM/PCS versions of the AHRQ tools were still under development.

For prior years, refer to the KID Description of Data Elements page on the HCUP-US website or to previous versions of the KID Introduction.

Type of Data Element	HCUP Name	Coding Notes
Clinical Classifications Software Refined (CCSR) Category	DXCCSR_aaannn[c]	Indication that at least one ICD-10-CM diagnosis on the record is included in CCSR aaannn
	DXCCSR_DEFAULT_DX1	Default Clinical Classifications Software Refined (CCSR) for principal diagnosis
	DXCCSR_VERSION	Version of CCSR for ICD-10-CM diagnoses
	PRCCSR_aaannn[d]	Indication that at least one ICD-10-PCS procedure code on the record is included in CCSR aaannn
	PRCCSR_VERSION	Version of the CCSR for ICD-10-PCS procedures
Procedure Classes Refined	PCLASSn[e]	Procedure Classes Refined for ICD-10-PCS procedures
	PCLASS_VERSION	Version of the Procedure Classes Refined for ICD-10-PCS procedures
Linkage Data Elements	HOSP_KID	KID hospital number (links to Hospital Weights file; does not link to previous years)
	RECNUM	HCUP record identifier (links to KID discharge level files; does not link to previous years)

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; CCSR, Clinical Classifications Software Refined; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; ICD-10-CM/PCS, International Classification of Diseases, Tenth Revision, Clinical Modification/Procedure Coding System; KID, Kids' Inpatient Database.

[c] Where aaa denotes the body system and nnn denotes the CCSR number within the body system.

[d] Where aaa denotes the clinical domain and nnn denotes the CCSR number within the clinical domain.

[e] PCLASSn was also available on the KID through data year 2012 and was specific to the coding of ICD-9-CM procedures.

Return to Introduction

Appendix IV: Teaching Hospital Indicator Assignment

The following data elements from the American Hospital Association Annual Survey Database™ were used to assign the KID Teaching Hospital Indicator:

AHA Data Element Name	= Description [HCUP Data Element Name].
BDH	= Number of short-term hospital beds [B001H].
BDTOT	= Number of total facility beds [B001].
FTRES	= Number of full-time employees: interns & residents (medical & dental) [E125].
PTRES	= Number of part-time employees: interns & residents (medical & dental) [E225].
MAPP8	= Council of Teaching Hospitals (COTH) indicator [A101].
MAPP3	= Residency training approval by the Accreditation Council for Graduate Medical Education (ACGME) [A102].

Beginning with the 2000 KID, the following SAS code was used to assign the teaching hospital status indicator, HOSP_TEACH:

```
/*******************************************************/
/* FIRST ESTABLISH SHORT-TERM BEDS DEFINITION */
/*******************************************************/
IF BDH NE . THEN BEDTEMP = BDH ;      /* SHORT TERM BEDS */
ELSE IF BDH =. THEN BEDTEMP = BDTOT ; /* TOTAL BEDS PROXY */
/*******************************************************/
/* ESTABLISH IRB NEEDED FOR TEACHING STATUS */
/* BASED ON F-T P-T RESIDENT INTERN STATUS */
/*******************************************************/
IRB = (FTRES + .5*PTRES) / BEDTEMP ;
/*******************************************************/
/* CREATE TEACHING STATUS DATA ELEMENT */
/*******************************************************/
IF (MAPP8 EQ 1) OR (MAPP3 EQ 1) THEN HOSP_TEACH = 1 ;
ELSE IF (IRB GE 0.25) THEN HOSP_TEACH = 1 ;
ELSE HOSP_TEACH = 0 ;
```

Return to Introduction