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)

2006

These pages provide only an introduction to the KID package.

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

Issued June 2008

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

Phone: (866) 290-HCUP (4287)
Fax: (301) 427-1430
E-mail: hcup@ahrq.gov
Website: http://www.hcup-us.ahrq.gov

KID Data and Documentation Distributed by:
HCUP Central Distributor
Phone: (866) 556-4287 (toll-free)
Fax: (866) 792-5313
E-mail: HCUPDistributor@ahrq.gov

Table of Contents

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

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

All users of the KID must take the on-line Data Use Agreement (DUA) training session, sign a Data Use Agreement, and send a copy to AHRQ. †

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

- Will not use the data for any purpose other than research or aggregate statistical reporting.
- Will not re-release any data to unauthorized users.
- Will not identify or attempt to identify any individual. **Will not 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.**
- Will not link HCUP data to data from another source that identifies individuals.
- Will not report information that could identify individual establishments (e.g., hospitals).
- Will not use the data concerning individual establishments for commercial or competitive purposes involving those establishments.
- Will not use the data to determine rights, benefits, or privileges of individual establishments.
- Will not identify or attempt to identify any establishment when its identity has been concealed on the database.
- Will not contact establishments included in the data.
- Will not attribute to data contributors any conclusions drawn from the data.
- Will not use data elements from the proprietary severity adjustment software packages (3M APR-DRGs, HSS APS-DRGs, and Thomson Reuters Disease Staging) for any commercial purpose or to disassemble, decompile, or otherwise reverse engineer the proprietary software.
- Must acknowledge the "Healthcare Cost and Utilization Project, (HCUP)", as described in the Data Use Agreement, in reports.

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

† The on-line Data Use Agreement training session and the Data Use Agreement are available on the HCUP User Support (HCUP-US) Website at http://www.hcup-us.ahrq.gov.
‡ Specific provisions are detailed in the Data Use Agreement for Kids' Inpatient Database.

HCUP CONTACT INFORMATION

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

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

After completing the on-line training tool, Please submit signed Data Use Agreements to HCUP at:

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

Phone: (866) 290-HCUP (4287)
Fax: (301) 427-1430
Website: http://www.ahrq.gov/data/hcup/

For technical assistance:

Visit the HCUP-US Website at

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

Or send an e-mail to HCUP User Support at

hcup@ahrq.gov

Or contact the HCUP Central Distributor at

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

Fax: (866) 792-5313
E-mail: HCUPDistributor@ahrq.gov

WHAT'S NEW IN THE 2006 KIDS' INPATIENT DATABASE (KID)?

- Arkansas and Oklahoma joined the KID in 2006.
- A new companion discharge-level KID file contains data elements from AHRQ software tools designed to facilitate the use of the ICD-9-CM diagnostic and procedure information in the HCUP databases.
- The patient location four-category urban rural data element (PL_UR_CAT4) has been replaced by a six-category data element created by the National Center for Health Statistics (PL_NCHS2006). The six-category data element can be collapsed to four categories that approximate the original data element.
- The version 18 DRG and MDC are replaced with the version 24 DRG and MDC.
- The data element HFIPSSTCO was added to the Hospital Weights file. HFIPSSTCO contains the hospital's FIPS State and county code. For more information, visit the HCUP-US Web site.
- The National Association of Children's Hospitals and Related Institutions (NACHRI) teaching status indicator was not available for 2006. Therefore, for all hospitals, teaching status was determined using only information from the American Hospital Association (AHA) Annual Survey Database (Health Forum, LLC © 2007).
- Fourth quarter data from sampled hospitals in Massachusetts were unavailable for inclusion in the 2006 KID; however, we adjusted the data to account for missing cases. For details, see the section on the Final KID Sample in this document.
- The KID Introduction and the KID Design Report were combined and reorganized for 2006. Data tables, figures, and State-specific restrictions now appear as appendices for the combined document.
- 2006 KID Documentation is available exclusively on the HCUP User Support (HCUP-US) Website (http://www.hcup-us.ahrq.gov) and is no longer included on the KID CD-ROMs. This ensures that documentation for your data will always be the most recent and up-to-date version.
- Users must complete an on-line Data Use Agreement training tool prior to receiving the data.

UNDERSTANDING THE KID
• This document, Introduction to the KID, 2006, summarizes the content of the KID and describes the development of the KID sample and weights. • Cumulative information for all previous years is included to provide a longitudinal view of the database. • Important considerations for data analysis are highlighted and references to detailed reports are provided. • In-depth documentation for the KID is available on the HCUP User Support (HCUP-US) Website (www.hcup-us.ahrq.gov). Please refer to detailed documentation before using the data.

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), formerly the Agency for Health Care Policy and Research.

The KID is the only dataset on hospital use, outcomes, and charges designed to study children's use of hospital services in the United States. The KID is a sample of 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.

The KID contains charge information on all patients, regardless of payer, including persons covered by private insurance, Medicaid, Medicare, and the uninsured. The KID's large sample size enables analyses of rare conditions, such as congenital anomalies and uncommon treatments, such as organ transplantation. 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.

Inpatient stay records in the KID include clinical and resource use information typically available from discharge abstracts. Discharge weights are provided for calculating national estimates. The KID can be linked to hospital-level data from the American Hospital Association's Annual Survey Database (Health Forum, LLC © 2007) and county-level data from the Bureau of Health Professions' Area Resource File, except in those States that do not allow the release of hospital identifiers.

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.

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) Website at http://www.hcup-us.ahrq.gov.

INTRODUCTION TO THE HCUP KIDS' INPATIENT DATABASE (KID)

Overview of KID Data

The Healthcare Cost and Utilization Project (HCUP) Kids' Inpatient Database (KID) was developed to enable analyses of hospital utilization by children across the United States. The target universe includes pediatric discharges from community, non-rehabilitation hospitals in the United States.[1]

The sampling frame is limited to pediatric discharges from community, non-rehabilitation hospitals in the participating HCUP Partner States shown in Figure 1 of Appendix I.

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:

- Uncomplicated in-hospital births (HOSPBRTH = 1 and UNCBRTH = 1)
- Complicated in-hospital births (HOSPBRTH = 1 and UNCBRTH = 0)
- All other pediatric cases (HOSPBRTH = 0).

In-hospital births (HOSPBRTH = 1) are identified by any principal or secondary diagnosis code 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. Uncomplicated births (UNCBRTH = 1) have a Diagnosis Related Group (DRG) equal to 391 indicating "Normal Newborn."

Unlike the HCUP Nationwide Inpatient Sample (NIS), the KID does not involve a two-stage sampling procedure. Instead, the KID includes a sample of pediatric discharges from all hospitals in the sampling frame – the State Inpatient Databases (SID) that agreed to participate in the KID). For sampling, pediatric discharges are stratified by uncomplicated in-hospital birth, complicated in-hospital birth, 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% of uncomplicated in-hospital births and 80% of complicated in-hospital births and other pediatric cases from each frame hospital.

[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.

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. These were the same characteristics used to define the NIS sampling strata (ownership/control, bedsize, teaching status, rural/urban location, and U.S. region), with the addition of a stratum for freestanding children's hospitals. To create weights, if there were fewer than two frame hospitals, 30 uncomplicated births, 30 complicated births, 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 Website (http://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 2006 KID, and discusses data analysis issues. This document highlights cumulative information for all previous KID releases to provide a longitudinal view of the database. 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. See Table 3 of Appendix I for a summary of KID releases. Each release of the KID includes:

- Data in fixed-width ASCII format on CD-ROM.
- 2 million to 3 million pediatric inpatient records per year.
- 2,500 to 3,500 hospitals per year (all SID hospitals with pediatric discharges).
- Discharge-level weights to calculate national estimates for discharges.
- Hospital File to link the KID to data from the AHA Annual Survey Database.
- 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.

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 data sources 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 CD-ROM

The KID is contained on one CD-ROM that include fixed-width ASCII formatted data files and a README.TXT file describing how to access related KID documentation on the HCUP-US Website (http://www.hcup-us.ahrq.gov).

The CD-ROM contains:

Inpatient Core File: The Core file contains pediatric discharges sampled from community, non-rehabilitation hospitals in participating HCUP States. The unit of observation is an *inpatient stay record*. The Core file contains data elements for linkage, patient demographics, clinical information, and payment information. Sample weights for the three types of records, uncomplicated in-hospital births, complicated in-hospital births, and all other pediatric cases, are calculated separately by stratum and are added to each discharge in the Core File, as appropriate, so that only one discharge weight variable (DISCWT) is needed. See Table 1 of Appendix III for a list of data elements in the Inpatient Core File. This file is available in all years of the KID.

Hospital File: The hospital-level file contains one observation for each hospital included in the KID and contains variance estimation data elements, linkage data elements, and data elements that describe basic characteristics about hospitals. The unit of observation is the *hospital*. The HCUP hospital identifier (HOSPID) provides the linkage between the KID Inpatient Core file and the Hospital file. See Table 2 of Appendix III for a list of data elements in the Hospital File. This file is available in all years of the KID.

Disease Severity Measures File: This discharge-level file contains information from four different sets of disease severity measures. Information from the severity file is to be used in conjunction with the inpatient Core file. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (RECNUM) provides the linkage between the Core file and the Disease Severity Measures files. See Appendix III, Table 3 for a list of data elements in the Severity Measures Files. This file is available beginning with the 2003 KID.

Diagnosis and Procedure Groups Files: These discharge-level files contain data elements from AHRQ software tools designed to facilitate the use of the ICD-9-CM diagnostic and procedure information in the HCUP databases. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (KEY) provides the linkage between the Core files and the Diagnosis and Procedure Groups files. Table 4 in Appendix III contains a list of data elements in the Diagnosis and Procedure Groups files. These files are available beginning with the 2006 KID.

On the HCUP-US Website (http://www.hcup-us.ahrq.gov), KID purchasers can access complete file documentation, including variable notes, file layouts, summary statistics, and related technical reports. Similarly, purchasers can also download SAS, SPSS, and Stata load programs. Available online documentation and supporting files are detailed in Appendix I, Table 4.

KID Data Elements

The KID contains two types of data: inpatient stay core records and hospital information. Appendix III identifies the data elements in each KID file:

- Table 1 for the Inpatient Core files (record = inpatient stay)
- Table 2 for the Hospital Weights files (record = hospital)
- Table 3 for the Disease Severity Measures files (record = inpatient stay). This file was added beginning with the 2003 KID.
- Table 4 for the Diagnosis and Procedure Groups files (record = inpatient stay). This file was added beginning with the 2006 KID.

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://hcup-us.ahrq.gov) for comprehensive information about data elements and the files

Getting Started

The KID data files are provided on one CD-ROM. Comprehensive documentation for the KID files is available on the HCUP-US Website (http://hcup-us.ahrq.gov).

KID Data Files

In order to load KID data onto your PC, you will need about five gigabytes of space available. Because of the size of the files, the data are distributed as self-extracting PKZIP compressed files. To decompress the data, you should follow these steps:

1. Create a directory for the KID on your hard drive.
2. Copy the self-extracting data files from the KID Data Files CD-ROM into the new directory.
3. Unzip each file by running the corresponding *.exe file.
   - Type the file name within DOS or click on the name within Windows Explorer.
   - Edit the name of the "Unzip To Folder" in the WinZip Self-Extractor dialog to select the desired destination directory for the extracted file.
   - Click on the "Unzip" button.

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

KID Documentation

KID documentation files on the HCUP-US Website (http://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 (http://www.hcup-us.ahrq.gov). The section labeled “KIDS’ Inpatient Database (KID) is specific to the KID.

Table 4 in Appendix I details both the KID related reports and the comprehensive KID documentation available on HCUP-US.

HOW TO USE THE KID FOR DATA ANALYSIS

This section provides a brief synopsis of special considerations when using the KID. For more details, refer to the comprehensive documentation on the HCUP-US Website (http://hcup-us.ahrq.gov/).

- 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 on-line Data Use Agreement Training Tool available on the HCUP-US Website (http://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.

Calculating National Estimates

- To produce national estimates, use one of the following discharge weights to weight discharges in the KID Core files to pediatric discharges from all U.S. community, non-rehabilitation hospitals. The name of the discharge weight data element depends on the year of data and the type of analysis. **In order to produce national estimates, you MUST use discharge weights.**

KID Data Year	Name of Discharge Weight on the **Core File** to Use for Creating Nationwide Estimates
2003 forward	• DISCWT for all analyses
2000	• DISCWT to create nationwide estimates for all analyses except those that involve total charges. • DISCWTCHARGE to create nationwide estimates of total charges.
1997	• DISCWT_U for all analyses

- Similar to the NIS, 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 Nationwide Inpatient Sample Variances* on the HCUP-US Website (www.hcup-us.ahrq.gov). A report specific to the KID, *Calculating Kids' Inpatient Database (KID) Variances,* is also available on www.hcup-us.ahrq.gov.

- The KID Comparison Reports (available on www.hcup-us.ahrq.gov) assess the accuracy of KID estimates. No comparison report was created for the 2000 KID. The updated report for

the current KID will be posted on the HCUP-US Website (www.hcup-us.ahrq.gov) as soon as it is completed.

- When creating national estimates, it is a good idea to check your estimates against other data sources, if available. For example, the National Hospital Discharge Survey (http://www.cdc.gov/nchs/products/pubs/pubd/series/ser.htm#sr13) can provide benchmarks against which to check your national estimates for hospitalizations with more than 5,000 discharges.

- To ensure that you are using the weights appropriately and calculating estimates and variances accurately, you can also use HCUPnet, the free online query system (http://www.hcupnet.ahrq.gov). 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

- When studying trends over time 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. In order to facilitate analysis of trends using multiple years of KID data, an alternate set of KID discharge and hospital weights for the 1997 HCUP KID were developed. These alternative weights were calculated in the same way as the weights for the 2000 and later years of the KID. The report, *Using the Kids' Inpatient Database (KID) to Estimate Trends*, includes details regarding the alternate weights and other recommendations for trends analysis. Both the KID trends report and the alternate weights are available on the HCUP-US Website under Methods Series (http://www.hcup-us.ahrq.gov/reports/methods.jsp).

- Short-term rehabilitation hospitals are included in the 1997 KID, but are excluded from later 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. The KID-Trends weights account for this change in KID sampling.

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 number of missing cases. Summary statistics for the entire KID are provided on the Summary Statistics page of the HCUP-US Website (http://hcup-us.ahrq.gov/db/nation/kid/kidsummarystats.jsp). When you detect anomalies (such as large numbers of missing cases), perform descriptive statistics by State for that variable to detect if there are State-specific differences. Sometimes 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 State data source. 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 a number of States; thus, national estimates using RACE should be interpreted and reported with caveats. Check the documentation and run frequencies by State to identify if a data element is not available in one or more States.

- 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. Be sure to read State-specific notes for each data element that you use in your analysis – this information can be found on the Description of Data Elements page on the HCUP-US Website (http://hcup-us.ahrq.gov/db/nation/kid/kiddde.jsp).

- Data elements with "_X" suffixes contain State-specific coding (i.e., these data elements are provided by the data sources and have not been altered in any way). For some data elements (e.g., LOS_X and TOTCHG_X) this means that no edit checks have been applied. For other data elements (e.g., PAY1_X), the coding is specific to each State and may not be comparable to any other State.

ICD-9-CM Diagnosis and Procedure Codes

- ICD-9-CM diagnosis and 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 codes change every October as new codes are introduced and some codes are retired. See the Conversion Table at http://www.cdc.gov/nchs/datawh/ftpserv/ftpicd9/ftpicd9.htm which shows ICD-9-CM code changes over time. **It is critical to check all ICD-9-CM code used for analysis to ensure the codes are in effect during the time period studied.**

- Although the KID contains up to 15 diagnoses and 15 procedures, the number of diagnoses and procedures varies by State. Some States provide as many as 30 diagnoses and 21 procedures, while other States provide as few as 9 diagnoses and 6 procedures. Because very few cases have more than 15 diagnoses or procedures, the diagnosis and procedure vectors were truncated to save space in the KID data files. Two variables are provided which tell you exactly how many diagnoses and procedures were on the original records (NDX and NPR).

- The collection and reporting of external cause of injury (E codes) varies greatly across States. Some States have laws or mandates for the collection of E codes; others do not. Some States do not require hospitals to report E codes in the range E870-E879 - "misadventures to patients during surgical and medical care" - which means that these occurrences will be underreported. Beginning with the 2003 KID, E codes have been separated from the other diagnoses stored in DX1-DX15 and placed in ECODE1-ECODE4. Be sure to read the State-specific notes on diagnoses for more details; this information can be found on the Description of Data Elements page on the HCUP-US Website (http://hcup-us.ahrq.gov/db/nation/kid/kiddde.jsp).

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. There are several techniques available to help overcome this bias. One strategy is to use imputation to replace missing values with acceptable values. Another strategy is to use sample weight adjustments to compensate for missing values.[1] 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.

On the other hand, if the cases with and without missing values are assumed to be similar with respect to their outcomes, no adjustment may be necessary for estimates of means and rates. This is because the non-missing cases would be representative of the missing cases. However, some adjustment may still be necessary for the estimates of totals. Sums of data elements (such as aggregate charges) containing missing values would be incomplete because cases with missing values would be omitted from the calculations.

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*. This report is available on the HCUP-US Website at http://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.

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 is in contrast to 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.[2] 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 Website at http://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 Website: http://www.hcp.med.harvard.edu/statistics/survey-soft/.

The KID database includes a Hospital file with variables required to calculate finite population statistics. The file includes hospital identifiers (Primary Sampling Units or PSUs), stratification variables, 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 variables are available for the entire universe of hospitals, including those outside the sampling frame. For instance, the variables 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 a number of 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 variables, 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), the KID does not involve sampling hospitals. Instead, the KID includes a sample of pediatric discharges from all hospitals in the sampling frame. For the sampling, pediatric discharges in all participating States are stratified by uncomplicated in-hospital birth, complicated in-hospital birth, 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% of uncomplicated in-hospital births and 80% of complicated in-hospital births and other 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. These were the same characteristics used to define the NIS sampling strata (ownership/control, bedsize, teaching status, rural/urban location, and U.S. region), with the addition of a stratum for freestanding children's hospitals. If there were fewer than two frame hospitals, 30 uncomplicated births, 30 complicated births, 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.

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.) Table 2 (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 HCUP-US Website. 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, as long as they were in operation during some part of the calendar year.

Stratification Variables

For the purpose of calculating discharge weights, we post-stratified hospitals on six characteristics contained in the AHA hospital files. These were the same characteristics used to define the HCUP Nationwide Inpatient Sample (NIS) sampling strata, with the addition of a stratum for stand-alone children's hospitals. The definitions of some of the NIS strata were revised for 1998 and subsequent data years, and we used the revised strata 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 Website at http://www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp.)

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

1. *Geographic Region – Northeast, Midwest, West, and South.* This is an important stratification variable 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 highlights the KID States by region, and Table 5 lists the States that comprise each region. Both can be found in Appendix I.

2. *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.

3. *Location – urban or rural.* Government payment policies often differ according to this designation. Also, rural hospitals are generally smaller and offer fewer services than urban hospitals. Beginning with the 2006 KID, we changed the classification of urban or rural hospital location for the sampling strata to use the newer Core Based Statistical Area (CBSA) codes rather than the older Metropolitan Statistical Area (MSA) codes. The CBSA groups are based on 2000 Census data, whereas the MSA groups were based on 1990 Census data. Also, the criteria for classifying the counties differ. For more information on the difference between CBSAs and MSAs, refer to the U.S. Census Bureau Website (http://www.census.gov/population/www/estimates/metroarea.html).

   Previously, we classified hospitals in an MSA as urban hospitals, while we classified hospitals outside a MSA as rural hospitals. Beginning with the 2006 KID, 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.

4. *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 National Association of Children's Hospitals and Related Institutions (NACHRI) was based on an indicator provided by NACHRI. The NACHRI teaching status indicator was not available for 2006. Therefore, for all hospitals, teaching status was determined using only information from the AHA Annual Survey Database.

   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:

   - 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 higher.[3]

5. *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 6 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 2% or less than 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.

6. *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, severity of illness, and available services between children's hospitals and other hospitals. Data from NACHRI were used to help verify and correct the AHA list of children's hospitals. Children's units in general hospitals were not stratified as children's hospitals.

Hospital 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 7 of Appendix I shows the number of AHA, HCUP SID, and KID hospitals by State. In most cases, the difference between the universe and the frame represents the difference between the number of community, non-rehabilitation hospitals in the 2006 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.

The largest discrepancy between HCUP data and AHA data is in Texas, as is evident in Table 7 of Appendix I. Certain Texas State-licensed hospitals are exempt from statutory reporting requirements. Exempt hospitals include:

- Hospitals that do not seek insurance payment or government reimbursement
- Rural providers.

The Texas statute that exempts rural providers from the requirement to submit data defines a hospital as a rural provider if it:

(I) Is located in a county that:

(A) Has a population estimated by the United States Bureau of the Census to be not

more than 35,000 as of July 1 of the most recent year for which county population estimates have been published; or

(B) Has a population of more than 35,000, but does not have more than 100 licensed hospital beds and is not located in an area that is delineated as an urbanized area by the United States Bureau of the Census; and

(II) Is not a State-owned hospital or a hospital that is managed or directly or indirectly owned by an individual, association, partnership, corporation, or other legal entity that owns or manages one or more other hospitals.

These exemptions apply primarily to smaller rural public hospitals and, as a result, these facilities are less likely to be included in the sampling frame than other Texas hospitals. While the number of hospitals omitted appears sizable, those available for the KID include more than 90% of inpatient discharges from Texas universe hospitals because excluded hospitals tended to have relatively few discharges.

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.

Hospital 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 and agree to contribute to the KID.

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, if available, because not all States contribute data to the KID. For example, the National Hospital Discharge Survey (http://www.cdc.gov/nchs/about/major/hdasd/nhds.htm) can provide benchmarks against which to check your national estimates for hospitalizations with more than 5,000 cases.

The *KID Comparison Report* assesses the accuracy of KID estimates by providing a comparison of the KID with other data sources. The most recent report is available on the HCUP-US Website (http://www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp).

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 uncomplicated births and complicated births. We sampled uncomplicated births at a lower rate because they have little variation in their outcomes.

To determine the best way to identify in-hospital births, we ran cross-tabulations of different

combinations of variables on all cases that had any of the following possible birth indicators: age of zero days (AGEDAY=0), neonatal diagnosis (NEOMAT>=2), neonatal Major Diagnostic Category (MDC 15), or admission type of birth (ATYPE=4).[4] Based on reviews of the cross-tabulations, the MDC 15 DRG definitions, and ICD-9-CM birth diagnosis codes, the following screen was devised for births: an in-hospital birth diagnosis code (any diagnosis code in the range V3000 - V3901 with a fourth digit of zero, indicating born in the hospital, and a fifth digit of zero or one, indicating delivered without mention of cesarean delivery, or delivered by cesarean delivery), without an admission source of another hospital or health facility (ASOURCE not equal to 2 or 3).

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.

Uncomplicated, in-hospital births are identified as cases that meet the above screen and are in DRG 391, "Normal Newborn." In the KID, a small percentage of the cases in DRG 391 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 actually 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 use the NIS community hospital universe and strata definitions for the KID. We revised some of the NIS hospital universe and strata definitions for 1998 and subsequent data years, and we used these revised definitions beginning with the 2000 KID. These changes included:

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

A full description of the evaluation and revision of the NIS sampling strategy for 1998 and subsequent data years can be found in the special report, *Changes in NIS Sampling and Weighting Strategy for 1998*. This document is available on the HCUP-US Website at http://www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp.

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 uncomplicated in-hospital birth, complicated in-hospital birth, and pediatric non-birth. 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% of uncomplicated in-hospital births and 80% of other pediatric cases from each frame hospital.

It should be observed that the NIS includes 100% of the discharges from hospitals in the NIS sample. Consequently, in the NIS outcomes can be estimated without sampling error for individual hospitals that are identified in the sample. However, the KID includes fewer than 100% 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, we post-stratified hospitals on six characteristics contained in the AHA hospital files. These were the same characteristics used to define the NIS sampling strata, with the addition of a stratum for freestanding children's hospitals. We also stratified the KID discharges according to whether the discharge was an uncomplicated in-hospital birth, a complicated in-hospital birth, or a non-newborn pediatric discharge. If there were fewer than two frame hospitals, 30 uncomplicated births, 30 complicated births, and 30 non-birth pediatric discharges sampled in a stratum, we merged that stratum with an "adjacent" stratum containing hospitals with similar characteristics.

The discharge weights were created by stratum, in proportion to the number of AHA discharges for newborns and non-newborns. Refer to the report *Design of the HCUP Kids' Inpatient Database (KID), 1997* for a discussion of the analysis and development of the KID weighting scheme. This report is available on the on the HCUP-US Website at http://www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp.

We used NACHRI data to help verify and correct the AHA list of children's hospitals in the target universe. Many of these children's hospitals are units of larger institutions (AHA hospital type 10). Consequently, we do not have separate reporting for them either in the AHA survey or in the HCUP SID. However, data analysts may find it useful to identify hospitals that contain children's units, which can be accomplished using the NACHTYPE variable in the KID.

Discharge Weights

The discharge weights usually are constant for all discharges of the same type (uncomplicated in-hospital birth, complicated in-hospital birth, and other pediatric discharge) 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 complicated and uncomplicated):

- 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 uncomplicated birth strata, R_k is the frame sampling rate for uncomplicated in-hospital births calculated as the sum of the adjusted number of uncomplicated births sampled divided by the sum of the adjusted number of uncomplicated births in the restricted frame.
- In the complicated birth strata, R_k is the frame sampling rate for complicated in-hospital births.

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.

Q_i 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.

Uncomplicated in-hospital births were sampled at a lower rate than other discharges because the variation in hospital outcomes for uncomplicated births is considerably less than that for other pediatric cases and because we expect research to focus much more on other pediatric patients. We sampled uncomplicated births at the nominal rate of 10% and sampled other pediatric discharges (complicated newborns and other pediatric cases) at the nominal rate of 80% from the discharges available in the (restricted) frame. To avoid rounding errors in the weights calculation, the actual sampling rate for a discharge type (uncomplicated in-hospital birth, complicated in-hospital birth, or non-birth 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 uncomplicated and complicated births. Therefore, the weights in the uncomplicated birth strata implicitly assume that the proportion of births that are

uncomplicated in the frame is representative of the proportion of births that are uncomplicated in the population for each stratum. A similar assumption is made for complicated newborns.

Similarly, the non-birth AHA discharge counts include all non-birth discharges, not just non-birth pediatric discharges. 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 discharges 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 extrapolate 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.

THE FINAL KID SAMPLE

In Appendix I, we present tables and figures that summarize the final KID sample. Table 8 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 discharges, including births
- Non-rehabilitation community hospitals in the SID and associated pediatric discharges
- Hospitals and pediatric discharges included in the KID.

Table 9 displays the unweighted and weighted number of uncomplicated births, complicated births, and pediatric non-births by hospital type in the KID.

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

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.

Because the KID sampling frame has a disproportionate representation of the more populous States and contains hospitals with more annual discharges, its comprehensiveness in terms of discharges is higher. Figure 3 summarizes the estimated U.S. population by geographic region on July 1, 2006. 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.

And, Figure 4 presents the number of discharges in the KID for each State in the sampling frame for 2006.

Special consideration was needed to handle the Massachusetts data in the 2006 KID. Fourth quarter data from sampled hospitals in Massachusetts were unavailable for inclusion in the 2006 KID. To account for the missing quarter of data, we sampled one fourth of the Massachusetts KID discharges from the first three quarters and modified the records to represent the fourth quarter. To ensure a representative sample, we sorted the Massachusetts KID discharges by hospital, discharge quarter, Clinical Classifications Software (CCS) diagnosis group for the principal diagnosis, gender, age, and a random number before selecting every fourth record. The following describes the adjustments made to the selected Massachusetts KID records:

1. We relabeled the discharge quarter (DQTR) to four and saved the original discharge quarter in a new data element (DQTR_X).
2. We adjusted the admission month (AMONTH) by the number of months corresponding to the change in the discharge quarter.
3. We adjusted the total charges (TOTCHG and TOTCHG_X) using quarter-specific adjustment factors calculated as the mean total charges in the fourth quarter for all Northeastern KID States (excluding Massachusetts) divided by the mean total charges in the first, second, or third quarter for all Northeastern KID States (excluding Massachusetts).

We then adjusted the discharge weights for the Massachusetts records to appropriately account for the shifting of quarter one through three discharges to quarter four.

Appendix I: Tables and Figures

Table 1. Data Sources

State	Data Organization
AR	Arkansas Department of Health & Human Services
AZ	Arizona Department of Health Services
CA	Office of Statewide Health Planning & Development
CO	Colorado Hospital Association
CT	Chime, Inc.
FL	Florida Agency for Health Care Administration
GA	Georgia Hospital Association
HI	Hawaii Health Information Corporation
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
MA	Division of Health Care Finance and Policy
MD	Health Services Cost Review Commission
MI	Michigan Health & Hospital Association
MN	Minnesota Hospital Association
MO	Hospital Industry Data Institute
NC	North Carolina Department of Health and Human Services
NE	Nebraska Hospital Association
NH	New Hampshire Department of Health & Human Services
NJ	New Jersey Department of Health & Senior Services
NV	Nevada Department of Health and Human Services

State	Data Organization
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
RI	Rhode Island Department of Health
SC	South Carolina State Budget & Control Board
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 & Family Services
WV	West Virginia Health Care Authority

Return to Introduction

Table 2. Summary of KID Data Sources, Hospitals, and Inpatient Stays, 1997, 2000, 2003, and 2006

	2006	2003	2000	1997
Number of States	38	36	27	22
Data Sources	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)*	AZ CA CO CT FL GA HI IA IL IN KS KY MD MA 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)*	AZ CA CO CT FL GA HI IA KS KY MD MA 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)*	AZ CA CO CT FL GA HI IL IA KS MD MA MO NJ NY OR PA SC TN UT WA WI
Hospitals	Community, *non-rehabilitation* hospitals	Community, *non-rehabilitation* hospitals	Community, *non-rehabilitation* hospitals	Community hospitals, *including rehabilitation hospitals*
Hospital Universe[5]	5,124	4,836	4,839	5,113
Number of KID Hospitals	3,739	3,438	2,784	2,521
Hospital identifiers	Available for 24 out of 38 States	Available for 23 out of 36 States	Available for 19 out of 27 States	None – only general descriptors of hospital types
Definition of pediatric discharges	Age at admission of 20 years or less	Age at admission of 20 years or less	Age at admission of 20 years or less	Age at admission of 18 years or less
Number of pediatric discharges (unweighted)	3,131,324	2,984,129	2,516,833	1,905,797
Number of pediatric discharges (weighted)	7,558,812	7,409,162	7,291,032	6,657,326

Figure 1. KID States, by Region

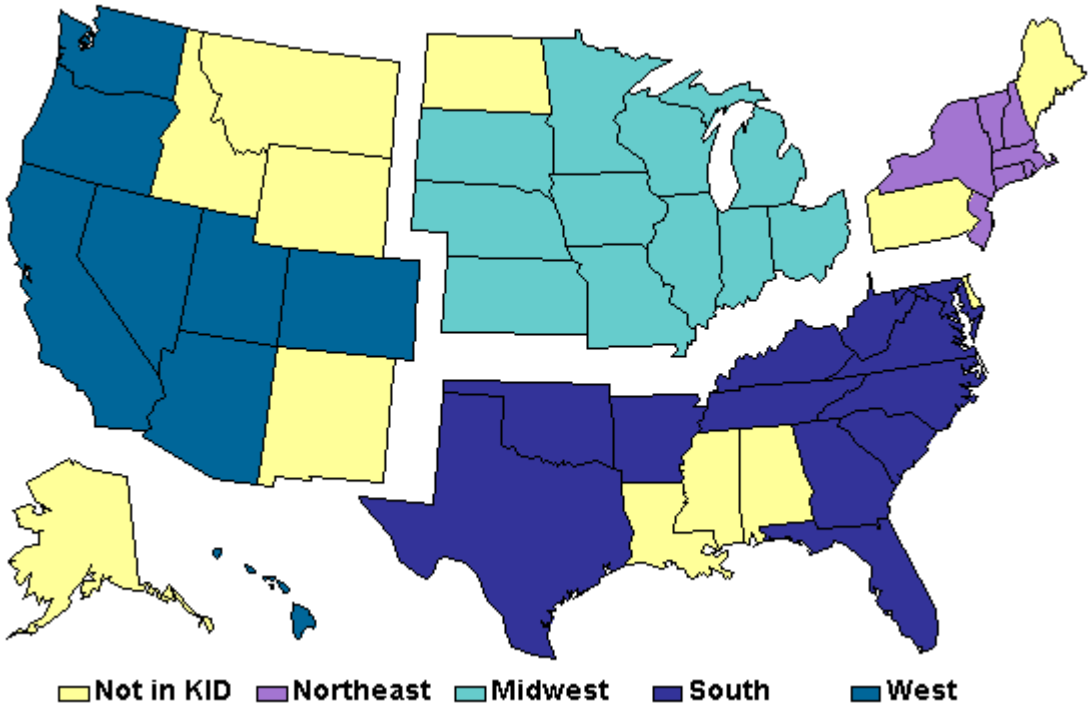


Table 3. Summary of KID Releases

Data from	Media/format options	Structure of Releases
▪ 1997 ▪ 22 States	On CD-ROM, in ASCII format	1 year of data on one CD, compressed files
▪ 2000 ▪ 27 States ▪ 2003 ▪ 36 States		Beginning in 2003, a companion file with four different sets of severity measures
▪ 2006 ▪ 38 States		Beginning in 2006, a companion file with diagnosis and procedure groups

Return to Introduction

Table 4. KID Related Reports and Database Documentation Available on HCUP-US

Restrictions on the Use of the KID
- Data Use Agreement for the KID

Description of the KID Files
- Introduction to the KID, 2006 – *this document*
- HCUP Quality Control Procedures – describes procedures used to assess data quality
- File Specifications – details data file names, number of records, record length, and record layout
- Sources of KID Data and State-Specific Restrictions (*included in this document beginning 2006*) – identifies the KID data sources and restrictions on sampling and the release of data elements

Availability of Data Elements
- Availability of KID data elements from 1988-2006

Description of Data Elements in the KID
- Description of Data Elements – details uniform coding and State-specific idiosyncrasies
- Summary Statistics – lists means and frequencies on nearly all data elements
- KID Severity Measures – provides detailed documentation on the different types of measures
- HCUP Coding Practices – describes how HCUP data elements are coded
- HCUP Hospital Identifiers – explains data elements that characterize individual hospitals

Corrections to the KID
- Information on corrections to the KID data sets
- Link to KID Trends Weights Files

Load Programs
Programs to load the ASCII data files into statistical software:
- SAS
- SPSS
- Stata

HCUP Tools: Labels and Formats
- Overview of Clinical Classifications Software (CCS), a categorization scheme that groups ICD-9-CM diagnosis and procedure codes into mutually exclusive categories
- Label file for CCS categories
- Label file for multiple versions of Diagnosis Related Groups (DRGs) and Major Diagnostic Categories (MDC)
- KID SAS format library program to create value labels

KID Related Reports
Links to HCUP-US page with various KID related reports such as the following:
- Design of the Kids' Inpatient Databases for 1997, 2000 and 2003 (*included in this document beginning 2006*)
- Changes in NIS Sampling and Weighting Strategy for 1998
- Calculating KID Variances
- File Composition by State
- KID Trends Report
- KID Comparison Reports
- HCUP E-Code Evaluation Report

HCUP Supplemental Files
- Cost-to-Charge Ratio files
- Hospital Market Structure files

Return to Introduction

Table 5. States, by Region

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

Return to Introduction

Table 6. 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+

Return to Introduction

Table 7. Number of AHA, HCUP SID, and KID Hospitals, by State, 2006[6]

State	AHA Universe Hospitals	SID Community, Non-Rehabilitation Hospitals	SID Community, Non-Rehabilitation Hospitals with Pediatric Discharges	KID Sampling-Frame Hospitals	KID Sample Hospitals
Non-Frame	840	0	05	0	0
Arizona	70	68	66	66	66
Arkansas	85	85	79	79	79
California	360	353	349	349	349
Colorado	75	70	69	69	69
Connecticut	34	29	29	28	28
Florida	203	200	195	195	194
Georgia	152	148	144	96	96
Hawaii	23	21	19	13	13
Iowa	118	117	116	116	116
Illinois	187	186	185	185	185
Indiana	122	109	108	108	108
Kansas	144	126	123	123	123
Kentucky	105	101	100	100	100
Massachusetts	78	67	65	65	65
Maryland	47	47	47	47	47
Michigan	149	140	134	100	99
Minnesota	132	124	124	124	124
Missouri	123	119	117	117	117
North Carolina	118	114	111	111	110
Nebraska	87	85	85	79	79
New Hampshire	26	26	26	26	26
New Jersey	78	77	73	73	73
Nevada	35	31	31	31	31
New York	200	200	198	198	197
Ohio	187	157	155	155	155
Oklahoma	132	127	118	116	116
Oregon	58	57	57	57	57
Rhode Island	11	11	11	11	11
South Carolina	61	57	57	53	53
South Dakota	57	47	44	42	41
Tennessee	131	111	110	110	109
Texas	476	388	340	340	339
Utah	46	42	42	42	42
Virginia	83	81	81	46	45
Vermont	14	14	13	13	13
Washington	89	88	85	85	85
Wisconsin	132	130	128	128	128

Table 7. Number of AHA, HCUP SID, and KID Hospitals, by State, 2006[6]

State	AHA Universe Hospitals	SID Community, Non-Rehabilitation Hospitals	SID Community, Non-Rehabilitation Hospitals with Pediatric Discharges	KID Sampling-Frame Hospitals	KID Sample Hospitals
West Virginia	56	56	52	52	51
Total	5,124	4,009	3,886	3,748	3,739

Return to Introduction

Table 8. Number of Hospitals and Discharges in the AHA Universe, SID, and KID, by Hospital Type, 2006

	AHA Universe		SID		KID	
Hospital Type	**Hospitals**	**Discharges (Including Births)**	**Hospitals with Pediatric Discharges**	**Pediatric Discharges**	**Hospitals**	**Pediatric Discharges**
Not a Children's Hospital	5,045	38,881,991	3,836	6,027,890	3,694	2,841,194
Children's Hospital	79	568,225	50	414,510	45	290,130
Total	5,124	39,450,216	3,886	6,442,400	3,739	3,131,324

Table 9. 2006 KID Discharges, by Hospital Type

Hospital Type	Uncomplicated Births	Complicated Births	Pediatric Non-Births	Total Pediatric Discharges
Unweighted:				
Not a Children's Hospital	255,908	740,183	1,845,103	2,841,194
Children's Hospital	707	3,405	286,018	290,130
Total	256,615	743,588	2,131,121	3,131,324
Weighted:				
Not a Children's Hospital	3,002,918	1,093,112	2,919,177	7,015,207
Children's Hospital	6,525	3,932	533,148	543,605
Total	3,009,443	1,097,044	3,452,325	7,558,812

Return to Introduction

Figure 2. Number of Hospitals in the 2006 AHA Universe, SID, and KID, by Region

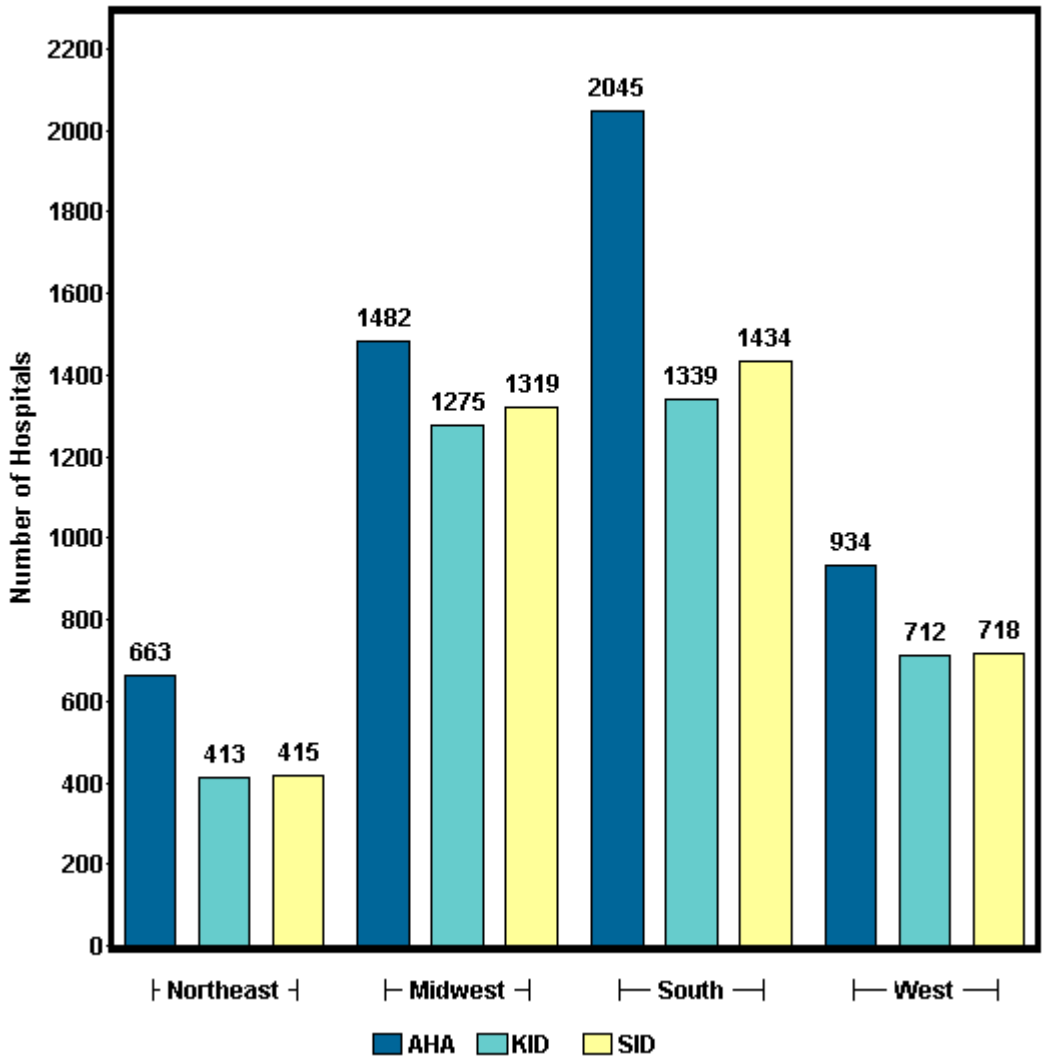


Figure 3. Percentage of U.S. Population in 2006 KID States, by Region
Calculated using the estimated U.S. population on July 1, 2006.[7]

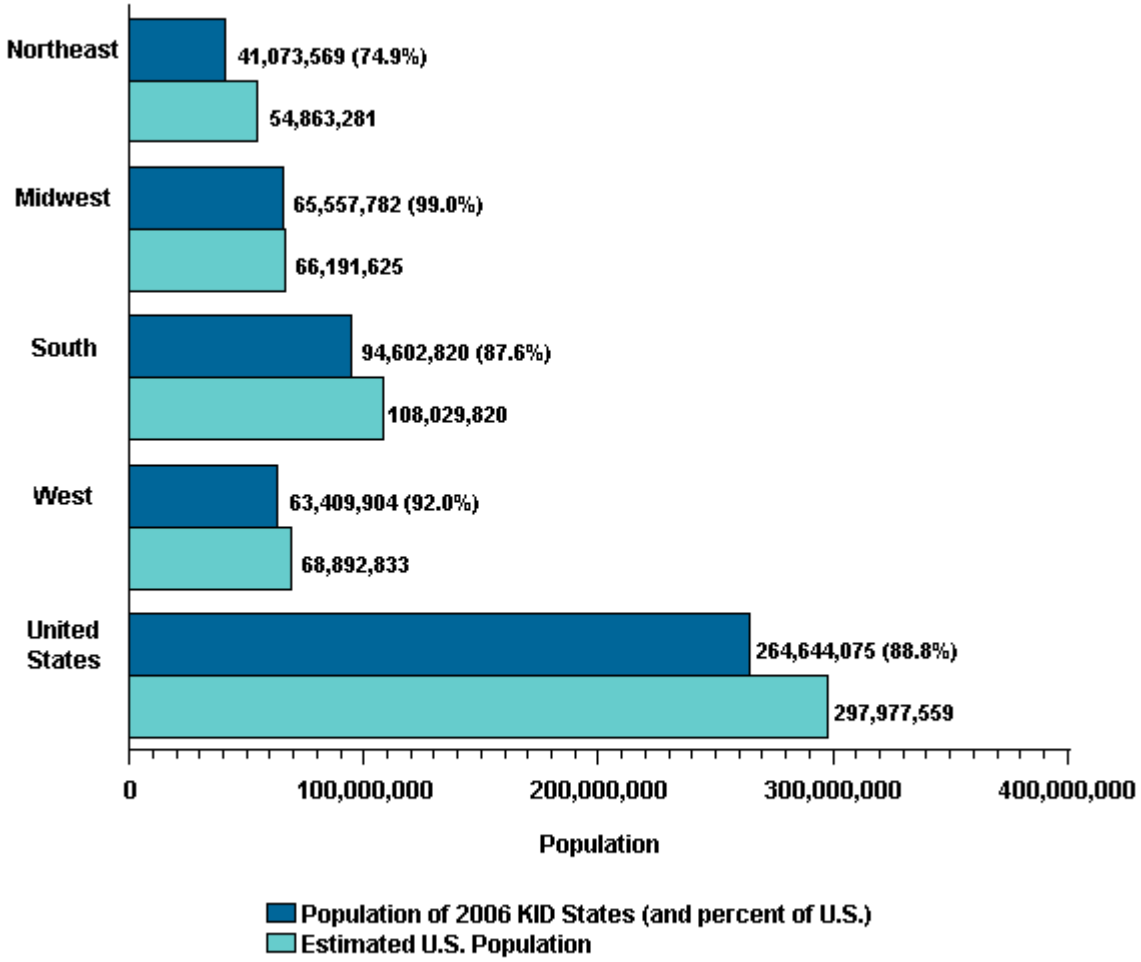
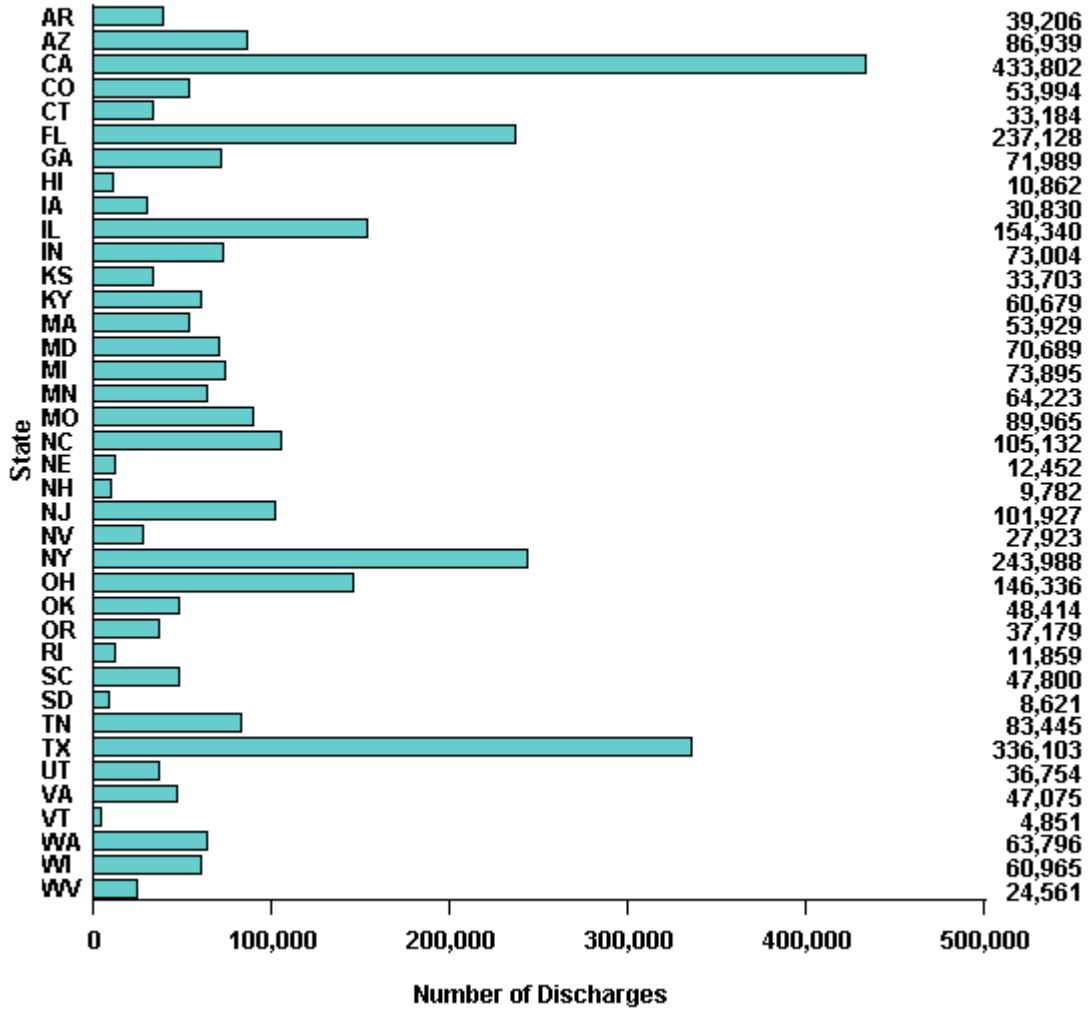


Figure 4. Number of Discharges in the 2006 KID, by State



Appendix II: State-Specific Restrictions

The table below enumerates the types of restrictions applied to the KIDS' Inpatient Database. Restrictions include the following types:

- Confidentiality of hospitals
    - Restricted identification of hospitals
    - Limitation on sampling
    - Restricted release of stratifiers
- Confidentiality of records
    - Restricted release of age in years, age in months, or age in days
    - Other restrictions
- Confidentiality of physicians
- Missing discharges.

For each restriction type the data sources are listed alphabetically by State. Only data sources that have restrictions are included. Data sources that do not have restrictions are not included.

Confidentiality of Hospitals - Restricted Identification of Hospitals

The following data sources required that hospitals not be identified in the KID:

- AR: Arkansas Department of Health & Human Services
- CT: Chime, Inc.
- GA: GHA: An Association of Hospitals & Health Systems
- HI: Hawaii Health Information Corporation
- IN: Indiana Hospital & Health Association
- KS: Kansas Hospital Association
- MI: Michigan Health & Hospital Association
- MO: Hospital Industry Data Institute
- NE: Nebraska Hospital Association
- OH: Ohio Hospital Association
- OK: Oklahoma State Department of Health
- SC: South Carolina State Budget & Control Board
- SD: South Dakota Association of Healthcare Organizations
- TN: Tennessee Hospital Association
- TX: Texas Department of State Health Services

In these States the following data elements are set to missing for all hospitals:

- DSHOSPID, data source hospital identifier
- HOSPSTCO, unmodified hospital State, county FIPS code
- HFIPSSTCO, modified hospital State, county FIPS code.
- IDNUMBER, AHA hospital identifier without leading 6
- AHAID, AHA hospital identifier with leading 6
- HOSPNAME, hospital name
- HOSPCITY, hospital city
- HOSPADDR, hospital address
- HOSPZIP, hospital ZIP Code

The following additional data elements are set to missing for all Georgia hospitals:

- PEDS_PCT, percent of hospital discharges, 20 years old or younger.
- PEDS_DISC, number of hospital discharges; 20 years or younger.
- TOTAL_DISC, total number of discharges.

Confidentiality of Hospitals - Limitation on Sampling

Limitations on sampling were needed for the following data sources:

- CT: Chime, Inc.
  - Chime requested that one hospital be excluded from the sampling frame.
- GA: GHA: An Association of Hospitals & Health Systems
  - GHA requested that no more than 60% of Georgia hospitals be included in the KID.
  - Ninety-six out of 161 Georgia hospitals (60%) were included in the 2006 KID.
- IL: Illinois Department of Public Health
  - Illinois Department of Public Health requested that no more than 40% of Illinois discharges appear in any discharge quarter of KID data.
  - 2006 KID – About 9% of the discharges in Illinois were sampled. No hospitals were dropped from the sampling frame.
- MI: Michigan Health & Hospital Association
  - Reporting of total charge is limited in the Michigan data. Thirty-three out of 134 hospitals were dropped from the sampling frame because they did not report total charges. These hospitals were fairly evenly distributed by hospital type. There were no sampling strata in the State containing only hospitals without total charges.
- NE: Nebraska Hospital Association
  - Nebraska Hospital Association requested that the two stand-alone children's hospitals be excluded from the sampling frame.
- OH: Ohio Hospital Association
  - Ohio masked the identities of three hospitals in the data provided to HCUP. These three hospitals were not included in the sampling frame because we were unable to match them to the AHA data.
- SC: South Carolina State Budget & Control Board
  - South Carolina requested that two hospitals be excluded from the sampling frame.
- VA: Virginia Health Information
  - The KID may not include more than 50% of the hospitals in Virginia.
  - Forty-six of 93 hospitals (49%) of the hospitals in Virginia were included in the 2006 KID.

Some States limit the hospitals that can be included in the KID. The following data sources requested that hospitals be dropped from the sampling frame whenever there were fewer than two hospitals in a sampling stratum. For more details about the number of hospitals included in the AHA Universe, Frame, and KID for each KID State, refer to Table 7 in Appendix I.

- GA: GHA: An Association of Hospitals & Health Systems
- HI: Hawaii Health Information Corporation
- IN: Indiana Hospital & Health Association
- MI: Michigan Health & Hospital Association
- NE: Nebraska Hospital Association
- OH: Ohio Hospital Association
- OK: Oklahoma State Department of Health
- SC: South Carolina State Budget & Control Board
- SD: South Dakota Association of Healthcare Organizations
- TN: Tennessee Hospital Association

Confidentiality of Hospitals - Restricted Release of Stratifiers

Stratifier data elements were restricted for the following data sources to further ensure hospital confidentiality in the KID:

- GA: GHA: An Association of Hospitals & Health Systems
- HI: Hawaii Health Information Corporation
- IN: Indiana Hospital & Health Association
- MI: Michigan Health & Hospital Association
- NE: Nebraska Hospital Association
- OK: Oklahoma State Department of Health
- OH: Ohio Hospital Association
- SC: South Carolina State Budget & Control Board
- SD: South Dakota Association of Healthcare Organizations
- TN: Tennessee Hospital Association

For the above States, stratifier data elements were set to missing if the cell, as defined by the data elements below, had fewer than two hospitals in the universe of the State's hospitals:

- HOSP_CONTROL, control/ownership of hospital
- HOSP_LOCATION, location (urban/rural) of hospital
- HOSP_TEACH, teaching status of hospital
- HOSP_BEDSIZE, bed size of hospital
- HOSP_LOCTEACH, location/teaching status of hospital
- NACHTYPE, National Association of Children's Hospitals and Related Institutions (NACHRI) hospital type

Confidentiality of Records - Restricted Release of Age in Years, Age in Months, or Age in Days

The following data sources restrict or limit the release of age:

- CT: Chime, Inc.
  - Age in days at admission (AGEDAY) is set to missing on all records.
  - Age in months at admission (AGEMONTH) is set to missing on all records.

- FL: Florida Agency for Health Care Administration
  - Age in days (AGEDAY) is set to missing on all records.
  - Age in months at admission (AGEMONTH) is set to missing on all records.

- MA: Division of Health Care Finance and Policy
  - Age in days (AGEDAY) is set to missing on all records.

- NH: New Hampshire Department of Health & Human Services
  - Age in days (AGEDAY) is set to missing on all records.

- SC: South Carolina State Budget & Control Board
  - Age in days (AGEDAY) is set to missing on all records.
  - Age in months at admission (AGEMONTH) is set to missing on all records.

- TX: Texas Department of State Health Services
  - Age in days (AGEDAY) is set to missing on all records.
  - Age in months at admission (AGEMONTH) is set to missing on all records.
  - Age in years (AGE) is set to the midpoints of age ranges defined by the data source. There were 6 age groups for the general patient population.

Texas Restriction on AGE for General Patient Population other than HIV or Drug/Alcohol Use Patients	
Age Range	New value of AGE
0	0
1-4	2
5-9	7
10-14	12
15-17	16
18-20	19

  - Texas also requested that age in years (AGE) be set missing for HIV or alcohol/drug use patients. The HIV or drug/alcohol use patients are identified by any principal or secondary diagnosis code on the record having the first four characters equal to one of the values in the following list: '2910', '2911', '2912', '2913', '2914', '2915', '2918', '2919', ‘2920', '2921', '2922', '2928', '2929', '3030', '3039', '3040', '3041', '3042', '3043', '3044', '3045', '3046', '3047', '3048','3049', '3050', '3052', '3053', '3054', '3055', '3056', '3057', '3058', '3059', ‘7903', 042, ‘V08''.

Confidentiality of Records – Other Restrictions

The following data sources restrict or limit the release of data elements for patient confidentiality:

- CA: Office of Statewide Health Planning & Development
  - Admission month (AMONTH), age in days, (AGEDAY), age in years (AGE), age in months (AGEMONTH), gender (FEMALE), and race (RACE), are suppressed for some records. In some cases, AGE is set to the midpoint of the age category.
- CT: Chime, Inc.
  - Admission month (AMONTH) is set to missing on all records.
- FL: Florida Agency for Health Care Administration
  - Admission month (AMONTH) is set to missing on all records
- GA: GHA: An Association of Hospitals & Health Systems
  - Patient race (RACE) is set to missing on all records
- NY: New York State Department of Health
  - Birth Weight (BWT) is set to missing on all records
- OK: Oklahoma State Department of Health
  - Days from admission to procedure (PRDAYn) is set to missing on all records.
  - Birth Weight (BWT) is set to missing on all records.

Confidentiality of Physicians

The following data sources restrict the release of physician identifiers:

- CA: Office of Statewide Health Planning & Development
- CT: Chime, Inc.
- GA: GHA: An Association of Hospitals & Health Systems
- IL: Illinois Department of Public Health
- IN: Indiana Hospital & Health Association
- MA: Division of Health Care Finance and Policy
- NC: North Carolina Department of Health and Human Services
- OH: Ohio Hospital Association
- OK: Oklahoma State Department of Health
- UT: Utah Department of Health
- VT: Vermont Association of Hospitals and Health Systems
- WV: West Virginia Health Care Authority

In these States the following data elements are set to missing for all records:

- MDNUM1_R
- MDNUM2_R

Missing Discharges

The following data sources may be missing discharge records for specific populations of patients:
 	
• IA: Iowa Hospital Association
  o The Iowa Hospital Association prohibits the release of two types of discharges: HIV Infections (defined by MDC of 25) and behavioral health including chemical dependency care or psychiatric care (defined by a service code of BHV). These discharges were not included in the source file provided to HCUP and are therefore not included in the KID.
 	
• NE: Nebraska Hospital Association
  o The Nebraska Hospital Association prohibits the release of discharge records for patients with HIV diagnoses. These discharges were not included in the source file provided to HCUP and are therefore not included in the KID.

Return to Introduction

Appendix III: Data Elements

Table 1. Data Elements in the KID Inpatient Core File

Note: Not all data elements in the KID are uniformly coded or available across all States. Each KID release differs in that some data elements were dropped, some were added, and the values of some data elements were changed.

Data elements that are *italicized* are not included in the 2006 KID, but are only available in previous years' files.

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
Admission day of week or weekend	AWEEKEND	2000, 2003, 2006	Admission on weekend: (0) admission on Monday-Friday, (1) admission on Saturday-Sunday	
	ADAYWK	*1997*	*Admission day of week: (1) Sunday, (2) Monday, (3) Tuesday, (4) Wednesday, etc.*	
Admission month	AMONTH	1997, 2000, 2003, 2006	Admission month coded from (1) January to (12) December	CT, FL
Admission source	ASOURCE	1997, 2000, 2003, 2006	Admission source, uniform coding: (1) ER, (2) another hospital, (3) another facility including long-term care, (4) court/law enforcement, (5) routine/birth/other	
	ASOURCE_X	2000, 2003, 2006	Admission source, as received from data source using State-specific coding	
	ASOURCEUB92	2003, 2006	Admission source (UB-92 standard coding). For newborn admissions (ATYPE = 4): (1) normal delivery, (2) premature delivery, (3) sick baby, (4) extramural birth; For non-newborn admissions (ATYPE NE 4): (1) physician referral, (2) clinic referral, (3) HMO referral, (4) transfer from a hospital, (5) transfer from a skilled nursing facility, (6) transfer from a another health care facility, (7) emergency room, (8) court/law enforcement, (A) transfer from a critical access hospital	CA, MD, RI
Admission type	ATYPE	1997, 2000, 2003, 2006	Admission type, uniform coding: (1) emergency, (2) urgent, (3) elective, (4) newborn, (5) trauma center beginning in 2003 data, (6) other	CA
	ELECTIVE	2003, 2006	Indicates elective admission: (1) elective, (0) non-elective admission	
Age at admission	AGE	1997, 2000, 2003, 2006	Age in years coded 0-124 years	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	AGEDAY	1997, 2000, 2003, 2006	Age in days coded 0-365 only when the age in years is less than 1	CT, FL, MA, NH, SC, TX
	AGEMONTH	1997, 2000, 2003, 2006	Age in months (when age < 11 years)	CT, FL, SC, TX, VA
Birth weight	BWT	2000, 2003, 2006	Birth weight in grams	CA, FL, IA, IL, KS, MI, MN, MO, NE, NH, NV, NY, OH, OK, OR, SC, SD, TN, TX, UT, WA, WI, WV
Clinical Classifications Software (CCS) category	DXCCS1 - DXCCS15	2000, 2003, 2006	CCS category for all diagnoses	
	DCCHPR1	*1997*	*CCS category for principal diagnosis in 1997. CCS was formerly called the Clinical Classifications for Health Policy Research (CCHPR)*	
	PRCCS1 - PRCCS15	2000, 2003, 2006	CCS category for all procedures	
	PCCHPR1	*1997*	*CCS category for principal procedure in 1997. CCS was formerly called the Clinical Classifications for Health Policy Research (CCHPR)*	
Diagnosis information	DX1 - DX15	1997, 2000, 2003, 2006	Diagnoses, principal and secondary (ICD-9-CM). Beginning in 2003, the diagnosis array does not include any of external cause of injury codes. These codes have been stored in a separate array ECODEn.	
	DXV1 - DXV15	*1997*	*Diagnosis validity flags*	
	HOSPBRTH	1997, 2000, 2003, 2006	Birth diagnosis, in this hospital	
	NDX	1997, 2000, 2003, 2006	Number of diagnoses coded on the original record	
	UNCBRTH	1997, 2000, 2003, 2006	Normal, uncomplicated birth in hospital	
Diagnosis Related Group (DRG)	DRG	1997, 2000, 2003, 2006	DRG in use on discharge date	
	DRGVER	2000, 2003, 2006	Grouper version in use on discharge date	
	DRG10	*1997*	*DRG Version 10 (effective October 1992 - September 1993)*	
	DRG18	*2000, 2003*	*DRG Version 18 (effective October 2000 - September 2001)*	
	DRG24	2006	DRG Version 24 (effective October 2006 - September 2007)	
Discharge quarter	DQTR	1997, 2000, 2003, 2006	Coded: (1) Jan - Mar, (2) Apr - Jun, (3) Jul - Sep, (4) Oct - Dec	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	DQTR_X	2006	Discharge quarter, as received from data source	
Discharge weights	DISCWT	2000, 2003, 2006	Weight to discharges in AHA universe for national estimates. In 2000, the discharge weight DISCWTcharge should be used for estimates of total charges.	
	DISCWT_U	*1997*	*Weight to discharges in AHA universe for national estimates.*	
	DISCWTcharge	*2000*	*Weight to discharges in AHA universe for total charge estimates.*	
Discharge year	YEAR	1997, 2000, 2003, 2006	Calendar year	
Disposition of patient (discharge status)	DIED	1997, 2000, 2003, 2006	Indicates in-hospital death: (0) did not die during hospitalization, (1) died during hospitalization	
	DISP	*1997*	*Disposition of patient, uniform coding in 1997: (1) routine, (2) short-term hospital, (3) skilled nursing facility, (4) intermediate care facility, (5) another type of facility, (6) home health care, (7) against medical advice, (20) died*	
	DISPUB92	2000, 2003, 2006	Disposition of patient (UB-92 standard coding)	CA, IN, MD
	DISPUNIFORM	2000, 2003, 2006	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	
External causes of injury and poisoning	ECODE1 – ECODE4	2003, 2006	External cause of injury and poisoning code, primary and secondary (ICD-9-CM). Beginning in 2003, external cause of injury codes are stored in a separate array ECODEn from the diagnosis codes in the array DXn. Prior to 2003, these codes are contained in the diagnosis array (DXn).	
	E_CCS1 - E_CCS4	2003, 2006	CCS category for the external cause of injury and poisoning codes	
	NECODE	2003, 2006	Number of external cause of injury codes on the original record.	
Gender of patient	FEMALE	2000, 2003, 2006	Indicates gender for KID beginning in 1998: (0) male, (1) female	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	SEX	*1997*	*Indicates gender in 1997 KID: (1) male, (2) female*	
Hospital information	DSHOSPID	2000, 2003, 2006	Hospital number as received from the data source	CT, GA, HI, IN, KS, MI, MO, NE, OH, OK, SC, SD, TN, TX
	HOSPID	2000, 2003, 2006	HCUP hospital number (links to Hospital file)	
	HOSPNUM	*1997*	*HCUP hospital number in 1997 (links to Hospital file)*	
	HOSPST	2000, 2003, 2006	State postal code for the hospital (e.g., AZ for Arizona)	
	HOSPSTCO	*2000*	*Modified Federal Information Processing Standards (FIPS) State/county code for the hospital links to Area Resource File (available from the Bureau of Health Professions, Health Resources and Services Administration). Beginning in 2003, this data element is available only on the hospital file.*	
	KID_STRATUM	2000, 2003, 2006	Hospital stratum used for weights.	
Length of Stay	LOS	1997, 2000, 2003, 2006	Length of stay, edited	
	LOS_X	1997, 2000, 2003, 2006	Length of stay, as received from data source	
Location of the patient	*PL_UR_CAT4*	*2003*	*Urban–rural designation for patient's county of residence: (1) large metropolitan, (2) small metropolitan, (3) micropolitan, (4) non-core*	
	PL_NCHS2006	2006	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	
Major Diagnosis Category (MDC)	MDC	1997, 2000, 2003, 2006	MDC in use on discharge date	
	MDC10	*1997*	*MDC Version 10 (effective October 1992 - September 1993)*	
	MDC18	*2000, 2003*	*MDC Version 18 (effective October 2000 - September 2001)*	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	MDC24	2006	MDC Version 24 (effective October 2006 - September 2007)	
Median household income for patient's ZIP Code	ZIPINC_QRTL	2003, 2006	Median household income quartiles for patient's ZIP Code. Because these estimates are updated annually, the value ranges for the ZIPINC_QRTL categories vary by year. Check the HCUP-US Website for details.	
	ZIPINC	*2000*	*Median household income category in files beginning in 1998: (1) $1-$24,999, (2) $25,000-$34,999, (3) $35,000-$44,999, (4) $45,000 and above*	
	ZIPINC4	*1997*	*Median household income category in 1997: (1) $1-$25,000, (2) $25,001-$30,000, (3) $30,001-$35,000, (4) $35,001 and above*	
Neonatal/ maternal flag	NEOMAT	1997, 2000, 2003, 2006	Assigned from diagnoses and procedure codes: (0) not maternal or neonatal, (1) maternal diagnosis or procedure, (2) neonatal diagnosis, (3) maternal and neonatal on same record	
Payer information Payer information (continued)	PAY1	1997, 2000, 2003, 2006	Expected primary payer, uniform: (1) Medicare, (2) Medicaid, (3) private including HMO, (4) self-pay, (5) no charge, (6) other	
	PAY1_N	*1997*	*Expected primary payer, nonuniform: (1) Medicare, (2) Medicaid, (3) Blue Cross, Blue Cross PPO, (4) commercial, PPO, (5) HMO, PHP, etc., (6) self-pay, (7) no charge, (8) Title V, (9) Worker's Compensation, (10) CHAMPUS, CHAMPVA, (11) other government, (12) other*	
	PAY1_X	2000, 2003, 2006	Expected primary payer, as received from the data source	
	PAY2	1997, 2000, 2003, 2006	Expected secondary payer, uniform: (1) Medicare, (2) Medicaid, (3) private including HMO, (4) self-pay, (5) no charge, (6) other	AZ, CA, CO, FL, HI, IA, NH, OH, OK, RI, SD, VA
	PAY2_N	*1997*	*Expected secondary payer, nonuniform: (1) Medicare, (2) Medicaid, (3) Blue Cross, Blue Cross PPO, (4) commercial, PPO, (5) HMO, PHP, etc., (6) self-pay, (7) no charge, (8) Title V, (9) Worker's Compensation, (10) CHAMPUS, CHAMPVA, (11) other government, (12) other*	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	PAY2_X	2000, 2003, 2006	Expected secondary payer, as received from the data source	AZ, CA, CO, FL, HI, IA, NH, OH, OK, RI, SD, VA
Physician identifiers, synthetic	MDNUM1_R	2003, 2006	Re-identified attending physician number in files starting in 2003	CA, CT, GA, HI, IL, IN, MA, NC, OH, OK, UT, VT, WI, WV
	MDID_S	*1997, 2000*	*Synthetic attending physician number in 1997 and 2000 KID*	
	MDNUM2_R	2003, 2006	Re-identified secondary physician number in files starting in 2003	CA, CT, GA, HI, IL, IN, MA, NC, OH, OK, UT, VT, WI, WV
	SURGID_S	*1997, 2000*	*Synthetic second physician number in 1997 and 2000 KID*	
Procedure information	PR1 - PR15	1997, 2000, 2003, 2006	Procedures, principal and secondary (ICD-9-CM)	
	PRV1 -PRV15	*1997*	*Procedure validity flag*	
	NPR	1997, 2000, 2003, 2006	Number of procedures coded on the original record	
	PRDAY1	1997, 2000, 2003, 2006	Number of days from admission to principal procedure.	IL, OH, OK, UT, WA, WV
	PRDAY2 - PRDAY15	2000, 2003, 2006	Number of days from admission to secondary procedures.	
Race of Patient	RACE	1997, 2000, 2003, 2006	Race, uniform coding: (1) white, (2) black, (3) Hispanic, (4) Asian or Pacific Islander, (5) Native American, (6) other	GA, IL, KY, MN, NV, OH, OR, WA, WV
Record identifier, synthetic	RECNUM	1997, 2003, 2006	HCUP unique record number	
	KEY	*2000*	*Unique record number for 2000 KID file*	
Total Charges	TOTCHG	1997, 2000, 2003, 2006	Total charges, edited	
	TOTCHG_X	1997, 2000, 2003, 2006	Total charges, as received from data source	

Return to Introduction

Table 2. Data Elements in the KID Hospital File

Note: Not all data elements in the KID are uniformly coded or available across all States. Each 2000 KID release differs in that some data elements were dropped, some were added, and the values of some data elements were changed.

Data elements that are *italicized* are not included in the 2006 KID, but are only available in previous years' files.

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
Universe Counts	N_DISC_U	1997, 2000, 2003, 2006	Number of universe discharges in the KID_STRATUM	
	N_BRTH_U	1997, 2000, 2003, 2006	Number of universe births in KID_STRATUM	
	N_HOSP_U	1997, 2000, 2003, 2006	Number of universe hospitals in KID_STRATUM	
Sample Counts	S_DISC_U	1997, 2000, 2003, 2006	Number of sampled discharges in the sampling stratum (KID_STRATUM or STRATUM)	
	S_BRTH_U	1997, 2000, 2003, 2006	Number of sample births in KID_STRATUM	
	S_CHLD_U	1997, 2000, 2003, 2006	Number of sample pediatric non-births in KID_STRATUM	
	S_CMPB_U	1997, 2000, 2003, 2006	Number of sample complicated births in KID_STRATUM	
	S_UNCB_U	1997, 2000, 2003, 2006	Number of sample uncomplicated births in KID_STRATUM	
	S_HOSP_U	1997, 2000, 2003, 2006	Number of sample hospitals in KID_STRATUM	
SID (Frame) Counts	PEDS_DISC	2000, 2003, 2006	Number of discharges, 20 years old or younger, from this hospital in the SID	GA
	PEDS_PCT	2000, 2003, 2006	Percentage of hospital discharges, 20 years old or younger, from this hospital in the SID	GA
	TOTAL_DISC	2000, 2003, 2006	Total number of discharges from this hospital in the SID	GA
	TOTDSCHG	*1997*	*Total number of discharges from this hospital in the SID*	
Hospital Identifiers	HOSPID	2000, 2003, 2006	HCUP hospital identification number (links to inpatient Core files)	
	HOSPNUM	*1997*	*HCUP hospital identification number (links to inpatient Core files)*	
	AHAID	2000, 2003, 2006	AHA hospital identifier that matches AHA Annual Survey Database	CT, GA, HI, IN, KS, MI, MO, NE, OH, OK, SC, SD, TN, TX
	IDNUMBER	2000, 2003, 2006	AHA hospital identifier without the leading 6	CT, GA, HI, IN, KS, MI, MO, NE, OH, OK, SC, SD, TN, TX

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	HOSPNAME	2000, 2003, 2006	Hospital name from AHA Annual Survey Database	AR, CT, GA, HI, IN, KS, MI, MO, NE, OH, OK, SC, SD, TN, TX
	NACHTYPE	1997, 2000, 2003, 2006	National Association of Children's Hospitals and Related Institutions (NACHRI) hospital type: (0) not identified as a children's hospital by NACHRI, (1) children's general hospital, (2) children's specialty hospital, (3) children's unit in a general hospital	GA, NE, OK
Hospital Location	HOSPADDR	2000, 2003, 2006	Hospital address from AHA Annual Survey Database	AR, CT, GA, HI, IN, KS, MI, MO, NE, OH, OK, SC, SD, TN, TX
	HOSPCITY	2000, 2003, 2006	Hospital city from AHA Annual Survey Database	AR, CT, GA, HI, IN, KS, MI, MO, NE, OH, OK, SC, SD, TN, TX
	HOSPST	2000, 2003, 2006	Hospital State postal code for hospital (e.g., AZ for Arizona)	
	HOSPSTCO	2003, 2006	Modified Federal Information Processing Standards (FIPS) State/county code for the hospital links to Area Resource File (available from the Bureau of Health Professions, Health Resources and Services Administration)	CT, GA, HI, IN, KS, MI, MO, NE, OH, OK, SC, SD, TN, TX
	HOSPZIP	2000, 2003, 2006	Hospital ZIP Code from AHA Annual Survey Database	AR, CT, GA, HI, IN, KS, MI, MO, NE, OH, OK, SC, SD, TN, TX
	HFIPSSTCO	2006	Unmodified Federal Information Processing Standards (FIPS) State/county code for the hospital. Links to the Area Resource File (available from the Bureau of Health Professions, Health Resources and Services Administration)	CT, GA, HI, IN, KS, MI, MO, NE, OH, OK, SC, SD, TN, TX
Hospital Characteristics	KID_STRATUM	2000, 2003, 2006	Hospital stratum used for weights	
	STRATUM	*1997*	*Hospital stratum used for weights in 1997*	
	HOSP_BEDSIZE	2000, 2003, 2006	Bed size of hospital: (1) small, (2) medium, (3) large	
	H_BEDSZ	*1997*	*Bed size of hospital: (1) small, (2) medium, (3) large*	
	HOSP_CONTROL	2000, 2003, 2006	Control/ownership of hospital: (0) government or private, collapsed category, (1) government, nonfederal, public, (2) private, non-profit, voluntary, (3) private, invest-own, (4) private, collapsed category	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	H_CONTRL	*1997*	*Control/ownership of hospital: (1) government, nonfederal (2) private, non-profit (3) private, invest-own*	
	HOSP_ LOCATION	2000, 2003, 2006	Location: (0) rural, (1) urban	
	H_LOC	*1997*	*Location: (0) rural, (1) urban*	
	HOSP_ LOCTEACH	2000, 2003, 2006	Location/teaching status of hospital: (1) rural, (2) urban non-teaching, (3) urban teaching	
	H_LOCTCH	*1997*	*Location/teaching status of hospital: (1) rural, (2) urban non-teaching, (3) urban teaching*	
	HOSP_REGION	2000, 2003, 2006	Region of hospital: (1) Northeast, (2) Midwest, (3) South, (4) West	
	H_REGION	*1997*	*Region of hospital: (1) Northeast, (2) Midwest, (3) South, (4) West*	
	HOSP_TEACH	2000, 2003, 2006	Teaching status of hospital: (0) non-teaching, (1) teaching	
	H_TCH	*1997*	*Teaching status of hospital: (0) non-teaching, (1) teaching*	
Discharge Year	YEAR	1997, 2000, 2003, 2006	Calendar year	
Note: Because the following variables are not needed for calculating national estimates, they are not included in the 2006 KID Hospital file.				
Discharge Weights	*CHLDWT*	*2000*	*Weight to pediatric non-births in universe for national estimates. In 2000, the discharge weight CHLDWTcharge should be used for estimates of total charges.*	
	CHLDWT_U	*1997*	*Weight to pediatric cases in universe for national estimates. In the 1997 data, one weight CHLDWT_U is used to create all estimates.*	
	CHLDWTCHARGE	*2000*	*Weight to pediatric non-births in universe for total charge estimates*	
	CMPBWT	*2000*	*Weight to complicated births in universe for national estimates. In 2000, the discharge weight CMPBWTcharge should be used for estimates of total charges.*	
	CMPBWTCHARGE	*2000*	*Weight to complicated births in universe for total charge estimates*	
	UNCBWT	*2000*	*Weight to uncomplicated births in universe for national estimates. In 2000, the discharge weight UNCBWTcharge should be used for estimates of total charges.*	
	UNCBWTCHARGE	*2000*	*Weight to uncomplicated births in universe for total charge estimates*	
Frame Counts	*H_BRTH_F*	*1997, 2000*	*Number of frame HCUP births in KID_STRATUM*	
	H_CHLD_F	*1997, 2000*	*Number of frame HCUP pediatric non-births in KID_STRATUM*	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	H_CMPB_F	*1997, 2000*	*Number of frame HCUP complicated births in KID_STRATUM*	
	H_UNCB_F	*1997, 2000*	*Number of frame HCUP uncomplicated births in KID_STRATUM*	
	H_DISC_F	*1997, 2000*	*Number of frame HCUP discharges in KID_STRATUM*	
	H_HOSP_F	*1997, 2000*	*Number of frame HCUP hospitals in KID_STRATUM*	
Sample Counts	*S_CHLD*	*1997, 2000*	*Pediatric non-births sampled*	
	S_CMPB	*1997, 2000*	*Complicated births sampled*	
	S_UNCB	*1997, 2000*	*Uncomplicated births sampled*	

Return to Introduction

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

All data elements listed below are available for all States in the 2006 KID Disease Severity Measures Files.

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes
AHRQ Comorbidity Software (AHRQ)	CM_AIDS	2003, 2006	AHRQ comorbidity measure: Acquired immune deficiency syndrome
	CM_ALCOHOL	2003, 2006	AHRQ comorbidity measure: Alcohol abuse
	CM_ANEMDEF	2003, 2006	AHRQ comorbidity measure: Deficiency anemias
	CM_ARTH	2003, 2006	AHRQ comorbidity measure: Rheumatoid arthritis/collagen vascular diseases
	CM_BLDLOSS	2003, 2006	AHRQ comorbidity measure: Chronic blood loss anemia
	CM_CHF	2003, 2006	AHRQ comorbidity measure: Congestive heart failure
	CM_CHRNLUNG	2003, 2006	AHRQ comorbidity measure: Chronic pulmonary disease
	CM_COAG	2003, 2006	AHRQ comorbidity measure: Coagulopathy
	CM_DEPRESS	2003, 2006	AHRQ comorbidity measure: Depression
	CM_DM	2003, 2006	AHRQ comorbidity measure: Diabetes, uncomplicated
	CM_DMCX	2003, 2006	AHRQ comorbidity measure: Diabetes with chronic complications
	CM_DRUG	2003, 2006	AHRQ comorbidity measure: Drug abuse
	CM_HTN_C	2003, 2006	AHRQ comorbidity measure: Hypertension, uncomplicated and complicated
	CM_HYPOTHY	2003, 2006	AHRQ comorbidity measure: Hypothyroidism
	CM_LIVER	2003, 2006	AHRQ comorbidity measure: Liver disease
	CM_LYMPH	2003, 2006	AHRQ comorbidity measure: Lymphoma
	CM_LYTES	2003, 2006	AHRQ comorbidity measure: Fluid and electrolyte disorders
	CM_METS	2003, 2006	AHRQ comorbidity measure: Metastatic cancer
	CM_NEURO	2003, 2006	AHRQ comorbidity measure: Other neurological disorders
	CM_OBESE	2003, 2006	AHRQ comorbidity measure: Obesity
	CM_PARA	2003, 2006	AHRQ comorbidity measure: Paralysis
	CM_PERIVASC	2003, 2006	AHRQ comorbidity measure: Peripheral vascular disorders
	CM_PSYCH	2003, 2006	AHRQ comorbidity measure: Psychoses
	CM_PULMCIRC	2003, 2006	AHRQ comorbidity measure: Pulmonary circulation disorders
	CM_RENLFAIL	2003, 2006	AHRQ comorbidity measure: Renal failure
	CM_TUMOR	2003, 2006	AHRQ comorbidity measure: Solid tumor without metastasis
	CM_ULCER	2003, 2006	AHRQ comorbidity measure: Peptic ulcer disease excluding bleeding

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes
	CM_VALVE	2003, 2006	AHRQ comorbidity measure: Valvular disease
	CM_WGHTLOSS	2003, 2006	AHRQ comorbidity measure: Weight loss
All Patient Refined DRG (3M)	APRDRG	2003, 2006	All Patient Refined DRG
	APRDRG_Risk_ Mortality	2003, 2006	All Patient Refined DRG: Risk of Mortality Subclass
	APRDRG_Severity	2003, 2006	All Patient Refined DRG: Severity of Illness Subclass
All-Payer Severity-adjusted DRG (HSS, Inc.)	APSDRG	2003, 2006	All-Payer Severity-adjusted DRG
	APSDRG_ Mortality_Weight	2003, 2006	All-Payer Severity-adjusted DRG: Mortality Weight
	APSDRG_LOS_ Weight	2003, 2006	All-Payer Severity-adjusted DRG: Length of Stay Weight
	APSDRG_Charge _Weight	2003, 2006	All-Payer Severity-adjusted DRG: Charge Weight
Disease Staging (Medstat)	DS_DX_ Category1	2003, 2006	Disease Staging: Principal Disease Category
	DS_Stage1	2003, 2006	Disease Staging: Stage of Principal Disease Category
	DS_LOS_Level	2003, 2006	Disease Staging: Length of Stay Level
	DS_LOS_Scale	2003, 2006	Disease Staging: Length of Stay Scale
	DS_Mrt_Level	2003, 2006	Disease Staging: Mortality Level
	DS_Mrt_Scale	2003, 2006	Disease Staging: Mortality Scale
	DS_RD_Level	2003, 2006	Disease Staging: Resource Demand Level
	DS_RD_Scale	2003, 2006	Disease Staging: Resource Demand Scale
Linkage Variables	HOSPID	2003, 2006	HCUP hospital identification number
	RECNUM	2003, 2006	HCUP record identifier

Return to Introduction

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

All data elements listed below are available for all States in the 2006 KID Diagnosis and Procedure Groups files.

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes
Clinical Classifications Software category for Mental Health and Substance Abuse (CCS-MHSA)	CCSMGN1 – CCSMGN15	2006	CCS-MHSA general category for all diagnoses
	CCSMSP1 – CCSMSP15	2006	CCS-MHSA specific category for all diagnoses
	ECCSMGN1 – ECCSMGN4	2006	CCS-MHSA general category for all external cause of injury codes
Chronic Condition Indicator	CHRON1 – CHRON15	2006	Chronic condition indicator for all diagnoses: (0) non-chronic condition, (1) chronic condition
	CHRONB1 – CHRONB15	2006	Chronic condition indicator body system for all diagnoses: (1) Infectious and parasitic disease, (2) Neoplasms, (3) Endocrine, nutritional, and metabolic diseases and immunity disorders, (4) Diseases of blood and blood-forming organs, (5) Mental disorders, (6) Diseases of the nervous system and sense organs, (7) Diseases of the circulatory system, (8) Diseases of the respiratory system, (9) Diseases of the digestive system, (10) Diseases of the genitourinary system, (11) Complications of pregnancy, childbirth, and the puerperium, (12) Diseases of the skin and subcutaneous tissue, (13) Diseases of the musculoskeletal system, (14) Congenital anomalies, (15) Certain conditions originating in the perinatal period, (16) Symptoms, signs, and ill-defined conditions, (17) Injury and poisoning, (18) Factors influencing health status and contact with health services
Procedure Class	PCLASS1 – PCLASS15	2006	Procedure Class for all procedures: (1) Minor Diagnostic, (2) Minor Therapeutic, (3) Major Diagnostic, (4) Major Therapeutic
Linkage Variables	HOSPID	2006	HCUP hospital identification number
	RECNUM	2006	HCUP record identifier

Return to Introduction

ENDNOTES

[1] Refer to Chapter 10 in Foreman, E.K., Survey Sampling Principles. New York: Dekker, 1991.

[2] 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.

[3] We used the following American Hospital Association Annual Survey Database (Health Forum, LLC © 2007) data elements 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].

Prior to the 1998 KID, we used the following SAS code to assign the KID teaching hospital status indicator, H_TCH:

```
/* FIRST ESTABLISH SHORT-TERM BEDS DEFINITION */
IF BDH NE . THEN BEDTEMP = BDH ;      /* SHORT TERM BEDS  */
ELSE IF BDH =. THEN BEDTEMP=BDTOT ;   /* TOTAL BEDS PROXY */

/*******************************************************/
/* NEXT ESTABLISH TEACHING STATUS BASED ON F-T & P-T   */
/* RESIDENT/INTERN STATUS FOR HOSPITALS.               */
/*******************************************************/
RESINT = (FTRES + .5*PTRES)/BEDTEMP ;
IF RESINT > 0 & (MAPP3=1 OR MAPP8=1) THEN H_TCH=1;/* 1=TEACHING */
ELSE H_TCH=0 ;                        /* 0=NONTEACHING */
```

Beginning with the 1998 KID, we used the following SAS code 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 VARIABLE */
/*******************************************************/
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 ;
```

[4] We performed this analysis during the development of the original 1997 KID.

[5] Most AHA surveys do not cover a January-to-December calendar year for every hospital. The numbers of hospitals for the KID are based on the AHA Annual Survey files.

[6] The columns in Table 7 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 before applying restrictions to the frame.
- "KID Sampling-Frame Hospitals" lists hospitals with pediatric discharges in the sampling frame after applying state-specific 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.

[7] Table 1: Annual Estimates of the Population for the United States, Regions, States, and Puerto Rico: April 1, 2000 to July 1, 2007 (NST-EST2007-01). Source: Population Division, U.S. Census Bureau. Release Date: December 27, 2007.