



**H·CUP**

HEALTHCARE COST AND UTILIZATION PROJECT

**DESIGN OF THE HCUP KIDS' INPATIENT  
DATABASE (KID), 2003**

**January 30, 2006**

**Healthcare Cost and Utilization Project  
Agency for Healthcare Research and Quality  
540 Gaither Road  
Rockville, MD 20850**

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>i</b>
Introduction .....	i
Sample Design.....	i
Design Considerations .....	i
Changes to Sampling and Weighting Strategy Beginning with the 2000 KID .....	ii
Sampling Frame .....	ii
Sampling Procedure.....	ii
Discharge Weights .....	ii
Weight Data Elements.....	iii
The 2003 KID Sample.....	iii
Data Analysis .....	iv
Missing Values .....	iv
Variance Calculations.....	iv
Studying Trends .....	iv
<b>INTRODUCTION .....</b>	<b>1</b>
<b>HOSPITAL UNIVERSE.....</b>	<b>2</b>
Hospital Merges, Splits, and Closures .....	2
Stratification Variables .....	2
<b>SAMPLING FRAME .....</b>	<b>7</b>
<b>SAMPLE DESIGN .....</b>	<b>11</b>
Design Considerations .....	11
Changes to Sampling and Weighting Strategy Beginning with the 2000 KID .....	12
Sampling Procedure.....	12
<b>SAMPLE WEIGHTS .....</b>	<b>12</b>
Discharge Weights .....	13
Weight Data Elements .....	14
<b>THE 2003 KID SAMPLE.....</b>	<b>15</b>
<b>DATA ANALYSIS.....</b>	<b>21</b>
Missing Values .....	21
Variance Calculations .....	21
Computer Software for Variance Calculations .....	22
Studying Trends .....	23
<b>CONCLUSION.....</b>	<b>23</b>
<b>ENDNOTES .....</b>	<b>24</b>

## INDEX OF TABLES

Table 1: Hospital Universe.....	2
Table 2: States, by Region.....	4
Table 3: Bed Size Categories, by Region .....	6
Table 4: Number of KID States, Hospitals, and Discharges, by Year .....	7
Table 5: Number of AHA, HCUP SID, and KID Hospitals, by State.....	10
Table 6: Number of Hospitals and Discharges in the AHA Universe, SID, and KID, by Hospital Type, 2003.....	15
Table 7: 2003 KID Discharges, by Hospital Type .....	20

## INDEX OF FIGURES

Figure 1: KID States, by Region .....	4
Figure 2: Number of Hospitals in the 2003 AHA Universe, SID, and KID, by Region.....	17
Figure 3: Percentage of U.S. Population in 2003 KID States, by Region .....	18
Figure 4: Number of Discharges in the 2003 KID, by State .....	19

## EXECUTIVE SUMMARY

### Introduction

The Kids' Inpatient Database (KID) is one of a family of databases and software tools developed as part of the Healthcare Cost and Utilization Project (HCUP), a Federal-State-Industry partnership sponsored by the Agency for Healthcare Research and Quality (AHRQ).

The KID is a unique and powerful nationwide database of hospital inpatient stays for children. It is a sample of pediatric discharges (age 20 or less at admission) from community, non-rehabilitation hospitals from states participating in HCUP. The HCUP KID team designed the database to permit researchers to study a broad range of conditions and procedures related to hospitalizations of children. Researchers and policy makers can use the KID to identify, track, and analyze national trends in hospital utilization, access, charges, quality, and outcomes for children.

This report describes the KID sample and weights, summarizes the contents of the 2003 KID, and discusses data analysis issues. Previous KID releases covered 1997 and 2000. This document highlights cumulative information for all previous years 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. The 2003 KID includes data from 36 states, nine more than the 2000 KID.

### Sample Design

#### Design Considerations

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

It should be possible, for example, to estimate Diagnosis Related Group (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. However, the 2003 KID includes data from only 36 states. Therefore, it is advisable to verify your estimates against other data sources, if available. 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. The most recent report is available on the KID Documentation CD-ROM and provides a comparison of the 1997 KID with other data sources. The updated report for the current KID is expected to be available on the HCUP User Support Website (<http://www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp>) in late January, 2006.

## Changes to Sampling and Weighting Strategy Beginning with the 2000 KID

We use the HCUP Nationwide Inpatient Sample (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 2003 KID Documentation CD-ROM and on the HCUP User Support Website at <http://www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp>.

### Sampling Frame

The KID sampling frame included all pediatric discharges from community, non-rehabilitation hospitals in the HCUP State Inpatient Databases (SID) that could be matched to the corresponding American Hospital Association (AHA) survey data (subject to state-specific restrictions). 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.

### Sampling Procedure

The KID includes a sample of pediatric discharges from all hospitals in the sampling frame. For the sample, 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 percent of uncomplicated in-hospital births and 80 percent of other pediatric cases from each frame hospital.

### Discharge 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 an additional stratum for freestanding children's hospitals. Hospital stratification variables were defined as follows:

1. Geographic Region – Northeast, Midwest, West, and South
2. Control – public, private not-for-profit, and proprietary
3. Location – urban or rural

4. Teaching Status – teaching or non-teaching
5. Bed Size – small, medium, and large
6. Hospital Type – children’s or other hospital.

We merged any stratum containing fewer than two frame hospitals, 30 uncomplicated births, 30 complicated births, and 30 non-birth pediatric discharges sampled with an "adjacent" stratum containing hospitals with similar characteristics. We created the discharge weights by hospital stratum in proportion to the number of AHA newborns for newborns and in proportion to the total number of non-newborn AHA discharges for non-newborns.

### 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. For 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 2003 KID Sample**

AHRQ obtained agreements from 36 HCUP State Partners to participate in the 2003 KID. More than 90% of the hospital universe is included in the sampling frame for 27 of these states. Texas, supplied data from only 73% of the state’s hospitals because some Texas hospitals, mostly small rural facilities, are exempt from statutory reporting requirements. Minnesota supplied data from only 88% of the state’s hospitals to HCUP because a few Minnesota hospitals do not participate in the project. There are no apparent significant differences between the characteristics of participating and non-participating Minnesota hospitals. We omitted from the sampling frame 32 Michigan hospitals that did not report total charges, leaving 68% of Michigan hospitals in the frame. Six State Partners – Connecticut, Georgia, Hawaii, South Carolina, South Dakota, and Virginia – imposed sampling restrictions that limited the percentage of state hospitals included in the frame to between 54 and 88 percent. (Restrictions from other states did not have an appreciable effect on the percentage of hospitals in the sampling frame.)

Although pediatric discharges from hospitals in each region are selected for the KID, the comprehensiveness of the sampling frame varies by region. The percentage of hospitals included in the sampling frame is highest in the Midwest (89.0%) and in the West (78.0%), while figures are lower for the Northeast (63.9%) and the South (56.9%).

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. The percentage of the regional population in KID states ranges from 99.0% in the Midwest to 74.9% in the Northeast. Overall, the states in the 2003 KID include an estimated 86.5% of the entire U.S population, up from 76.2% in the 2000 KID.

The final 2003 KID sample included 2,984,129 discharges of children from 3,438 hospitals, drawn from 36 frame states representing each region of the United States. The 2003 KID is larger than the 2000 KID across several dimensions:

- The number of states increased from 27 to 36.
- The number of hospitals increased from 2,784 to 3,438.
- The number of discharges increased from 2.5 million to 3.0 million.

## **Data Analysis**

### 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 will not accurately represent the discharge population. Also, when estimating totals for non-negative variables with missing values, sums would tend to be underestimated because the cases with missing values would be omitted from the calculations. Several techniques are available to help overcome this bias. One strategy is to impute acceptable values to replace missing values. Another strategy is to use sample weight adjustments to compensate for missing values. 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.

### 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 were 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 sample without replacement may be used to calculate statistics and their variances in most applications.

Examples of the use of SAS, SUDAAN, and STATA to calculate variances in the KID are presented in the special report: *Calculating Kids' Inpatient Database Variances*. This report is available on the KID Documentation CD-ROM and on the HCUP User Support Website at <http://www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp>.

### Studying Trends

When studying trends over time using the KID, be aware that the sampling frame for the KID changes over time. Because more states have been added, estimates from earlier years of the KID may be subject to more sampling bias than later years of the KID.

Short-term rehabilitation hospitals are included in the 1997 KID, but are excluded from the 2000 and 2003 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 three percent of community hospitals are short-term rehabilitation hospitals and not all state data sources included short term rehabilitation hospitals.

## INTRODUCTION

The Kids' Inpatient Database (KID) is one of a family of databases and software tools developed as part of the Healthcare Cost and Utilization Project (HCUP), a Federal-State-Industry partnership sponsored by the Agency for Healthcare Research and Quality (AHRQ).

The KID is a unique and powerful nationwide database of hospital inpatient stays for children. It is a sample of pediatric discharges (age 20 or less at admission) from community, non-rehabilitation hospitals from states participating in HCUP. The HCUP KID team developed the database to permit researchers to study a broad range of conditions and procedures related to hospitalizations of children. Researchers and policy makers can use the KID to identify, track, and analyze national trends in hospital utilization, access, charges, quality, and outcomes for children.

Potential research issues focus on both discharge- and hospital-level outcomes. Discharge outcomes of interest include:

- Frequency
- Costs
- Lengths of stay
- Effectiveness
- Quality of care
- Appropriateness
- Access to hospital care.

Hospital-level outcomes of interest include:

- Mortality rates
- Complication rates
- Patterns of care
- Diffusion of technology
- Trends toward specialization.

These and other outcomes are of interest for the nation as a whole and for policy-relevant inpatient subgroups defined by diagnoses and procedures, geographic region, patient demographics, hospital characteristics, and pay sources.

This report focuses on the KID sample and weights, summarizes the contents of the 2003 KID, and discusses data analysis issues. Previous KID releases covered 1997 and 2000. We have enhanced the nationwide representation of the sample by incorporating data from additional HCUP State Partners. The 2003 KID includes data from 36 states, nine more than the 2000 KID. This document highlights cumulative information for all previous years to provide a longitudinal view of the database.



## 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 American Hospital Association (AHA) Annual Survey. The AHA defines community hospitals as follows: "All nonfederal short-term general and other specialty hospitals, excluding hospital units of institutions." 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 1 displays the number of hospitals in the universe for each year, based on the corresponding AHA Annual Survey.

**Table 1: Hospital Universe<sup>1</sup>**

<b>Year</b>	<b>Number of Hospitals</b>
<b>1997</b>	5,113
<b>2000</b>	4,839
<b>2003</b>	4,836

### **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 for the 2003 KID. Therefore, when two or more hospitals merged to create a new hospital, the original hospitals and the newly-formed hospital were all considered separate entities in the universe during the year they merged. Similarly, if a hospital split, the original hospital and all newly-created hospitals were separate entities in the universe during the year they split. Finally, hospitals that closed during a year were included as long as they were in operation during some part of the calendar year.

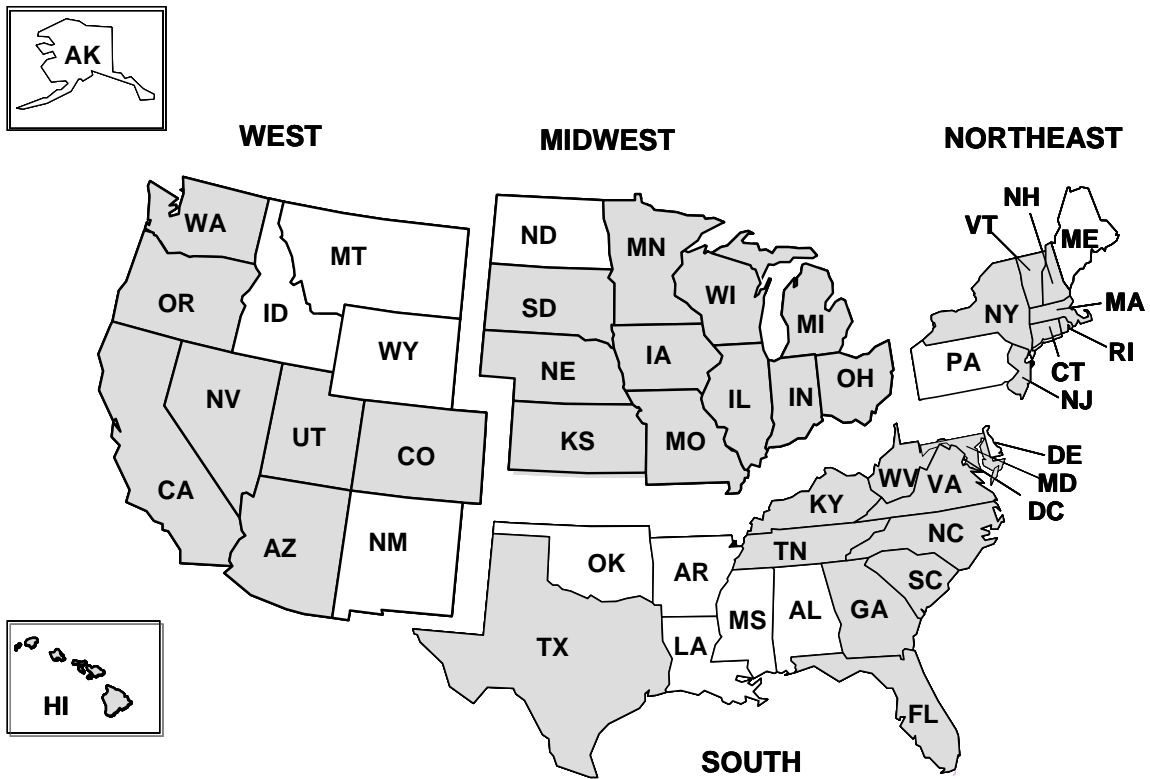
### **Stratification Variables**

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 1997 KID Documentation CD-ROM and on the HCUP 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 in gray, while Table 2 lists the states in each region.
2. *Control – government nonfederal (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.
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. The teaching status of hospitals identified as children’s hospitals by the National Association of Children’s Hospitals and Related Institutions (NACHRI) is based on an indicator provided by NACHRI. Other hospitals are considered to be teaching hospitals if they have an American Medical Association (AMA)-approved residency program, are members of the Council of Teaching Hospitals (COTH), or have a ratio of full-time equivalent interns and residents to beds of .25 or higher.<sup>2</sup>

**Figure 1: KID States, by Region**



**Table 2: States, by Region**

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

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 3. 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, in 2003, rural teaching hospitals comprised only 1% of the total hospital universe. The bed size categories were defined within location and teaching status because they would otherwise have been redundant. Rural hospitals tend to be small; urban non-teaching hospitals tend to be medium-sized; and urban teaching hospitals tend to be large. Yet it was important to recognize gradations of size within these types of hospitals. For example, in serving rural discharges, the role of "large" rural hospitals (particularly rural referral centers) often differs from the role of "small" rural hospitals.

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.

**Table 3: Bed Size Categories, by Region**

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

## 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. AHRQ obtained agreements with 36 HCUP State Partner organizations to include their data in the 2003 KID. The number of State Partners and hospitals contributing data to the KID has expanded over the years, as shown in Table 4.

**Table 4: Number of KID States, Hospitals, and Discharges, by Year**

Calendar Year	States in the Frame	Number of States	Sample Hospitals	Sample Discharges
1997	Arizona, California, Colorado, Connecticut, Florida, Georgia, Hawaii, Illinois, Iowa, Kansas, Maryland, Massachusetts, Missouri, New Jersey, New York, Oregon, Pennsylvania, South Carolina, Tennessee, Utah, Washington, and Wisconsin.	22	2,521	1,905,797
2000	Added Kentucky, Maine, North Carolina, Texas, Virginia, and West Virginia. Dropped Illinois.	27	2,784	2,516,833
2003	Added Illinois, Indiana, Michigan, Minnesota, Nebraska, New Hampshire, Nevada, Ohio, Rhode Island, South Dakota, and Vermont. Dropped Maine and Pennsylvania.	36	3,438	2,984,129

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

Table 5 shows the number of AHA, HCUP SID, and KID hospitals by state. The columns in Table 5 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. Seven hospitals in the frame were not included in the KID because they had so few pediatric discharges that none were randomly sampled.

In most cases, the difference between the universe and the frame represents the difference between the number of community, non-rehabilitation hospitals in the 2003 AHA Annual Survey of Hospitals 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 5, only 303 out of 414 Texas community, non-rehabilitation hospitals supplied data to HCUP for 2003. 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 91.8% of inpatient discharges from Texas universe hospitals.

The Minnesota frame contains 15 fewer hospitals than the state universe because a few of the state’s hospitals do not participate in HCUP. There are no apparent significant differences between the characteristics of participating and non-participating Minnesota hospitals.

The Ohio frame contains 11 fewer hospitals than the state universe, including three hospitals that could not be matched to the AHA data because the Partner masked their identities in the data.

For Connecticut, Georgia, Hawaii, Indiana, Michigan, Nebraska, South Carolina, South Dakota, and Virginia, we had to exclude several HCUP hospitals from the frame, as described below.

To help ensure the confidentiality of hospitals in Connecticut, the Partner requested that we exclude the only stand-alone community children's hospital in the state from the database.

Georgia, Hawaii, Indiana, Michigan, Nebraska, South Carolina, and South Dakota stipulated that only hospitals appearing in sampling strata with two or more hospitals from the state were to be included in the KID. Because of this restriction, two Georgia hospitals (including the only freestanding community, non-rehabilitation children's hospital in the state), four Hawaii hospitals, one Indiana hospital, one Michigan hospital, three Nebraska hospitals, five South Carolina hospitals, and two South Dakota hospitals were excluded from the KID sampling frame. An additional 43 Georgia hospitals were randomly omitted from the KID sampling frame because Georgia allowed no more than 60% of the state's hospitals to be included in the KID. The two stand-alone children's hospitals in Nebraska were also excluded at the Partner's request. Two additional South Carolina hospitals were removed from the sampling frame because they have unique characteristics that might make them identifiable.

Because charges represent a critical outcome variable in the KID, we decided to omit 32 additional Michigan hospitals from the frame that did not provide total charges. By excluding these hospitals, we avoid having to adjust the weights or create another weighting variable specifically for total charges. These hospitals are fairly evenly distributed by hospital type. There are no sampling strata in the state containing only hospitals without charges. The total charge data reported for Michigan is similar to total charge data reported by other Midwestern states. Thus, there does not seem to be an obvious bias in the type of cases for which charges are reported. The stratification and weighting scheme should adjust for the hospitals that are being excluded.

Because Virginia allowed only 50% or fewer of the state's hospitals to be included in the KID, 35 hospitals from the state were randomly omitted from the KID sampling frame.

A total of 133 hospitals were restricted from the KID sampling frame, leaving 3,445 hospitals with pediatric discharges in the frame. Seven of these hospitals were not represented in the KID because they had so few pediatric discharges that none were sampled.

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.



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

<b>State</b>	<b>AHA Universe Hospitals</b>	<b>SID Community, Non- Rehabilitation Hospitals</b>	<b>SID Community, Non-Rehabilitation Hospitals with Pediatric Discharges</b>	<b>KID Sampling- Frame Hospitals</b>	<b>KID Sample Hospitals</b>
Non-Frame	987	215	215	0	0
Arizona	59	58	58	58	58
California	367	361	359	359	358
Colorado	66	65	64	64	64
Connecticut	34	31	31	30	30
Florida	194	191	188	188	188
Georgia	143	142	141	96	96
Hawaii	23	20	20	16	16
Iowa	116	116	113	113	113
Illinois	189	188	187	187	187
Indiana	110	107	106	105	105
Kansas	136	128	128	128	127
Kentucky	98	96	96	96	96
Massachusetts	73	67	66	66	66
Maryland	48	46	46	46	46
Michigan	142	132	129	96	95
Minnesota	131	116	115	115	114
Missouri	118	115	114	114	114
North Carolina	113	109	109	109	109
Nebraska	84	83	82	77	77
New Jersey	26	26	26	26	26
New Jersey	73	73	73	73	73
Nevada	26	25	25	25	25
New York	204	203	200	200	200
Ohio	161	150	150	150	150
Oregon	58	56	56	56	56
Rhode Island	11	11	11	11	11
South Carolina	59	57	57	50	50
South Dakota	50	47	45	43	43
Tennessee	121	116	114	114	114
Texas	414	303	275	275	273
Utah	41	40	40	40	40
Virginia	81	79	79	44	44
Vermont	14	14	14	14	14
Washington	85	84	84	84	83
Wisconsin	127	127	125	125	125
West Virginia	54	54	52	52	52
<b>Total</b>	<b>4,836</b>	<b>3,851</b>	<b>3,793</b>	<b>3,445</b>	<b>3,438</b>

## 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. However, the 2003 KID includes data from only 36 states. Therefore, it is advisable to verify your estimates against other data sources, if available. 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. The most recent report is available on the KID Documentation CD-ROM and provides a comparison of the 1997 KID with other data sources. The updated report for the current KID is expected to be available on the HCUP User Support Website (<http://www.hcup-us.ahrq.gov/db/nation/kid/kidrelatedreports.jsp>) in late January, 2006.

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 $\geq$ 2), neonatal Major Diagnostic Category (MDC 15), or admission type of birth (ATYPE=4).<sup>3</sup> 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 2003 KID, approximately 1.2% 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. Less than 0.2% of cases in DRG 391 have an admission type of newborn (ATYPE = 4) but do not meet the in-hospital birth screen.

Using the above in-hospital birth screen, we identified 3,249,450 in-hospital births in community, non-rehabilitation hospitals in the 2003 KID, as compared with 3,238,145 births reported by the AHA in these hospitals. There were 11,305 fewer births reported by the AHA, a difference of less than .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 2003 KID Documentation CD-ROM and on the HCUP User Support 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 percent of uncomplicated in-hospital births and 80 percent 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 1997 KID Documentation CD-ROM and on the HCUP 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  $4 \div Q$ , 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 percent and sampled other pediatric discharges (complicated newborns and other pediatric cases) at the nominal rate of 80 percent from the discharges available in the (restricted) frame. To avoid rounding errors in the weights calculation, the actual sampling rate for a discharge type (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. For 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 2003 KID SAMPLE

There were 3,445 hospitals in the 2003 sampling frame, a 23.6% increase from the 2000 KID. The final 2003 KID sample included 2,984,129 discharges of children from 3,438 hospitals drawn from 36 frame states representing all regions of the United States. The 2003 KID is larger than the 2000 KID across several dimensions:

- The number of states increased from 27 to 36.
- The number of hospitals increased from 2,784 to 3,438.
- The number of discharges increased from 2.5 million to 3.0 million.

Table 6 summarizes the numbers 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
- KID nationwide hospitals and discharges.

**Table 6: Number of Hospitals and Discharges in the AHA Universe, SID, and KID, by Hospital Type, 2003**

Hospital Type	AHA Universe		SID		KID	
	Hospitals	Discharges (Including Births)	Hospitals with Pediatric Discharges	Pediatric Discharges	Hospitals	Pediatric Discharges
<b>Not a Children's Hospital</b>	4,757	37,689,130	3,742	6,034,209	3,396	2,714,359
<b>Children's Hospital</b>	79	531,529	51	422,072	42	269,770
<b>Total</b>	4,836	38,220,659	3,793	6,456,281	3,438	2,984,129

Figure 2 summarizes the 2003 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. Figure 2 reveals that the percentage of hospitals included in the sampling frame is highest in the Midwest (89.0%) and in the West (78.0%), while figures are lower for the Northeast (63.9%) and the South (56.9%).

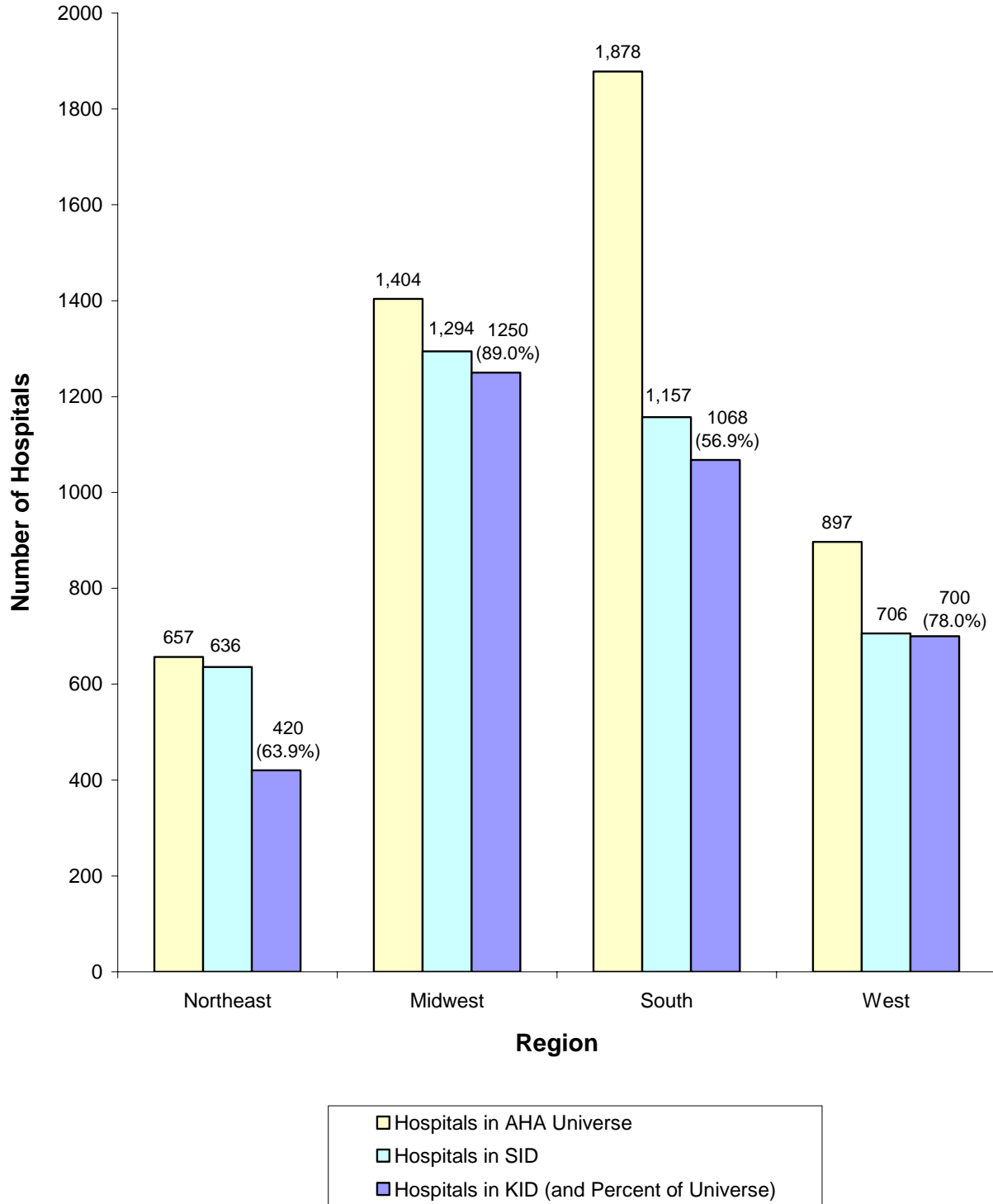
Because the KID sampling frame has a disproportionate representation of the more populous states and contains hospitals with more annual discharges, its comprehensiveness in terms of discharges is higher. Figure 3 summarizes the estimated U.S. population by geographic region on July 1, 2003. For each region, the figure reveals:

- The estimated U.S. population<sup>4</sup>
- The estimated population of states in the 2003 KID
- The percentage of estimated U.S. population included in KID states.

For example, the estimated population of the Northeast region on July 1, 2003 was 54,426,252. On that same date, the estimated population of states in the Northeast region that were included in the 2003 KID was 40,746,286. This represents 74.9% of the total Northeast region's population. The percentage of the Northeast population represented declined from 94.6% in the 2000 KID because the addition of New Hampshire, Rhode Island, and Vermont was offset by the loss of Maine and Pennsylvania. However, the seven Midwestern states added for 2003 have substantially increased the percentage of the Midwestern population represented, from 25.7% in the 2000 KID to 99.0% in the 2003 KID. The percentage of estimated U.S. population included in the West (92.0%) and in the South (81.3%) was slightly higher than for 2000. Overall, the states in the 2003 KID include an estimated 86.5% of the entire U.S population, up from 76.2% in the 2000 KID.

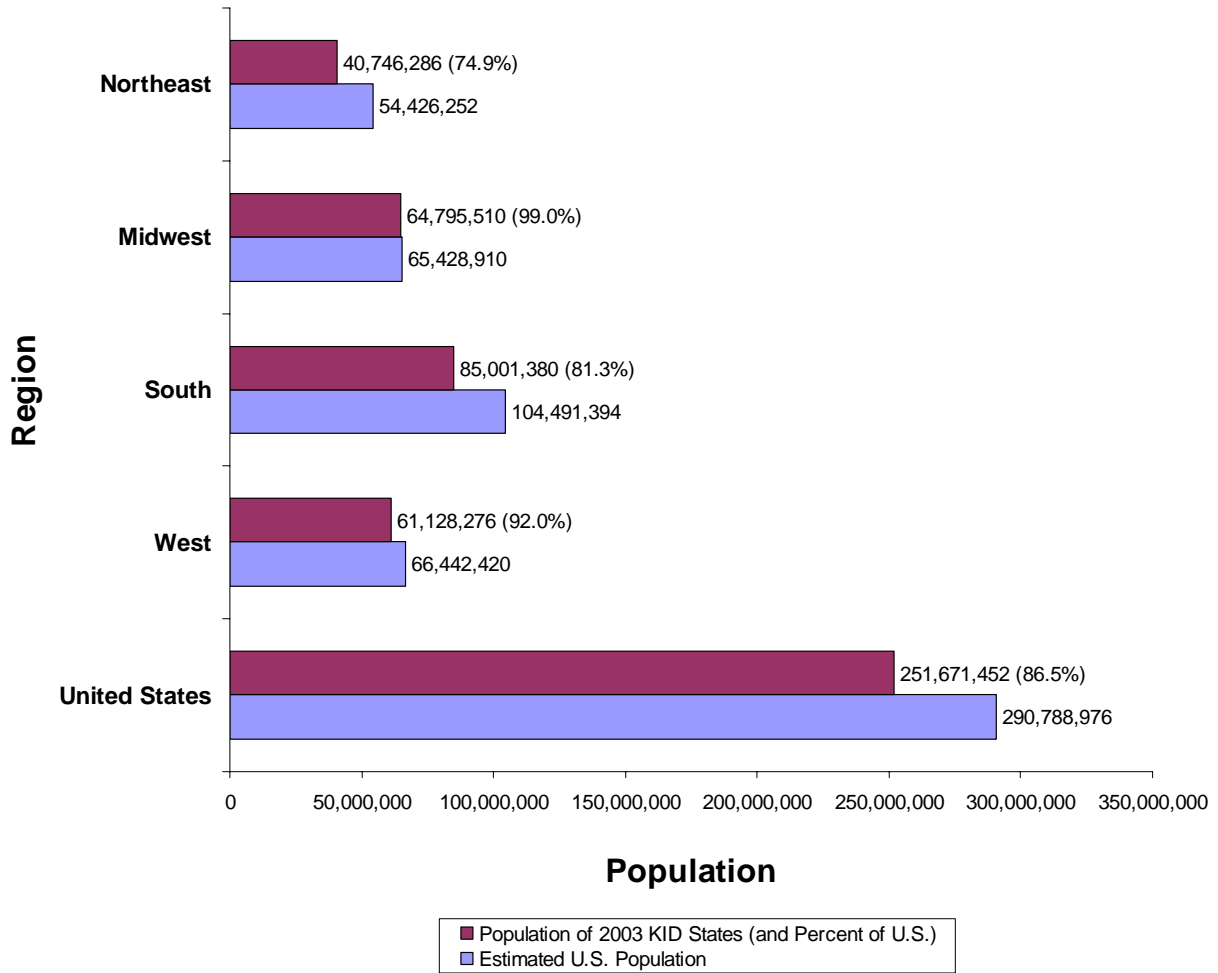
Figure 4 presents the number of discharges in the KID for each state in the sampling frame for 2003. The number of discharges ranges from 4,724 in Vermont to 428,650 in California.

**Figure 2: Number of Hospitals in the 2003 AHA Universe, SID, and KID, by Region**





**Figure 3: Percentage of U.S. Population in 2003 KID States, by Region**



**Figure 4: Number of Discharges in the 2003 KID, by State**

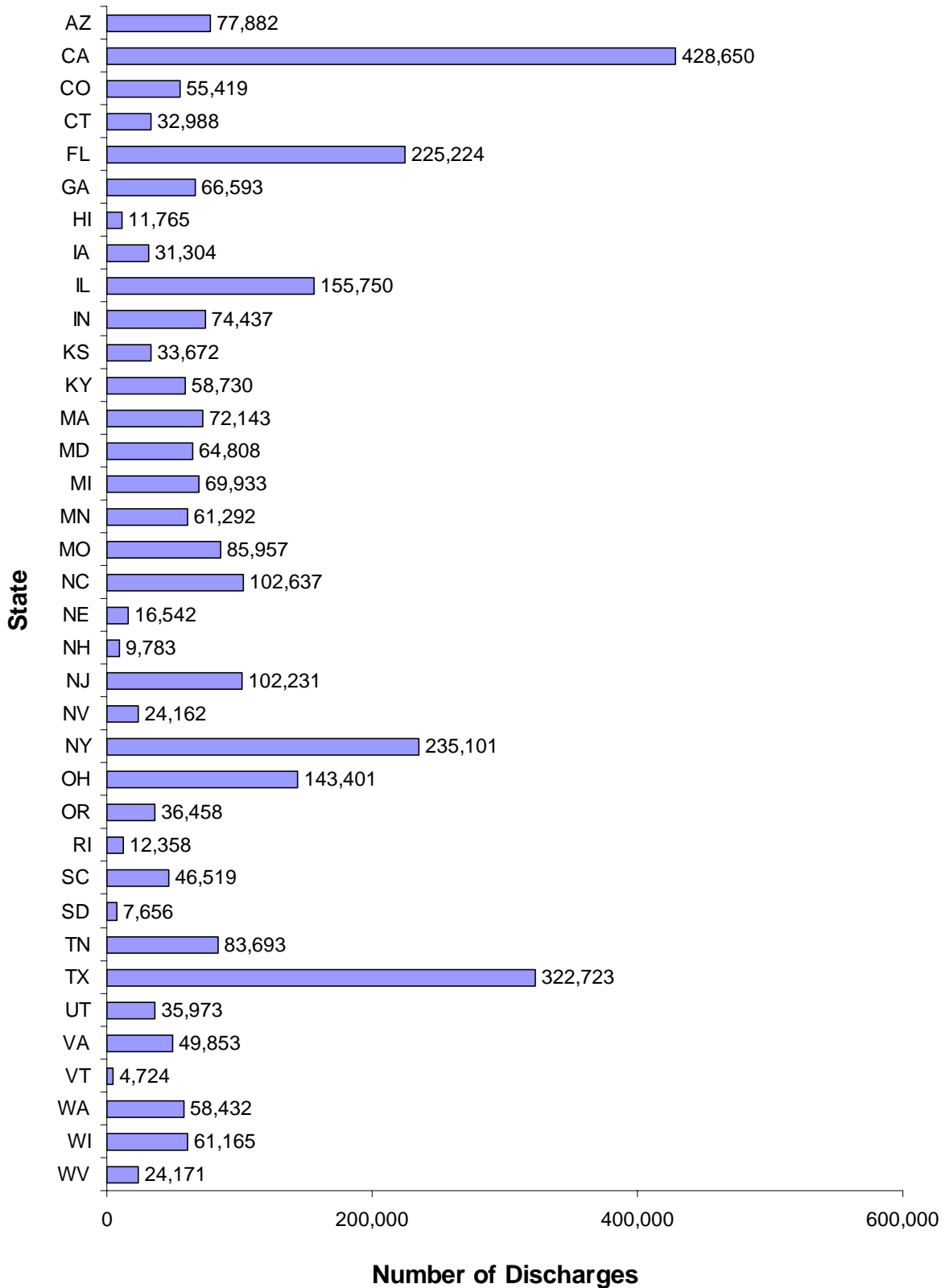


Table 7 displays the unweighted and weighted number of uncomplicated births, complicated births, and pediatric non-births by hospital type in the 2003 KID.

**Table 7: 2003 KID Discharges, by Hospital Type**

<b>Hospital Type</b>	<b>Uncomplicated Births</b>	<b>Complicated Births</b>	<b>Pediatric Non-Births</b>	<b>Total Pediatric Discharges</b>
<b>Unweighted:</b>				
<b>Not a Children's Hospital</b>	243,494	645,064	1,825,801	2,714,359
<b>Children's Hospital</b>	344	1,680	267,746	269,770
<b>Total</b>	243,838	646,744	2,093,547	2,984,129
<b>Weighted:</b>				
<b>Not a Children's Hospital</b>	2,937,048	979,523	2,976,017	6,892,588
<b>Children's Hospital</b>	2,616	1,600	512,358	516,574
<b>Total</b>	2,939,665	981,122	3,488,375	7,409,162

## DATA ANALYSIS

### 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.<sup>5</sup> 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.

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.<sup>6</sup> For example, nearly all Statistical Analysis System (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 Variances*. This report is available on the 2003 KID Documentation CD-ROM and on the HCUP User Support 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.

### **Studying Trends**

When studying trends over time using the KID, be aware that the sampling frame for the KID changes over time. Because more states have been added, estimates from earlier years of the KID may be subject to more sampling bias than later years of the KID.

Short-term rehabilitation hospitals are included in the 1997 KID, but are excluded from the 2000 and 2003 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 three percent of community hospitals are short-term rehabilitation hospitals and not all state data sources included short term rehabilitation hospitals.

### **CONCLUSION**

In this report, we have described the development and use of the KID sample and weights and summarized the contents of the 2003 KID. We have included cumulative information for all previous years to provide a longitudinal view of the database. Once again, the nationwide representation of the database has been enhanced by incorporating data from additional HCUP State Partners, a total of 36 participants for the year 2003. We have highlighted important considerations for data analysis and have provided references to detailed reports on this subject.

## ENDNOTES

- 1 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.
- 2 We used the NACHRI Teaching Hospital Indicator, if provided. Otherwise, we used the following AHA Annual Survey 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 = AMA approved residency program indicator [A102].

For the 1997 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 (MAPP3 = . AND MAPP8 = .) THEN DO ;
  IF RESINT > .10 THEN ST_TEACH = 1 ;
  ELSE ST_TEACH = 0 ;
END ;
IF (MAPP3=1 OR MAPP8=1) THEN ST_TEACH=1 ; /* 1=TEACHING */
ELSE ST_TEACH=0 ; /* 0=NONTEACHING */
/*****/
/* CREATE TEACHING CATEGORY VARIABLES TO FURTHER */
/* REFINE TEACHING STATUS DEFINITION. */
/*****/
IF ST_TEACH = 1 THEN DO ;
  IF 0 < RESINT < .15 THEN TEACHCAT=0 ; /* MINOR CATEGORY */
  ELSE IF RESINT GE .15 THEN TEACHCAT=1 ; /* MAJOR CATEGORY */
  ELSE ST_TEACH = 0 ; /* NONTEACH STATUS*/
END ;
```

Beginning with the 2000 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 ;

```

- 3 We performed this analysis during the development of the original 1997 KID.
- 4 U.S. Census Bureau, Population Division. "Table NST-EST2004-01 – Annual Estimates of the Population for the United States and States, and for Puerto Rico: April 1, 2000 to July 1, 2004." Internet Release Date: December 22, 2004.
- 5 Refer to Chapter 10 in Foreman, E.K., Survey Sampling Principles. New York: Dekker, 1991.
- 6 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.