HCUP KIDS' INPATIENT DATABASE

DESIGN REPORT, 1997

TABLE OF CONTENTS

INTRODUCTION

The Kids' Inpatient Database (KID) of the Healthcare Cost and Utilization Project (HCUP) was developed to enable analyses of hospital utilization by children across the United States. The target universe includes all pediatric discharges from all community hospitals in the United States. The KID is a nationwide sample of pediatric discharges from the HCUP State Inpatient Database (SID) community hospitals weighted to all pediatric discharges in the target universe.

Kids' Inpatient Database (KID)

Calendar Year	States	Number of Hospitals	Number of Discharge Records (unweighted)	Number of Discharge Records (weighted)
1997	22	2,521	1,905,797	6,657,326

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, and
- access to hospital care.

Hospital outcomes of interest include:

- mortality rates,
- complication rates,
- patterns of care,
- diffusion of technology, and
- trends toward specialization.

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

This report provides a detailed description of the KID sample design, as well as a summary of the resultant sample. Sample weights were developed to obtain national estimates of inpatient parameters. These weights are described in detail.

THE KID HOSPITAL UNIVERSE

The hospital universe is defined by all hospitals that were open during any part of the calendar year and were designated as community hospitals in the American Hospital Association (AHA) Annual Survey of Hospitals. Community hospitals, as defined by the AHA, include "all nonfederal, short-term, general and other specialty hospitals, excluding hospital units of institutions." Included among community hospitals are academic medical centers and specialty hospitals such as obstetrics-gynecology, ear-nose-throat, short-term rehabilitation, orthopedic, and pediatric hospitals. Excluded are federal hospitals (Veterans Administration, Department of Defense, and Indian Health Service hospitals), long term hospitals, psychiatric hospitals, alcohol/chemical dependency treatment facilities and hospitals units within institutions such as prisons. There were 5,113 universe hospitals based on the 1997 AHA Annual Survey.

Hospital Merges, Splits, and Closures

All hospital entities that were designated community hospitals in the AHA hospital file were included in the hospital universe. Therefore, if 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 for the year of the merge. Likewise, if a community hospital split, the original hospital and all newly created community hospitals were separate entities in the universe for the year of the split. Finally, community 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 purposes of calculating discharge weights, we post-stratified hospitals on six characteristics contained in the AHA hospital files. The stratification variables were as follows:

1) *Geographic Region – Northeast, Midwest, West, and South*. This is an important stratifier 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.

2) *Control – government nonfederal, private not-for-profit, and private investor-owned.* These types of hospitals tend to have different missions and different responses to government regulations and policies.

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 nonteaching*. The missions of teaching hospitals differ from nonteaching hospitals. In addition, financial considerations differ between these two hospital groups. A hospital is considered to be a teaching hospital if it has an AMA-approved residency program or is a member of the Council of Teaching Hospitals (COTH).

5) *Bedsize – small, medium, and large.* Bedsize categories are based on hospital beds, and are specific to the hospital's location and teaching status, as shown in Table 1.

Table 1. Bedsize Categories

Location and Teaching Status	Hospital Bedsize		
	Small	**Medium**	**Large**
Rural	1-49	50-99	100+
Urban, non-teaching	1-99	100-199	200+
Urban, teaching	1-299	300-499	500+

Rural hospitals were not split according to teaching status, because rural teaching hospitals were rare. For example, in 1997 there were only 12 rural teaching hospitals in the HCUP SID states. The bedsize categories were defined within location and teaching status because they would otherwise have been redundant. Rural hospitals tend to be small; urban nonteaching 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. The cut-off points for the bedsize categories are consistent with those used in *Hospital Statistics,* published annually by the AHA.

6) Hospital Type – 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 the National Association of Children's Hospitals and Related Institutions (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.

SAMPLING FRAME

The target universe is all pediatric discharges from community hospitals in the 1997 American Hospital Association (AHA) survey data. The *universe* of hospitals was established as all community hospitals located in the U.S. However, it was not feasible to obtain and process all-payer discharge data from a random sample of the entire universe of hospitals for at least two reasons. First, all-payer discharge data were not available from all hospitals for research purposes. Second, based on the experience of prior hospital discharge data collections, it would have been too costly to obtain data from individual hospitals, and it would have been too burdensome to process each hospital's unique data structure.

Therefore, the KID *sampling frame* was constructed from the subset of universe hospitals that released their discharge data to HCUP for research use. Two sources for all-payer discharge data were state agencies and private data organizations, primarily state hospital associations. At the time when the sample was drawn, the Agency for Healthcare Research and Quality (AHRQ) had agreements with 22 data sources that maintain statewide, all-payer discharge data files to include their data in the HCUP databases as shown in Table 2.

Table 2. States in the Frame for KID

Year	States in the Frame
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

The list of the entire frame of hospitals was composed of all AHA community hospitals in each of the frame states *that could be matched to the discharge data provided to HCUP*. If an AHA community 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). As described below, further restrictions were put on the sampling frames for Connecticut, Georgia, Hawaii, Illinois, South Carolina, Missouri, and Tennessee, for the nationwide sample hospitals.

The data source for Illinois data stipulated that no more than 40 percent of the discharges provided by Illinois could be included in a nationwide sample. However, the total number of pediatric discharges in Illinois represents only about 20 percent of all discharges. So, no changes were made to the Illinois sampling frame due to this restriction.

The data sources for Georgia, Hawaii, South Carolina and Tennessee data stipulated that only hospitals that appear in sampling strata with two or more hospitals were to be included in a nationwide sample. Due to this restriction, two Georgia hospitals, six Hawaii hospitals, six South Carolina hospitals and six Tennessee hospitals were excluded from the 1997 nationwide sampling frame, leaving 157 Georgia community hospitals, 11 Hawaii community hospitals, 54 South Carolina community hospitals and 92 Tennessee community hospitals in the 1997 sampling frame.

The data source for Missouri data stipulated that only hospitals that had signed agreements for public release should be included in a nationwide sample. For 1997, thirty-five Missouri hospitals signed releases for confidential HCUP use only. These hospitals were excluded from the nationwide sampling frame, leaving 75 hospitals in the 1997 sampling frame.

Table 3 shows the number of AHA, HCUP SID and KID hospitals by state. A total of 55 community hospitals were restricted from the KID nationwide sampling frame, leaving a total of 2,526 hospitals with pediatric discharges in the KID nationwide sampling frame. Of the 2,526 hospitals with pediatric discharges in the sampling frame, five had so few pediatric discharges that none were randomly selected for the nationwide sample, leaving 2,521 hospitals in the KID nationwide sample.

Table 3. Number of AHA, HCUP SID and KID Hospitals, by State

State	AHA Community Hospitals	All Hospitals in the SID	Community Hospitals in the SID	Community Hospitals in the SID with Sampling Restrictions Applied	Community Hospitals in the SID with any Pediatric Discharges	Community Hospitals in the SID with Pediatric Discharges after Sampling Restrictions Applied	KID Hospitals
Non-frame States	2,402	0	0	0	0	0	0
Arizona	64	69	62	62	62	62	62
California	415	543	411	411	407	407	407
Colorado	67	70	66	66	65	65	65
Connecticut	34	32	32	32	32	32	32
Florida	210	240	198	198	197	197	196
Georgia	159	188	159	157	158	156	156
Hawaii	20	22	17	11	17	11	11
Iowa	115	117	115	115	114	114	114
Illinois	203	217	202	202	201	201	201
Kansas	131	122	120	120	119	119	118
Massachusetts	84	75	73	73	72	72	72
Maryland	51	52	51	51	51	51	51
Missouri	125	115	110	75	110	75	75
New Jersey	85	79	78	78	78	78	78
New York	225	230	222	222	222	222	222
Oregon	61	64	59	59	59	59	59
Pennsylvania	217	239	211	211	210	210	209
South Carolina	65	61	60	54	60	54	53
Tennessee	126	98	98	92	98	92	92
Utah	41	51	40	40	40	40	40
Washington	89	93	88	88	86	86	85
Wisconsin	124	143	124	124	123	123	123
Total	5,113	2,920	2,596	2,541	2,581	2,526	2,521

Column definitions are as follows:

- "Community Hospitals in the SID" lists the sampling frame before applying any sampling restrictions and before selecting hospitals with pediatric discharges.
- "Community Hospitals in the SID with Sampling Restrictions Applied " lists the sampling frame with sampling restrictions applied, but before selecting hospitals with pediatric discharges.
- "Community Hospitals in the SID with Pediatric Discharges" lists the sampling frame for community hospitals with pediatric discharges before applying any sampling restrictions. These are the hospitals in the pediatric extract.
- "Community Hospitals in the SID with Pediatric Discharges with Sampling Restrictions Applied" lists the sampling frame for community hospitals with pediatric discharges after applying sampling restrictions.
- "KID Nationwide Sample Hospitals" lists the sampled hospitals in the KID from the sampling frame of community hospitals with pediatric discharges after the sampling restrictions were applied. Five hospitals were not selected for the KID because they had so few pediatric discharges that none were randomly selected.

Design Requirements

For the KID nationwide sample, we sampled 10 percent of uncomplicated in-hospital births, and 80 percent of other pediatric discharges from all hospitals contained in the restricted frame. The overall objective was to select a sample of pediatric discharges "generalizable" to the target universe, which includes pediatric discharges outside the frame (zero probability of selection). Moreover, this sample was to be geographically dispersed, yet drawn from the subset of states with inpatient discharge data that agreed to provide such data to the project.

It should be possible, for example, to estimate DRG-specific average lengths of stay over all U.S. hospitals using weighted average lengths of stay, based on averages or regression estimates from the KID. Ideally, relationships among outcomes and their correlates estimated from the KID should generally hold across all U.S. hospitals. However, since only 22 states contributed data to this release, some estimates may differ from estimates from comparative data sources. When possible, estimates based on the KID should be checked against national benchmarks, such as data from the National Hospital Discharge Survey to determine the appropriateness of the KID for specific analyses (see the Special Report: *HCUP KIDS' Inpatient Database Comparative Analysis, 1997*).

1997 KID Sampling Procedure

The nationwide sampling frame includes all pediatric discharges in the 22-state 1997 HCUP State Inpatient Databases (SID) from community hospitals matched to the 1997 AHA survey data (subject to state-specific restrictions discussed above). Pediatric discharges were defined to include all discharges that had an age at admission of 18 years or less. Discharges with missing, invalid or inconsistent ages were excluded.

Unlike the Nationwide Inpatient Sample (NIS), we did not execute a two-stage sampling procedure for the KID. Instead the KID includes a sample of pediatric discharges from all hospitals in the (restricted) sampling frame. Pediatric discharges are stratified by uncomplicated in-hospital birth, complicated in-hospital birth, and pediatric non-birth. The stratum-specific sampling rates are constant across all hospitals in the sampling frame. We sampled 10 percent of uncomplicated in-hospital births and 80 percent of other pediatric cases from each frame hospital. If we had fewer than: two frame hospitals, 30 uncomplicated births, 30 complicated births, or 30 non-birth pediatric discharges sampled in a stratum, then that stratum was merged with an "adjacent" stratum containing hospitals with similar characteristics for the purpose of calculating discharge weights.

For use in sampling and weighting births to the AHA, which reports in-hospital births, we wanted to identify all in-hospital births in the KID data. We also wanted to further separate the in-hospital births into uncomplicated "normal" births and complicated births. Uncomplicated births have little variation in their outcomes. Consequently, they could be sampled at a low rate.

To identify 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 MDC (MDC 15) or admission type of birth (ATYPE=4). Based on reviews of the cross-tabulations, the MDC 15 DRG definitions, and ICD-9-CM birth codes, the following screen was selected for births: an in-hospital birth diagnosis code (any DX code in the range V3000 - V3901 with a fourth digit of zero and a fifth digit of zero or one), 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 since 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. Since the HCUP data are already edited for neonatal diagnoses inconsistent with age, we did not include any age criteria in the in-hospital birth screen.

"Normal" uncomplicated in-hospital births are identified as cases that meet the above screen and are in DRG 391, "Normal Newborn." Around 0.6% 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 V3nnn code. Others, however, may be readmissions or may have been born before admission to the hospital, and did not receive a V3nnn code. Less than 0.1% 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 2,256,161 in-hospital births in community hospitals in the HCUP SID data compared to 2,312,557 births reported by the AHA in community hospitals in the HCUP SID states. There were 56,396 more births reported by the AHA, a difference of less than 2.5%.

The state-imposed restrictions on sampling did not affect the discharge sampling rates. All frame hospitals with pediatric discharges were included in the sample.

It should be observed that, for the NIS, states wanted to make it difficult or impossible to identify individual hospitals in part because the NIS included 100 percent of the discharges from hospitals in the NIS sample. Consequently, outcomes could have been estimated without sampling error for individual hospitals that could be identified in the sample. However, the KID includes fewer than 100 percent of the pediatric discharges for each hospital in the database. Therefore, researchers will not be able to calculate hospital-specific outcomes with certainty.

A systematic random sample was drawn from each stratum, after sorting discharges by hospital within each state, then by DRG within each hospital, and then by a random number within each DRG. These sorts ensured that the sample case-mix is representative of each hospital’s pediatric discharges.

FINAL KID SAMPLE

The actual numbers of hospitals and discharges in the KID are shown in table 4.

Table 4. Number of Hospitals and Discharges in 1997 KID

	Nationwide Sample	
Hospital Type	**Number of Hospitals**	**Number of Discharges**
Children's Hospital	26	169,008
Not a children's Hospital	2,495	1,736,789
Total	2,521	1,905,797

A more detailed breakdown of the 1997 KID hospital sample by geographic region is shown in Table 5. For each geographic region, Table 5 shows the number of:

- AHA universe hospitals and total discharges including births, and
- KID nationwide sample hospitals and discharges.

Table 5. Number of Hospitals and Discharges in AHA Universe and KID by Region, 1997

	AHA Universe		KID		
Region	**Hospitals**	**Total Discharges**	**Hospitals**	**% of AHA Hospitals**	**Pediatric Discharges**
Northeast	737	7,424,738	613	83.2	561,446
Midwest	1,453	8,332,143	631	43.4	309,365
South	1,968	13,098,721	548	27.8	421,174
West	955	6,552,605	729	76.3	613,812
Total	5,113	35,408,207	2,521	49.3	1,905,797

For example, in 1997 the Northeast region contained 737 hospitals in the AHA universe. It also contained 614 hospitals in the pediatric extract, of which 613 hospitals were in the KID nationwide sample.

Table 6 shows the number of hospitals and discharges in the AHA universe, in HCUP hospitals with any pediatric discharges (the "pediatric extract"), and in the KID for each state in the nationwide sampling frame for 1997. Some states have fewer hospitals in the nationwide sample than in the pediatric extract for one or both of the following reasons: 1) five hospitals have so few pediatric discharges that none were selected for the nationwide sample; and 2) as previously described, Georgia, Hawaii, Illinois, South Carolina, Tennessee and Missouri restricted which hospitals could be included in the nationwide sample.

- The number of Georgia hospitals in the KID nationwide sample is two less than in the Georgia pediatric extract. Two hospitals were excluded because of the sampling restrictions stipulated by Georgia.

- The number of Hawaii hospitals in the KID nationwide sample is six fewer than in the Hawaii pediatric extract. Six hospitals were excluded because of sampling restrictions stipulated by Hawaii.

- The number of South Carolina hospitals in the KID nationwide sample is seven fewer than in the South Carolina pediatric extract. Six hospitals were excluded because of sampling restrictions stipulated by South Carolina, and one hospital had so few pediatric discharges that none were selected for the nationwide sample.

- The number of Tennessee hospitals in the KID nationwide sample is six fewer than in the Tennessee pediatric extract. Six hospitals were excluded because of sampling restrictions stipulated by Tennessee.

- The number of Missouri hospitals in the KID nationwide sample is 35 fewer than the Missouri universe. Thirty-five hospitals were excluded because they signed releases restricted to confidential use only.

Table 6. Number of Hospitals and Discharges in AHA Universe, the Pediatric Extract, and KID, by State, 1997

	AHA Universe		Pediatric Extract		KID	
State	Hospitals	Total Discharges	Hospitals	Pediatric Discharges	Hospitals	Pediatric Discharges
Arizona	64	549,425	62	115,372	62	55,149
California	415	3,630,246	407	809,769	407	390,875
Colorado	67	420,792	65	88,382	65	44,822
Connecticut	34	381,404	32	66,641	32	31,032
Florida	210	2,093,989	197	322,300	196	174,103
Georgia	159	972,795	158	186,767	156	98,048
Hawaii	20	115,158	17	22,571	11	10,201
Iowa	115	403,312	114	66,379	114	34,536
Illinois	203	1,622,223	201	296,024	201	152,176
Kansas	131	333,668	119	54,486	118	26,625
Massachusetts	84	844,920	72	129,801	72	64,945
Maryland	51	633,450	51	104,593	51	61,267
Missouri	125	805,507	110	127,430	75	39,951
New Jersey	85	1,199,691	78	186,778	78	95,157
New York	225	2,628,619	222	437,371	222	229,360
Oregon	61	352,191	59	66,814	59	29,520
Pennsylvania	217	1,898,842	210	259,586	209	140,952
South Carolina	65	489,886	60	90,021	53	39,402
Tennessee	126	842,510	98	121,825	92	48,354
Utah	41	228,498	40	63,886	40	29,569
Washington	89	549,617	86	112,790	85	53,676
Wisconsin	124	617,300	123	111,683	123	56,077
Non-HCUP States	2,402	13,794,164	0	0	0	0
Total	5,113	35,408,207	2,581	3,841,269	2,521	1,905,797

Table 7 shows the non-weighted and weighted number of uncomplicated births, complicated births and pediatric non-births by hospital type in the 1997 KID nationwide sample.

Table 7. Kid Discharges

Hospital Type	Uncomplicated Births	Complicated Births	Pediatric Non-births	Total Pediatric Discharges
Non-Weighted:				
Not a Children's Hospital	154,881	512,447	1,069,461	1,736,789
Children's Hospital	210	1,418	167,380	169,008
Total	155,091	513,865	1,236,841	1,905,797
Weighted:				
Not a Children's Hospital	2,621,701	1,120,599	2,479,115	6,221,415
Children's Hospital	2,628	2,217	431,066	435,911
Total	2,624,329	1,122,816	2,910,181	6,657,326

SAMPLING WEIGHTS

Although the sampling design was simple and straightforward, it is necessary to incorporate sample weights to obtain national estimates. Therefore, sample weights were developed to weight the KID nationwide sample discharges to the AHA universe.

Weighting Options

Using the HCUP SID data from all 22 states, we summarized counts of pediatric discharges for each AHA hospital identifier and correlated those counts with AHA hospital characteristics. For example, we found that for the hospitals in the HCUP SID identified by the AHA as children's hospitals (service code = 50), the total number of discharges recorded in the AHA survey data is a good estimate of the total number of pediatric discharges observed in the HCUP SID for those hospitals. Intuitively, this makes sense because children's hospitals primarily serve only pediatric patients.

We considered the following three weighting options for the KID:

1. Weights in proportion to the total number of AHA discharges, with post-stratification on the standard NIS hospital stratification variables.
2. Weights 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, with post-stratification on the standard NIS hospital stratification variables.
3. Weights 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, with hospital type added to the standard post-stratification.

We selected Option 3, separate weights for newborns with hospital type added to the stratification. For this option, in addition to the standard NIS stratification variables, we considered the AHA hospital type in developing pediatric discharge weights.

Figure 1 contains a plot of the (logarithm of) discharge counts reported in the 1997 AHA survey versus the (logarithm of) discharge counts calculated from the HCUP SID data. Clearly, the AHA discharge count appears to be a reliable estimate of the HCUP SID discharge count. Therefore, the AHA count (variable B005) is also likely to be a good estimate of total discharges for the universe. These discharge counts include all discharges, not just pediatric discharges. However, weights for non-newborn pediatrics implicitly assume that, in the aggregate, the proportion of non-newborn pediatrics across the HCUP SID hospitals is the same as the proportion of non-newborn pediatrics in the universe of AHA hospitals within each stratum.

Figure 2 contains a plot of the (logarithm of) birth count from the AHA survey versus the (logarithm of) birth count for each hospital in the HCUP SID data. Most hospitals are clustered around the 45 degree diagonal, indicating good agreement between the AHA and the HCUP SID. However, 25 hospitals have HCUP SID births but zero AHA birth counts. Further, 365 hospitals have zero HCUP SID births but nonzero AHA births. Consultation with other sources revealed that some of these hospitals contract with other facilities for deliveries. In any case, we assume that the AHA count of births is sufficiently accurate in the aggregate (within each stratum) for these weight calculations.

Figure 1 - AHA vs. SID Total Hospital Discharges (Log Scale)

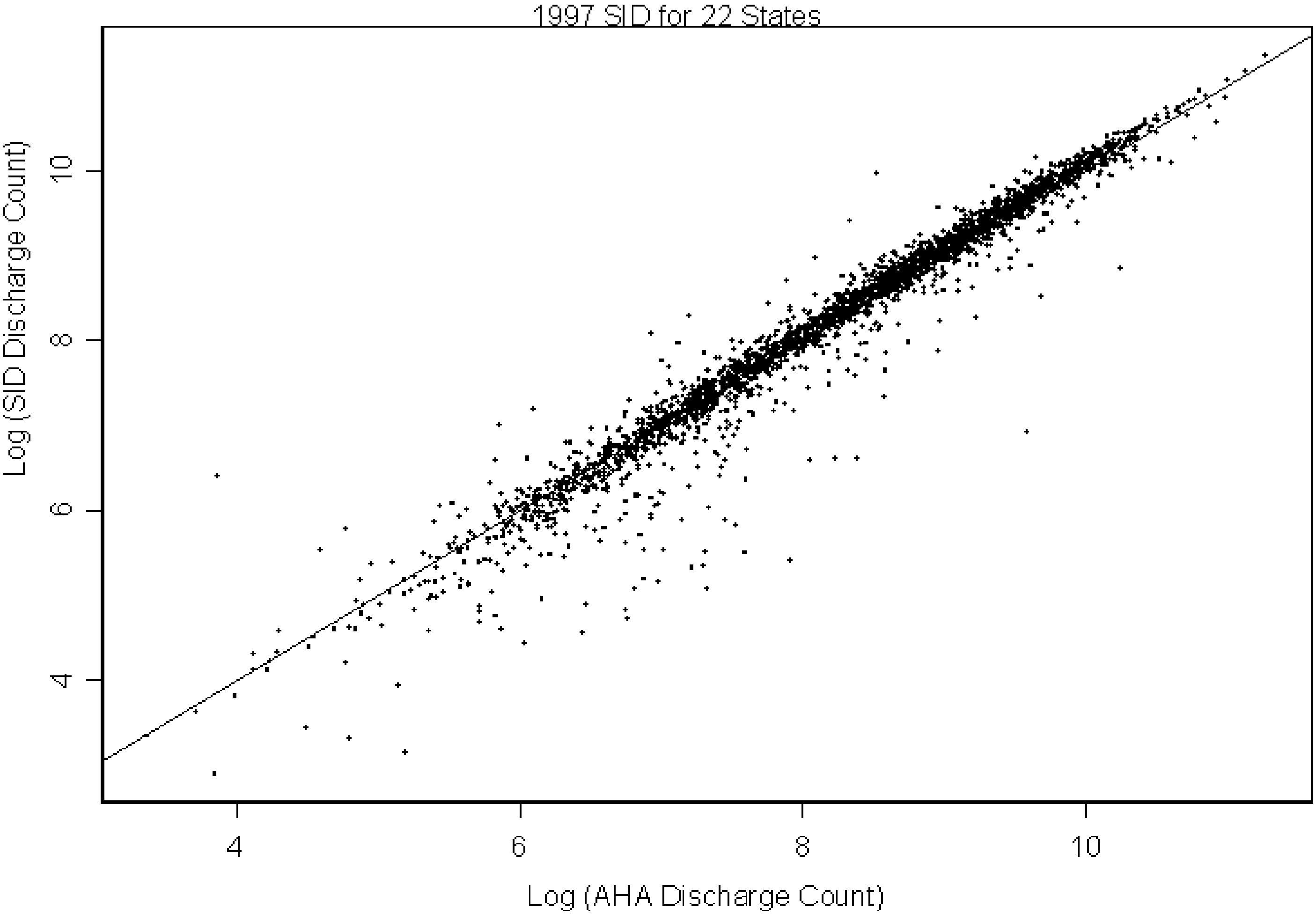


Figure 2 - AHA Birth Count vs. SID Birth Count (Log Scale)

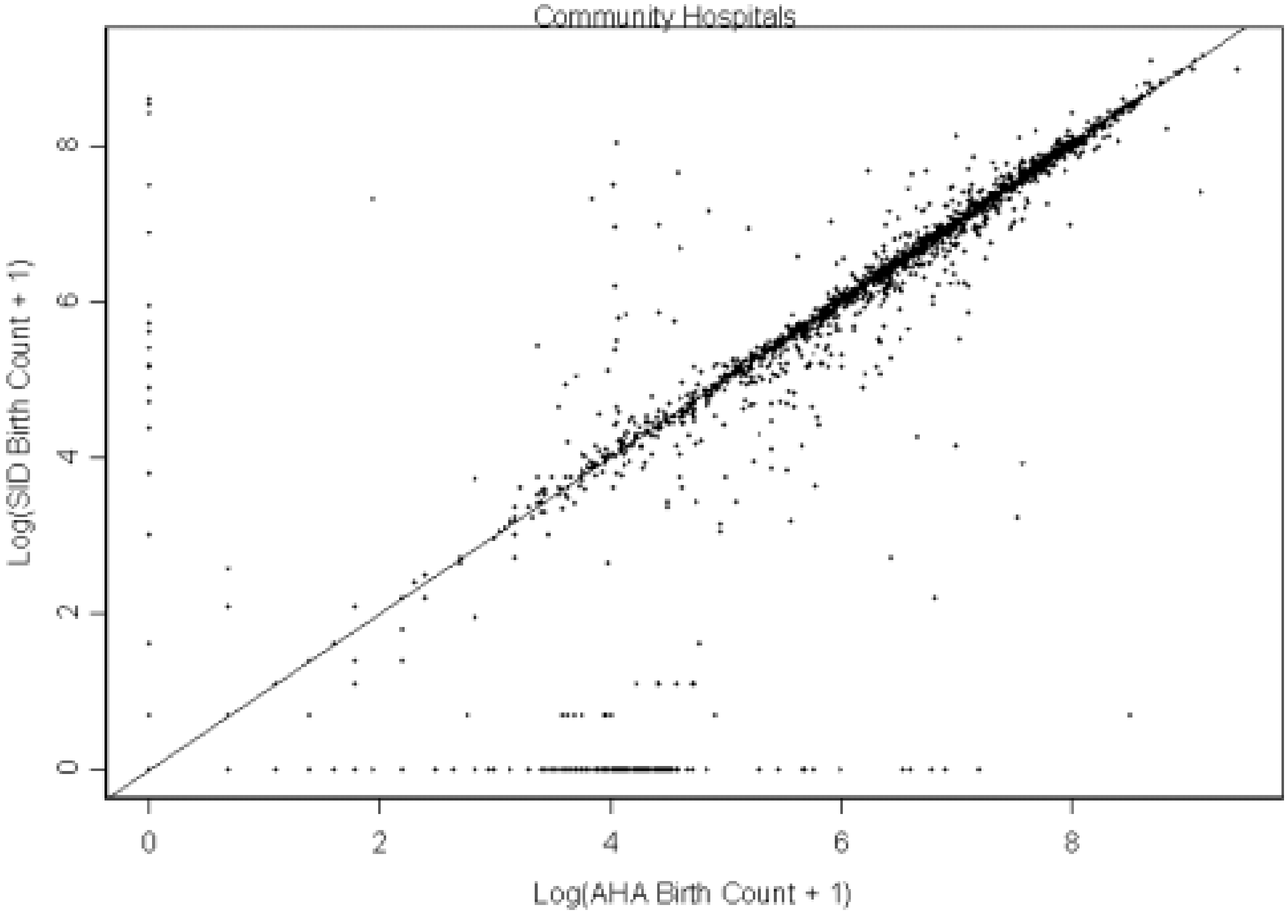


Table 8 shows the variation in the percentage of pediatric discharges and in the pediatric average lengths of stay (LOS) in the 2,581 HCUP SID community hospitals with pediatric discharges for each of the NIS stratification variables. Both the percentage of pediatric discharges and the average lengths of stay vary with the hospital strata. For example, the pediatric average length of stay is higher for urban hospitals than for rural hospitals, higher for teaching hospitals than for nonteaching hospitals, and higher for larger hospitals than for smaller hospitals within each location. Consequently, discharge weights based on these stratification variables should yield more precise estimates than discharge weights without post-stratification.

We post-stratified the HCUP SID hospitals using the same stratification variables that were used for the NIS sample plus the hospital type (children's or other); and stratified HCUP SID discharges by whether the discharge was an uncomplicated in-hospital birth, a complicated in-hospital birth, or a non-newborn pediatric discharge.

In the HCUP SID, approximately 60 percent of pediatric discharges are for births. The cost and utilization for uncomplicated births differ from those of other pediatrics. Therefore, stratification on uncomplicated birth status should pay dividends in terms of better weights and better estimates for pediatrics as a whole.

The proportion of discharges that are pediatric cases at each hospital varies with the type of hospital. Table 9 lists the count of hospitals in the HCUP SID for each type. Overall, 92 percent of pediatric discharges occur in general medical and surgical hospitals (type=10). Actually, more than 8 percent of pediatric discharges occur in children's hospitals because many children's hospitals are units of larger institutions, and their discharges are reported as a part of the larger institution, which is not classified as a children's hospital.

Table 8. HCUP SID Pediatric Statistics by Strata

STRATA	Community Hospitals	Total Discharges	Percent Pediatric	Average length of stay
Region				
Northeast	614	6,397,846	16.9	3.85
Midwest	667	3,464,336	18.9	3.35
South	564	4,683,011	17.6	3.56
West	736	5,635,200	22.7	3.02
Ownership				
Public	542	2,540,652	21.0	3.83
Voluntary	1,655	15,395,689	19.1	3.43
Proprietary	384	2,244,052	16.3	2.79
Teaching/Location/Bedsize				
Rural Total	880	2,210,839	17.5	2.41
Small	498	462,278	16.0	2.03
Medium	217	597,780	17.4	2.18
Large	165	1,150,781	18.2	2.68
Urban Nonteaching Total	1,253	9,943,179	17.5	2.91
Small	356	835,078	16.8	2.22
Medium	478	3,152,822	17.7	2.62
Large	419	5,955,279	17.6	3.16
Urban Teaching Total	448	8,026,375	21.3	4.18
Small	190	1,903,166	23.9	3.86
Urban	159	2,976,032	20.7	3.94
Large	99	3,147,177	20.3	4.65
TOTAL	2,581	20,180,393	19.0	3.47

Table 9. Breakdown by AHA Service Code Descriptions

Type Code	Description	HCUP SID Hospitals	SID Total Discharges	SID Pediatric Discharges	Percent Pediatric
10	General medical and surgical	2,471	19,738,688	3,585,445	18.2
44	Obstetrics and gynecology	4	61,837	19,243	31.1
45	Eye, ear, nose and throat	5	6,912	1,063	15.4
46	Rehabilitation	42	52,775	1,090	2.1
47	Orthopedic	5	16,949	2,232	13.2
49	Other specialty	25	61,069	2,837	4.6
50	Children’s general	21	181,871	174,680	96.1
57	Children’s orthopedic	2	1,722	1,624	94.2
59	Children’s other specialty	6	58,570	53,055	90.6
TOTAL		2,581	20,180,393	3,841,269	19.0

Figure 3 contains box plots that summarize the distribution of the proportion of pediatric discharges across hospitals of each type. In the plot, the vertical axis runs from 0 to 1 indicating the proportion of pediatric discharges. The shaded area of each box is bounded below by the 25th percentile and is bounded above by the 75th percentile. The white line within the shaded area marks the median (50 th percentile). The thin lines extending from the top and bottom of the boxes extend to upper and lower outlier thresholds, respectively. The horizontal lines drawn above and below the outlier thresholds mark the locations of the outliers themselves. For example, for general hospitals (hospital type 10) the 25th percentile is about .10, the median is about .15, the 75th percentile is about .21, and the upper outlier threshold is about .39. Thus, 25 percent of general hospitals have fewer than 10 percent pediatric discharges, 50 percent of general hospitals have between 10 percent and 21 percent pediatric discharges, and 25 percent have more than 21 percent pediatric discharges.

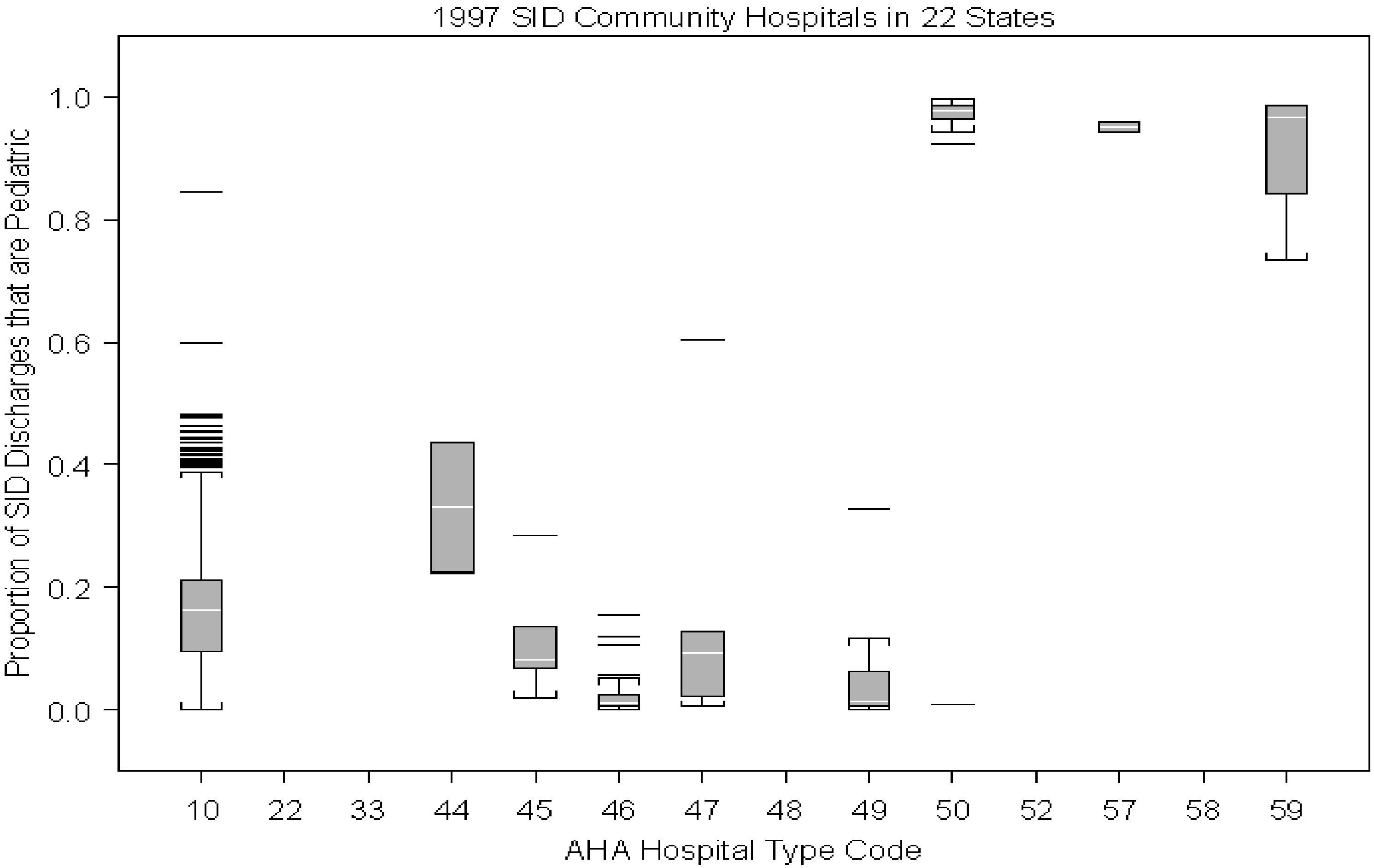
Figure 3 makes it clear that nearly all discharges are pediatric discharges in pediatric hospitals (types 50 through 59). We noted that one type 50 hospital has nearly zero pediatric discharges in the HCUP SID (marked by a horizontal line near zero). We learned that the AHA hospital type is wrong for this institution and we corrected it in the stratification. We also noted one type 10 hospital with 85 percent pediatric discharges. We learned that this hospital is a children's hospital, so the AHA hospital type was also wrong for it. We also corrected it for the stratification.

Based on the plot in Figure 3, we would benefit by two separate strata for:

1. children's hospitals (types 50 through 59), and
2. other hospitals (all other types).

NACHRI data were used 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 within them.

Figure 3 - Distribution of Pediatric Proportion by Hospital Type



Discharge-Level Sampling Weights

The discharge weights usually are constant for all discharges of the same type (uncomplicated in-hospital birth, complicated in-hospital birth, 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 NIS. For those hospitals, we *adjusted* the number of observed discharges by a factor 4 ÷ Q, where Q was the number of calendar quarters that the hospital contributed discharges to the NIS. For example, when a sample hospital contributed only two quarters of discharge data to the NIS, 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.

Universe Discharge Weights

Discharge weights to the universe were calculated by post-stratification. Hospitals were stratified on geographic region, urban/rural location, teaching status, bedsize, 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):
1.T_k is the total number of births reported in the AHA survey, and
2.A_k is the total number of adjusted births in the restricted sampling frame.
3.In the uncomplicated birth strata, R_k is the frame sampling rate for uncomplicated in-hospital births calculated as (sum of adjusted number sampled)/(sum of adjusted number in the restricted frame).
4.In the complicated birth strata, R_k is the frame sampling rate for complicated in-hospital births.

In the non-newborn strata:
5.T_k is the total number of non-newborns reported in the AHA survey,
6.A_k is the total number of adjusted non-newborn discharges in the sampling frame, and
7.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 pediatric cases because the variation in hospital outcomes for uncomplicated births is considerably less than that for other pediatrics and because we expect research to focus much more on other pediatrics. We sampled uncomplicated births at the nominal rate of 10 percent and sampled other pediatric discharges 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:

$$R_k, = S_k / H_k$$

8.S_k is the number of adjusted discharges sampled for the discharge type in stratum k
9.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 discharges that are non-birth pediatric across the HCUP SID hospitals is the same as the proportion of discharges that are non-birth pediatric across the universe of AHA hospitals, in the aggregate within each stratum.

DATA ANALYSIS

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 could 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 sampling without replacement could 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 are listed below that calculate statistics and their variances from sample survey data. 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.

In most cases, computer programs are readily available to perform these calculations. For instance, Stata and SUDAAN do calculations for numerous statistics arising from the stratified sampling design.

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 specific to 1997 should be governed by finite-sample theory.

Alternatively, in the study of hypothetical population outcomes not limited to a specific point in time, 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, since the population is defined at that point in time, and because the estimate is for a characteristic as it existed at the time of sampling. 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. Under the superpopulation (stochastic) model, procedures (such as those described by Potthoff, Woodbury, and Manton[1]) have been developed to draw inferences using weights from complex samples. In this context, the survey weights are not used to weight the sampled cases to the universe, because the universe is conceptually infinite in size. Instead, these weights are used to produce unbiased estimates of parameters that govern the superpopulation.

In summary, 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 should be used to weight the sample data in estimating population statistics.

Several statistical programming packages allow weighted analyses.[2] For example, nearly all SAS (Statistical Analysis System) procedures incorporate weights.

In addition, several statistical analysis programs have been developed that specifically calculate statistics and their standard errors from survey data. For an excellent review of such programs, visit the following web site: http://www.fas.harvard.edu/~stats/survey-soft/.

The KID database includes a Hospital Weights file with variables required by these programs to calculate finite population statistics. In addition to the sample weights described earlier, a hospital identifier (HOSPID), stratification variables, and stratum-specific totals for the numbers of discharges and hospitals are included so that finite-population corrections (FPCs) 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. At 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. To the extent that hospital-level outcomes correlate with these variables, they may be used to sharpen regional and nationwide estimates.

ENDNOTES

1. Potthoff, R.F., M.A. Woodbury, and K.G. Manton (1992). "Equivalent Sample Size" and "Equivalent Degrees of Freedom" Refinements for Inference Using Survey Weights Under Superpopulation Models. *Journal of the American Statistical Association*, Vol. 87, 383-396.

2. Carlson, B.L., A.E. Johnson, and S.B. Cohen (1993). An Evaluation of the Use of Personal Computers for Variance Estimation with Complex Survey Data. *Journal of Official Statistics*, Vol. 9, No. 4, 795-814.