HEALTHCARE COST AND UTILIZATION PROJECT — HCUP
A FEDERAL-STATE-INDUSTRY PARTNERSHIP IN HEALTH DATA
Sponsored by the Agency for Healthcare Research and Quality

INTRODUCTION TO

THE HCUP NATIONWIDE INPATIENT SAMPLE (NIS)

2006

These pages provide only an introduction to the NIS package.

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

Issued May 2008
Updated November 2015

Agency for Healthcare Research and Quality
Healthcare Cost and Utilization Project (HCUP)

Phone: (301) 427-1410
E-mail: hcup@ahrq.gov
Website: http://www.hcup-us.ahrq.gov

NIS Data and Documentation Distributed by:
HCUP Central Distributor

Phone: (866) 556-4287 (toll-free)
Fax: (866) 792-5313
E-mail: HCUPDistributor@ahrq.gov

Table of Contents

SUMMARY OF DATA USE LIMITATIONS 1

HCUP CONTACT INFORMATION 2

WHAT'S NEW IN THE 2006 NATIONWIDE INPATIENT SAMPLE (NIS)? 3

UNDERSTANDING THE NIS 3

ABSTRACT 4

INTRODUCTION TO THE HCUP NATIONWIDE INPATIENT SAMPLE (NIS) 5

- Overview of NIS Data 5
  - *NIS Data Sources, Hospitals, and Inpatient Stays 6*
  - *State-Specific Restrictions 6*
  - *Contents of CD-ROM Set 6*
  - *NIS Data Elements 7*
- Getting Started 8
  - *NIS Data Files 8*
  - *NIS Documentation 8*

HOW TO USE THE NIS FOR DATA ANALYSIS 8

- Calculating National Estimates 9
- Why the NIS Should Not Be Used to Make State-Level Estimates 10
- Studying Trends 11
- Choosing Data Elements for Analysis 11
- ICD-9-CM Diagnosis and Procedure Codes 12
- Missing Values 12
- Variance Calculations 13
- Computer Software for Variance Calculations 14
- Longitudinal Analyses 15
- Discharge Subsamples 15

SAMPLING OF HOSPITALS 16

- Sampling of Hospitals Included in the NIS 16
  - *The NIS Hospital Universe 16*
  - *Hospital Merges, Splits, and Closures 16*
  - *Stratification Variables 17*
- Hospital Sampling Frame 19
- Hospital Sample Design 20
  - *Design Considerations 20*
  - *Overview of the Sampling Procedure 21*
  - *Subsamples 21*
  - *Change to Hospital Sampling Procedure Beginning with the 1998 NIS 21*
  - *Zero-Weight Hospitals 22*

Final Hospital Sample....22
SAMPLE WEIGHTS....23
Hospital Weights....23
Discharge Weights....24
APPENDIX I: TABLES AND FIGURES....1
Table 1: 2006 Data Sources....1
Table 2: Number of NIS States, Hospitals, and Discharges, by Year....3
Table 3. Summary of NIS Releases....4
Table 4. Summary of NIS Data Sources, Hospitals, and Inpatient Stays, 1988-2006....5
Table 5. NIS Related Reports and Database Documentation Available on HCUP-US....7
Figure 1: Hospital Universe, by Year....8
Figure 2: NIS States, by Region....9
Table 6: All States, by Region....10
Table 7: Bed Size Categories, by Region....11
Figure 3: NIS Hospital Sampling Frame, by Year....12
Figure 4: Number of Hospitals in the 2006 Universe, Frame, and Sample for Frame States....13
Part A: Arkansas – North Carolina....13
Part B: Nebraska – West Virginia....14
Table 8: Number of Hospitals and Discharges in 2006 AHA Universe, Frame, and NIS, by State....15
Figure 5: Number of Hospitals Sampled, by Year....17
Figure 6: Number of NIS Discharges, Unweighted, by Year....18
Figure 7: Number of NIS Discharges, Weighted, by Year....19
Figure 8: Number of Hospitals in the 2006 Universe, Frame, Sample, Target, and Surplus, by Region....20
Figure 9: Percentage of U.S. Population in 2006 NIS States, by Region....21
Figure 10: Number of Discharges in the 2006 NIS, by State....22
APPENDIX II: STATE-SPECIFIC RESTRICTIONS....1
APPENDIX III: DATA ELEMENTS....1
Table 1. Data Elements in the NIS Inpatient Core Files....1
Table 2. Data Elements in the NIS Hospital Weights Files....9
Table 3. Data Elements in the NIS Disease Severity Measures Files....12
Table 4. Data Elements in the NIS Diagnosis and Procedure Groups Files....14

HCUP NATIONWIDE INPATIENT SAMPLE (NIS)
SUMMARY OF DATA USE LIMITATIONS

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

All users of the NIS must take the on-line HCUP Data Use Agreement (DUA) training course, and read and sign a Data Use Agreement.[†]

Authorized users of HCUP data agree to the following restrictions:[‡]

- Will not use the data for any purpose other than research or aggregate statistical reporting.
- Will not re-release any data to unauthorized users.
- Will not redistribute HCUP data by posting on any Web site or publicly-accessible online repository.
- Will not identify or attempt to identify any individual, including by the use of vulnerability analysis or penetration testing. Methods that could be used to identify individuals directly or indirectly shall not be disclosed or published.
- Will not publish information that could identify individual establishments (e.g., hospitals) and will not contact establishments.
- Will not use the data concerning individual establishments for commercial or competitive purposes involving those establishments and will not use the data to determine rights, benefits, or privileges of individual establishments.
- Will not use data elements from the proprietary severity adjustment software packages (3M APR-DRGs, HSS APS-DRGs, and Truven Health Analytics Disease Staging) for any commercial purpose or to disassemble, decompile, or otherwise reverse engineer the proprietary software.
- Will acknowledge in reports that data from the "Healthcare Cost and Utilization Project (HCUP)," were used, including names of the specific databases used for analysis.
- Will acknowledge that risk of individual identification of persons is increased when observations (i.e., individual discharge records) in any given cell of tabulated data is less than or equal to 10.

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

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

HCUP CONTACT INFORMATION

All HCUP data users, including data purchasers and collaborators, must complete the online HCUP Data Use Agreement (DUA) Training Tool, and read and sign the HCUP Data Use Agreement. Proof of training completion and signed Data Use Agreements must be submitted to the HCUP Central Distributor as described below.

The on-line DUA training course is available at: http://www.hcup-us.ahrq.gov/tech_assist/dua.jsp

The HCUP Nationwide Data Use Agreement is available on the AHRQ-sponsored HCUP User Support (HCUP-US) Web site at: http://www.hcup-us.ahrq.gov

HCUP Central Distributor

Data purchasers will be required to provide their DUA training completion code and will execute their DUAs electronically as a part of the online ordering process. The DUAs and training certificates for collaborators and others with access to HCUP data should be submitted directly to the HCUP Central Distributor using the contact information below.

The HCUP Central Distributor can also help with questions concerning HCUP database purchases, your current order, training certificate codes, or invoices, if your questions are not covered in the Purchasing FAQs on the HCUP Central Distributor Web site.

Purchasing FAQs:
https://www.distributor.hcup-us.ahrq.gov/Purchasing-Frequently-Asked-Questions.aspx

Phone: (866) 290-HCUP (4287)
Email: HCUPDistributor@AHRQ.gov
Fax: 866-792-5313 (toll free in the United States)

Mailing address:
HCUP Central Distributor
Social & Scientific Systems, Inc.
8757 Georgia Ave, 12th Floor
Silver Spring, MD 20910

HCUP User Support:

Information about the content of the HCUP databases is available on the HCUP User Support (HCUP-US) Web site (http://www.hcup-us.ahrq.gov). If you have questions about using the HCUP databases, software tools, supplemental files, and other HCUP products, please review the HCUP Frequently Asked Questions or contact HCUP User Support:

HCUP FAQs: http://www.hcup-us.ahrq.gov/tech_assist/faq.jsp

Phone: 866-290-HCUP (4287) (toll free)
Email: hcup@ahrq.gov

WHAT'S NEW IN THE 2006 NATIONWIDE INPATIENT SAMPLE (NIS)?

- Virginia returned to the NIS in 2006.
- The data element HOSPBRTH, which indicates an in-hospital birth, is added to the NIS Core file beginning in 2006.
- Also in the 2006 NIS Core file, the version 18 DRG and MDC are replaced with the version 24 DRG and MDC.
- Fourth quarter data from sampled hospitals in Massachusetts were unavailable for inclusion in the 2006 NIS; however, we adjusted the data to account for missing cases. For details, see the section on the Final Hospital Sample in this document.
- The NIS Introduction and the NIS Design Report were combined and reorganized for 2006. Data tables, figures, and State-specific restrictions now appear as appendices for the combined document.
- 2006 NIS Documentation is available exclusively on the HCUP User Support (HCUP-US) Website (http://www.hcup-us.ahrq.gov) and is no longer included on the NIS CD-ROMs. This ensures that documentation for your data will always be the most recent and up-to-date version.
- Users must complete an on-line Data Use Agreement training tool prior to receiving the data.

UNDERSTANDING THE NIS

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

The Agency for Healthcare Research and Quality and the staff of the Healthcare Cost and Utilization Project (HCUP) thank you for purchasing the HCUP Nationwide Inpatient Sample (NIS).

HCUP Nationwide Inpatient Sample (NIS)

ABSTRACT

The Nationwide Inpatient Sample (NIS) is part of the Healthcare Cost and Utilization Project (HCUP), sponsored by the Agency for Healthcare Research and Quality (AHRQ), formerly the Agency for Health Care Policy and Research.

The NIS is a database of hospital inpatient stays. Researchers and policy makers use the NIS to identify, track, and analyze national trends in health care utilization, access, charges, quality, and outcomes.

The NIS is the largest all-payer inpatient care database that is publicly available in the United States, containing data from 5 to 8 million hospital stays from about 1,000 hospitals sampled to approximate a 20-percent stratified sample of U.S. community hospitals. The NIS is drawn from those States participating in HCUP and weights are provided to calculate national estimates. See Table 1 in Appendix I for a list of the statewide data organizations participating in the NIS. The number of sample hospitals and discharges by State and year are available in Table 2 in Appendix I.

The NIS is available yearly, beginning with 1988, allowing analysis of trends over time. (Analyses of time trends are recommended from 1993 forward. See the report, *Using the HCUP Nationwide Inpatient Sample to Estimate Trends*, available on the HCUP User Support (HCUP-US) Website, for details.)

The NIS is the only national hospital database with charge information on all patients, regardless of payer, including persons covered by Medicare, Medicaid, private insurance, and the uninsured. The NIS' large sample size enables analyses of rare conditions, such as congenital anomalies; uncommon treatments, such as organ transplantation; and special patient populations, such as the uninsured.

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

Beginning in 1998, the NIS differs from previous NIS releases: some data elements were dropped; some were added; for some data elements, the coding was changed; and the sampling and weighting strategy was revised to improve the representativeness of the data. (See the report, *Changes in the NIS Sampling and Weighting Strategy for 1998*, which describes these changes, available on the HCUP-US Website.) Periodically, new data elements are added to the NIS and some are dropped; see Appendix III for a summary of data elements and when they are effective.

Access to the NIS is open to users who sign data use agreements. Uses are limited to research and aggregate statistical reporting.

For more information on the NIS, please visit the AHRQ-sponsored HCUP-US Website at http://www.hcup-us.ahrq.gov.

INTRODUCTION TO THE HCUP NATIONWIDE INPATIENT SAMPLE (NIS)

Overview of NIS Data

The Nationwide Inpatient Sample (NIS) contains all-payer data on hospital inpatient stays from States participating in the Healthcare Cost and Utilization Project (HCUP). Each year of the NIS provides information on approximately 5 million to 8 million inpatient stays from about 1,000 hospitals. All discharges from sampled hospitals are included in the NIS database.

The NIS contains clinical and resource use information included in a typical discharge abstract. The NIS can be linked directly to hospital-level data from the American Hospital Association (AHA) Annual Survey Database (Health Forum, LLC © 2007) and to county-level data from the Health Resources and Services Administration Bureau of Health Professions' Area Resource File (ARF), except in those States that do not allow the release of hospital identifiers.

The NIS is designed to approximate a 20-percent sample of U.S. community hospitals, defined by the AHA to be "all non-Federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions." Included among community hospitals are specialty hospitals such as obstetrics-gynecology, ear-nose-throat, short-term rehabilitation, orthopedic, and pediatric institutions. Also included are public hospitals and academic medical centers. Starting in 2005, the AHA included long term acute care facilities in the definition of community hospitals. These facilities provide acute care services to patients who need long term hospitalization (stays of more than 25 days). Excluded from the NIS are short-term rehabilitation hospitals (beginning with 1998 data), long-term non-acute care hospitals, psychiatric hospitals, and alcoholism/chemical dependency treatment facilities.

This universe of U.S. community hospitals is divided into strata using five hospital characteristics: ownership/control, bed size, teaching status, urban/rural location, and U.S. region.

The NIS is a stratified probability sample of hospitals in the frame, with sampling probabilities proportional to the number of U.S. community hospitals in each stratum. The frame is limited by the availability of inpatient data from the data sources currently participating in HCUP.

In order to improve the representativeness of the NIS, the sampling and weighting strategy was modified beginning with the 1998 data. The full description of this process can be found in the special report on *Changes in NIS Sampling and Weighting Strategy for 1998.* This report is available on the AHRQ-sponsored HCUP-US Website at http://www.hcup-us.ahrq.gov. To facilitate the production of national estimates, both hospital and discharge weights are provided, along with information necessary to calculate the variance of estimates. Detailed information on the design of the NIS prior to 2006 is available in the year-specific special reports on *Design of the Nationwide Inpatient Sample* found on the HCUP-US Website (http://hcup-us.ahrq.gov/db/nation/nis/nisrelatedreports.jsp). Starting with the 2006 NIS, the information on the design of the NIS was incorporated into this report.

NIS data sets are available yearly, beginning with 1988. See Table 3 in Appendix I for a summary of NIS releases. Each release of the NIS includes:

- Data in fixed-width ASCII format on CD-ROM.
- Patient-level hospital discharge abstract data for 100% of discharges from a sample of hospitals in participating States.
- 5 million to 8 million inpatient records per year.
- 800-1,100 hospitals per year.
- Two 10% subsamples of discharges from all NIS hospitals (only available prior to the 2005 NIS).
- Discharge-level weights to calculate national estimates for discharges.
- Hospital Weights File to produce national estimates for hospitals and to link the NIS to data from the AHA Annual Survey Database (Health Forum, LLC © 2007)
- NIS Documentation and tools – including file specifications, programming source code for loading ASCII data into SAS and SPSS, and value labels. Beginning in 2005, code is also provided for loading the NIS ASCII file into Stata.

NIS Data Sources, Hospitals, and Inpatient Stays

Table 4 in Appendix I contains a summary of the data sources, number of hospitals, and number of unweighted and weighted inpatient stays in NIS data.

State-Specific Restrictions

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

Contents of CD-ROM Set

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

CD-ROM #1 contains:

Inpatient Core File: This inpatient discharge-level file contains data for 100% of the discharges from a sample of hospitals in participating States. The unit of observation is an *inpatient stay record*. Refer to Table 1 in Appendix III for a list of data elements in the Inpatient Core File. This file is available in all years of the NIS.

Hospital Weights File: This hospital-level file contains one observation for each hospital included in the NIS and contains weights and variance estimation data elements, as well as linkage data elements. The unit of observation is the *hospital*. The HCUP hospital identifier (HOSPID) provides the linkage between the NIS Inpatient Core files and the Hospital Weights file. A list of data elements in the Hospital Weights File is provided in Table 2 of Appendix III. This file is available in all years of the NIS.

CD-ROM #2 contains:

Disease Severity Measures Files: These discharge-level files contain information from four different sets of disease severity measures. Information from these severity files is to be used in conjunction with the Inpatient Core files. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (KEY) provides the linkage between the Core files and the Disease Severity Measures files. Refer to Table 3 in Appendix III for a list of data elements in the Severity Measures files. These files are available beginning with the 2002 NIS.

Diagnosis and Procedure Groups Files: These discharge-level files contain data elements from AHRQ software tools designed to facilitate the use of the ICD-9-CM diagnostic and procedure information in the HCUP databases. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (KEY) provides the linkage between the Core files and the Diagnosis and Procedure Groups files. Table 4 in Appendix III contains a list of data elements in the Diagnosis and Procedure Groups files. These files are available beginning with the 2005 NIS.

On the HCUP-US Website (http://www.hcup-us.ahrq.gov), NIS purchasers can access complete file documentation, including variable notes, file layouts, summary statistics, and related technical reports. Similarly, purchasers can also download SAS, SPSS, and Stata load programs. Available online documentation and supporting files are detailed in Appendix I, Table 5.

NIS Data Elements

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

- Table 1 for the Inpatient Core files (record = inpatient stay)
- Table 2 for the Hospital Weights files (record = hospital)
- Table 3 for the Disease Severity Measures files (record = inpatient stay)
- Table 4 for the Diagnosis and Procedure Groups files (record = inpatient stay).

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

Getting Started

The NIS data files are provided on CD-ROMs. The NIS Inpatient Core and Hospital Weights files are on CD-ROM #1, while the Disease Severity Measures and Diagnosis and Procedure Groups files are on CD-ROM #2. Comprehensive documentation for the NIS files is available on the HCUP-US Website (http://www.hcup-us.ahrq.gov).

NIS Data Files

In order to load and analyze the NIS data onto your PC, you will need 13 gigabytes of space available. Because of the size of the files, the data are distributed as self-extracting PKZIP compressed files. To decompress the data, you should follow these steps:

1. Create a directory for the 2006 NIS on your hard drive.
2. Copy the self-extracting data files from the NIS CD-ROMs into the new directory.
3. Unzip each file by running the corresponding *.exe file.
   - Type the file name within DOS or click on the name within Windows Explorer.
   - Edit the name of the "Unzip To Folder" in the WinZip Self-Extractor dialog to select the desired destination directory for the extracted file.
   - Click on the "Unzip" button.

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

NIS Documentation

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

- To locate the NIS documentation on HCUP-US, choose “HCUP Databases” from the home page (http://www.hcup-us.ahrq.gov). The first section under Nationwide HCUP Databases is specific to the NIS.

Table 5 in Appendix I details both the NIS related reports and the comprehensive NIS database documentation available on HCUP-US.

HOW TO USE THE NIS FOR DATA ANALYSIS

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

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

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

Calculating National Estimates

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

NIS Year	Name of Discharge Weight on the **Core File** to Use for Creating Nationwide Estimates	Name of Discharge Weight on the **10% Subsample Core File** to Use for Creating Nationwide Estimates
2005 forward	• DISCWT for all analyses	• *The 10% Subsample Core File was discontinued with the 2005 NIS.*
2001 - 2004	• DISCWT for all analyses	• DISCWT10 for all analyses
2000	• DISCWT to create nationwide estimates for all analyses except those that involve total charges. • DISCWTCHARGE to create nationwide estimates of total charges.	• DISCWT10 to create nationwide estimates for all analyses, except those that involve total charges. • DISCWTCHARGE10 to create nationwide estimates of total charges.
1998-1999	• DISCWT for all analyses	• DISCWT10 for all analyses
1988-1997	• DISCWT_U for all analyses	• D10CWT_U for all analyses

- Because the NIS is a stratified sample, proper statistical techniques must be used to calculate standard errors and confidence intervals. For detailed instructions, refer to the special report *Calculating Nationwide Inpatient Sample Variances* on the HCUP-US Website.

- The NIS Comparison Report assesses the accuracy of NIS estimates. The updated report for the current NIS will be posted on the HCUP-US Website (www.hcup-us.ahrq.gov) as soon as it is completed.

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

against which to check your national estimates for hospitalizations with more than 5,000 discharges.

- To ensure that you are using the weights appropriately and calculating estimates and variances accurately, you can also use HCUPnet, the free online query system (http://www.hcupnet.ahrq.gov). HCUPnet is a Web-based query tool for identifying, tracking, analyzing, and comparing statistics on hospitals at the national, regional, and State level. HCUPnet offers easy access to national statistics and trends and selected State statistics about hospital stays. This tool provides step-by-step guidance, helping researchers to quickly obtain the statistics they need. HCUPnet generates statistics using the NIS, KID, and SID for those States that have agreed to participate. In addition, HCUPnet provides Quick Statistics – ready-to-use tables on commonly requested information – as well as national statistics based on the AHRQ Quality Indicators.

Why the NIS Should Not Be Used to Make State-Level Estimates

AHRQ strongly advises researchers against using the NIS to estimate State-specific statistics. Prior to 2012, State is available as a NIS data element. However, these NIS samples were not designed to yield a representative sample of hospitals at the State level. AHRQ recommends that researchers employ the SID for State-level estimates.

Each NIS sample is drawn from the sampling frame consisting of discharge data submitted by HCUP Partners-statewide data organizations that agree to participate in the NIS. Data from non-Partner States are missing completely from the sampling frame, and data from Partner States are sometimes incomplete because of different State reporting requirements, different State restrictions, or other data omissions. The NIS is designed to represent hospitals and discharges nationally, including those outside the sampling frame.

To accomplish this, within each hospital sampling stratum the NIS draws a number of hospitals from the sampling frame required to net a total of 20 percent of hospitals nationally. The sampling strata are defined by census region (4 regions), hospital ownership (3 categories), urban-rural location, teaching status, and bed size (3 categories). As a result, the proportion of NIS hospitals in a stratum that are from a given State is unlikely to equal the State's actual proportion of hospitals in that stratum. Consequently, the sample of NIS hospitals is unlikely to be representative of hospitals in the State, and the NIS sample weights will not be appropriate at the State level.

The level of this “misrepresentation” varies across the States in any given year of the NIS, which further confounds State-to-State comparisons on the basis of State-specific estimates from the NIS. Moreover, for a given State the level of misrepresentation changes from year to year as States (and hospitals) enter and exit the sampling frame over time. This further confounds State-specific trends on the basis of State-specific estimates from the NIS.

Finally, because the NIS was not designed to be representative at the State level, design-based estimates of standard errors are not possible, which severely hampers State-level inferences. Moreover, the NIS is composed of all discharges from a sample of hospitals (a cluster sample). The hospital-to-hospital variation and the small number of hospitals available in the NIS for many States make Statelevel estimates very imprecise at best and biased at worst.

Studying Trends

- When studying trends over time using the NIS, be aware that the sampling frame for the NIS changes almost annually (i.e., more States have been added over time). Estimates from earlier years of the NIS may be subject to more sampling bias than later years of the NIS. In order to facilitate analysis of trends using multiple years of NIS data, an alternate set of NIS discharge and hospital weights for the 1988-1997 HCUP NIS were developed. These alternative weights were calculated in the same way as the weights for the 1998 and later years of the NIS. The report, *Using the HCUP Nationwide Inpatient Sample to Estimate Trends* includes details regarding the alternate weights and other recommendations for trends analysis. Both the NIS trends report and the alternate weights are available on the HCUP-US Website under Methods Series (http://www.hcup-us.ahrq.gov/reports/methods.jsp).

- To ease the burden on researchers conducting analyses that span multiple years, NIS trends supplemental files (NIS-Trends) are available through the HCUP Central Distributor. The NIS-Trends annual files contain the alternate trend weights for data prior to 1997, in addition to renamed, recoded, and new data elements consistent with the later years of the NIS. More information on these files is available on the HCUP-US Website under NIS database documentation (http://www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp).

- Short-term rehabilitation hospitals are included in the 1988-1997 NIS, but are excluded from the NIS beginning in 1998. 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 impact trends but the effect is likely small since only about 3% of community hospitals are short-term rehabilitation hospitals and not all State data sources included these hospitals. The NIS-Trends weights account for this change in NIS sampling.

Choosing Data Elements for Analysis

- For all data elements you plan to use in your analysis, first perform descriptive statistics and examine the range of values, including the number of missing cases. Summary statistics for the entire NIS are provided on the Summary Statistics page of the HCUP-US Website (http://www.hcup-us.ahrq.gov/db/nation/nis/nissummstats.jsp). When you detect anomalies (such as large numbers of missing cases), perform descriptive statistics by State for that variable to detect if there are State-specific differences. Sometimes performing descriptive statistics by hospital can be helpful in detecting hospital-specific data anomalies.

- Not all data elements in the NIS are provided by each State data source. These data elements are provided on the NIS because they can be valuable for research purposes but they should be used cautiously. For example, RACE is missing for a number of States; thus, national estimates using RACE should be interpreted and reported with caveats. Check the documentation and run frequencies by State to identify if a data element is not available in one or more States.

- Differences exist across the State data sources in the collection of information that could not be accounted for during HCUP processing to make the data uniform. Be sure to read State-specific notes for each data element that you use in your analysis – this information can be found on the Description of Data Elements page on the HCUP-US Website (http://www.hcup-us.ahrq.gov/db/nation/nis/nisdde.jsp).

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

ICD-9-CM Diagnosis and Procedure Codes

- ICD-9-CM diagnosis and procedure codes provide valuable insights into the reasons for hospitalization and what procedures patients receive, but these codes need to be carefully used and interpreted. ICD-9-CM codes change every October as new codes are introduced and some codes are retired. See the Conversion Table at http://www.cdc.gov/nchs/datawh/ftpserv/ftpicd9/ftpicd9.htm which shows ICD-9-CM code changes over time. **It is critical to check all ICD-9-CM code used for analysis to ensure the codes are in effect during the time period studied.**

- Although the NIS contains up to 15 diagnoses and 15 procedures, the number of diagnoses and procedures varies by State. Some States provide as many as 30 diagnoses and 21 procedures, while other States provide as few as 10 diagnoses and 6 procedures. Because very few cases have more than 15 diagnoses or procedures, the diagnosis and procedure vectors were truncated to save space in the NIS data files. Two variables are provided which tell you exactly how many diagnoses and procedures were on the original records (NDX and NPR).

- The collection and reporting of external cause of injury (E codes) varies greatly across States. Some States have laws or mandates for the collection of E codes; others do not. Some States do not require hospitals to report E codes in the range E870-E879 - “misadventures to patients during surgical and medical care” - which means that these occurrences will be underreported. Be sure to read the State-specific notes on diagnoses for more details; this information can be found on the Description of Data Elements page on the HCUP-US Website (http://www.hcup-us.ahrq.gov/db/nation/nis/nisdde.jsp).

Missing Values

Missing data values can compromise the quality of estimates. If the outcome for discharges with missing values is different from the outcome for discharges with valid values, then sample estimates for that outcome will be biased and inaccurately represent the discharge population. There are several techniques available to help overcome this bias. One strategy is to use imputation to replace missing values with acceptable values. Another strategy is to use sample weight adjustments to compensate for missing values.[1] Descriptions of such data preparation and adjustment are outside the scope of this report; however, it is recommended that researchers evaluate and adjust for missing data, if necessary.

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

Variance Calculations

It may be important for researchers to calculate a measure of precision for some estimates based on the NIS sample data. Variance estimates must take into account both the sampling design and the form of the statistic. The sampling design consisted of a stratified, single-stage cluster sample. A stratified random sample of hospitals (clusters) was drawn and then *all* discharges were included from each selected hospital. **To accurately calculate variances from the NIS, you must use appropriate statistical software and techniques. For details,** see the special report, *Calculating Nationwide Inpatient Sample Variances*. This report is available on the HCUP-US Website at http://www.hcup-us.ahrq.gov/db/nation/nis/nisrelatedreports.jsp.

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. Standard formulas for a stratified, single-stage cluster sample without replacement could be used to calculate statistics and their variances in most applications.

A multitude of statistics can be estimated from the NIS 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.

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 NIS, any estimates that attempt to accurately describe characteristics and interrelationships among hospitals and discharges during a specific year should be governed by finite-sample theory. Examples of this would be estimates of expenditure and utilization patterns or hospital market factors.

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 hospital weights are useful for producing hospital-level statistics for analyses that use the *hospital* as the unit of analysis, while the discharge weights are useful for producing discharge-level statistics for analyses that use the *discharge* as the unit of analysis. The discharge weights may be used to estimate nationwide population statistics.

In most cases, computer programs are readily available to perform these calculations. Several statistical programming packages allow weighted analyses.[2] For example, nearly all SAS procedures incorporate weights. In addition, several statistical analysis programs have been developed to specifically calculate statistics and their standard errors from survey data. Version eight or later of SAS contains procedures (PROC SURVEYMEANS and PROC SURVEYREG) for calculating statistics based on specific sampling designs. STATA and SUDAAN are two other common statistical software packages that perform calculations for numerous statistics arising from the stratified, single-stage cluster sampling design. Examples of the use of SAS, SUDAAN, and STATA to calculate NIS variances are presented in the special report, *Calculating Nationwide Inpatient Sample Variances*. This report is available on the HCUP-US Website at http://www.hcup-us.ahrq.gov/db/nation/nis/nisrelatedreports.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 NIS database includes a Hospital Weights file with variables required by these programs 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 analytic file is too small to set aside a large validation sample, cross-validation techniques may be used. For example, ten-fold cross-validation would split the data into ten subsets of equal size. 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 NIS 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.

As a simple example, the number of Cesarean sections performed in each hospital would be correlated with their total number of deliveries. The figure for Cesarean sections must be obtained from discharge data, but the number of deliveries is available from AHA data. Thus, if a regression model can be fit predicting this procedure from deliveries based on the NIS data, that regression model can then be used to obtain hospital-specific estimates of the number of Cesarean sections for all hospitals in the AHA universe.

Longitudinal Analyses

Hospitals that continue in the NIS for multiple consecutive years are a subset of the hospitals in the NIS for any one of those years. Consequently, longitudinal analyses of hospital-level outcomes may be biased, if they are based on any subset of NIS hospitals limited to continuous NIS membership. In particular, such subsets would tend to contain fewer hospitals that opened, closed, split, merged, or changed strata. Further, the sample weights were developed as annual, cross-sectional weights, rather than longitudinal weights. Therefore, different weights might be required, depending on the statistical methods employed by the analyst.

One approach to consider in hospital-level longitudinal analyses is to use repeated-measure models that allow hospitals to have missing values for some years. However, the data are not actually missing for some hospitals, such as those that closed during the study period. In any case, the analyses may be more efficient (e.g., produce more precise estimates) if they account for the potential correlation between repeated measures on the same hospital over time, yet incorporate data from all hospitals in the sample during the study period.

Discharge Subsamples

Prior to the 2005 NIS, two non-overlapping 10% subsamples of NIS discharges were provided each year for analytic purposes. Beginning with the 2005 NIS, 10% subsamples are no longer provided on the NIS CD-ROMs. However, users may still draw their own subsamples, if desired. One use of 10% subsamples would be to validate models and obtain unbiased estimates of standard errors. That is, one subsample may be used to estimate statistical models, while the other subsample may be used to test the fit of those models on new data. This is a very important analytical step, particularly in exploratory studies, where one runs the risk of fitting noise in the data.

It is well known that the percentage of variance explained by a regression, R^2, is generally overestimated by the data used to fit a model. The regression model could be estimated from the first subsample and then applied to the second subsample. The squared correlation between the actual and predicted value in the second subsample is an unbiased estimate of the model's true explanatory power when applied to new data.

SAMPLING OF HOSPITALS

Sampling of Hospitals Included in the NIS

The NIS Hospital Universe

The hospital universe is defined as all hospitals located in the U.S. that are open during any part of the calendar year and designated as community hospitals in the AHA Annual Survey Database (Health Forum, LLC © 2007). The AHA defines community hospitals as follows: "All non-Federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions." Starting in 2005, the AHA included long term acute care facilities in the definition of community hospitals. These facilities provide acute care services to patients who need long term hospitalization (stays of more than 25 days). Consequently, Veterans Hospitals and other Federal facilities (Department of Defense and Indian Health Service) are excluded. Beginning with the 1998 NIS, we excluded short-term rehabilitation hospitals 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. Figure 1 in Appendix I displays the number of universe hospitals for each year based on the AHA Annual Survey Database (Health Forum, LLC © 2007).

For more information on how hospitals in the data set were mapped to hospitals as defined by the AHA, refer to the special report, *HCUP Hospital Identifiers*. For a list of all data sources, refer to Table 1 in Appendix I. Detailed information on the design of the NIS prior to 2006 is available in the year-specific special reports on *Design of the Nationwide Inpatient Sample* found on the HCUP-US Website. Starting with the 2006 NIS, the design information was incorporated into this report.

Hospital Merges, Splits, and Closures

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

Stratification Variables

Given the increase in the number of contributing States, the NIS team evaluated and revised the sampling and weighting strategy for 1998 and subsequent data years, in order to best represent the U.S. This included changes to the definitions of the strata variables, the exclusion of rehabilitation hospitals from the NIS hospital universe, and a change to the calculation of hospital universe discharges for the weights. A full description of this process can be found in the special report on Changes in NIS Sampling and Weighting Strategy for 1998. This report is available on the HCUP-US Website at http://www.hcup-us.ahrq.gov/db/nation/nis/nisrelatedreports.jsp. (A description of the sampling procedures and definitions of strata variables used from 1988 through 1997 can be found in the special report: *Design of the HCUP Nationwide Inpatient Sample, 1997*. This report is also available on the HCUP-US Website.)

The NIS sampling strata were defined based on five hospital characteristics contained in the AHA hospital files. Beginning with the 1998 NIS, 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 2 highlights the NIS States by region, and Table 6 lists the States that comprise each region. Both can be found in Appendix I.

2. *Control – government non-Federal (public), private not-for-profit (voluntary), and private investor-owned (proprietary)*. Depending on their control, 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, we stratified hospitals as public, voluntary, and proprietary. We used this stratification for Southern rural, Southern urban non-teaching, and Western urban non-teaching hospitals. For smaller strata – the Midwestern rural and Western rural hospitals – we used a collapsed stratification of public versus private, with the voluntary and proprietary hospitals combined to form a single “private” category. For all other combinations of region, location, and teaching status, no stratification based on control was advisable, given the number of hospitals in these cells.

3. *Location – urban or rural.* Government payment policies often differ according to this designation. Also, rural hospitals are generally smaller and offer fewer services than urban hospitals. Beginning with the 2004 NIS, we changed the classification of urban or rural hospital location for the sampling strata to use the newer Core Based Statistical Area (CBSA) codes, rather than the older Metropolitan Statistical Area (MSA) codes. The CBSA groups are based on 2000 Census data, whereas the MSA groups were based on 1990 Census data. Also, the criteria for classifying the counties differ. For more information on the difference between CBSAs and MSAs, refer to the U.S. Census Bureau Website (http://www.census.gov/population/www/estimates/metroarea.html).

   Previously, we classified hospitals in a MSA as urban hospitals, while we classified hospitals outside a MSA as rural hospitals. Beginning with the 2004 NIS, we categorized hospitals with a CBSA type of *Metropolitan* or *Division* as urban, while we designated hospitals with a CBSA type of *Micropolitan* or *Rural* as rural. This change contributed to a slight decline in the number of hospitals that were classified as rural and a

corresponding increase in the number of hospitals categorized as urban. For the 2003 NIS, 44.9% of hospitals in the AHA universe were classified as rural hospitals; for 2004, only 41.3% of AHA universe hospitals were classified as rural.

4. *Teaching Status – teaching or non-teaching.* The missions of teaching hospitals differ from non-teaching hospitals. In addition, financial considerations differ between these two hospital groups. Currently, the Medicare Diagnosis Related Group (DRG) payments are uniformly higher to teaching hospitals. Prior to the 1998 NIS, we considered a hospital to be a teaching hospital if it had any residents or interns and met one of the following two criteria:

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

   Beginning with the 1998 NIS, we considered a hospital to be a teaching hospital if it met any one of the following three criteria:

   - Residency training approval by the Accreditation Council for Graduate Medical Education (ACGME)
   - Membership in the Council of Teaching Hospitals (COTH)
   - A ratio of full-time equivalent interns and residents to beds of .25 or higher.[3]

5. *Bed Size – small, medium, and large.* Bed size categories were based on the number of hospital beds and were specific to the hospital's region, location, and teaching status, as shown in Table 7 in Appendix I. We chose the bed size cutoff points 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). We used different cutoff points for rural, urban non-teaching, and urban teaching hospitals 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.

   We did not split rural hospitals according to teaching status, because rural teaching hospitals were rare. For example, in 2006, rural teaching hospitals comprised less than 2% of the total hospital universe. We defined the bed size categories 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.

   To further ensure geographic representativeness, implicit stratification variables included State and three-digit ZIP Code (the first three digits of the hospital’s five-digit ZIP Code). The hospitals were sorted according to these variables prior to systematic random sampling. Detailed information on the design of the NIS prior to 2006 is available in the

year-specific special reports on *Design of the Nationwide Inpatient Sample* found on the HCUP-US Website. Starting with the 2006 NIS, the design information was incorporated into this report.

Hospital Sampling Frame

The *universe* of hospitals was established as all community hospitals located in the U.S. with the exception, beginning in 1998, of short-term rehabilitation hospitals. However, some hospitals do not supply data to HCUP. Therefore, we constructed the NIS *sampling frame* from the subset of universe hospitals that released their discharge data to AHRQ for research use. The number of State Partners contributing data to the NIS has expanded over the years, as shown in Table 2 of Appendix I. As a result, the number of hospitals included in the NIS sampling frame has also increased over the years, as depicted in Figure 3, also in Appendix I.

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).

Figure 4 in Appendix I illustrates the number of hospitals in the universe, frame, and sample and the percentage of universe hospitals in the frame for each State in the sampling frame for 2006. In most cases, the difference between the universe and the frame represents the difference in the number of community, non-rehabilitation hospitals in the 2006 AHA Annual Survey Database (Health Forum, LLC © 2007) and the hospitals for which data 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 Figure 4 (Appendix I). Certain Texas State-licensed hospitals are exempt from statutory reporting requirements. Exempt hospitals include:

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

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

(I) Is located in a county that:

(A) Has a population estimated by the United States Bureau of the Census to be not more than 35,000 as of July 1 of the most recent year for which county population estimates have been published; or

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

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

These exemptions apply primarily to smaller rural public hospitals and, as a result, these

facilities are less likely to be included in the sampling frame than other Texas hospitals. While the number of hospitals omitted appears sizable, those available for the NIS include more than 93% of inpatient discharges from Texas universe hospitals because excluded hospitals tended to have relatively few discharges.

Refer to Table 8 of Appendix I for a full list of the number of hospitals, and discharges included in the 2006 AHA universe, frame, and NIS by State. Fewer hospitals may be in a State's frame than in the universe because data is not always received from every hospital and hospitals are sometimes excluded because of State requirements.

Hospital Sample Design

Design Considerations

The NIS is a stratified probability sample of hospitals in the frame, with sampling probabilities calculated to select 20% of the universe of U.S. community, non-rehabilitation hospitals contained in each stratum. This sample size was determined by AHRQ based on their experience with similar research databases. The overall design objective was to select a sample of hospitals that accurately represents the target universe, which includes hospitals outside the frame (i.e., having zero probability of selection). Moreover, this sample was to be geographically dispersed, yet drawn only from data supplied by HCUP Partners.

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 NIS. Ideally, relationships among outcomes and their correlates estimated from the NIS should accurately represent all U.S. hospitals. It is advisable to verify your estimates against other data sources, if available, because not all States contribute data to the NIS. Table 2 in Appendix I lists the number of NIS States, hospitals, and discharges by year. 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 *NIS Comparison Report* assesses the accuracy of NIS estimates by providing a comparison of the NIS with other data sources. The most recent report is available on the HCUP-US Website (http://www.hcup-us.ahrq.gov/db/nation/nis/nisrelatedreports.jsp).

The NIS team considered alternative stratified sampling allocation schemes. However, allocation proportional to the number of hospitals was preferred for several reasons:

- AHRQ researchers wanted a simple, easily understood sampling methodology. The concept that the NIS sample could represent a "miniaturization" of the hospital universe was appealing. There were, however, obvious geographic limitations imposed by data availability.

- AHRQ statisticians considered other optimal allocation schemes, including sampling hospitals with probabilities proportional to size (number of discharges). They ultimately concluded that sampling with probability proportional to the number of hospitals was preferable. While this approach was admittedly less efficient, the extremely large sample sizes yield reliable estimates. Furthermore, because the data are to be used for

purposes other than producing nationwide estimates, (e.g., regression modeling), it is critical that all hospital types, including small hospitals, are adequately represented.

Overview of the Sampling Procedure

To further ensure accurate geographic representation, we implicitly stratified the hospitals by State and three-digit ZIP Code (the first three digits of the hospital's five-digit ZIP Code). This was accomplished by sorting by three-digit ZIP Code within each stratum prior to drawing a systematic random sample of hospitals.

After stratifying the universe of hospitals, we sorted hospitals by stratum, the three-digit ZIP Code within each stratum, and by a random number within each three-digit ZIP Code. These sorts ensured further geographic generalizability of hospitals within the frame States, as well as random ordering of hospitals within three-digit ZIP Codes. Generally, three-digit ZIP Codes that are proximal in value are geographically near one another within a State. Furthermore, the U.S. Postal Service locates regional mail distribution centers at the three-digit level. Thus, the boundaries tend to be a compromise between geographic size and population size.

We then drew a systematic random sample of up to 20% of the total number of U.S. hospitals within each stratum. If too few frame hospitals appeared in a cell, we selected all frame hospitals for the NIS, subject to sampling restrictions specified by States. To simplify variance calculations, we drew at least two hospitals from each stratum. If fewer than two frame hospitals were available in a stratum, we merged it with an "adjacent" cell containing hospitals with similar characteristics.

Subsamples

Prior to the 2005 NIS, we drew two non-overlapping 10% subsamples of discharges from the NIS file for each year. The subsamples were selected by drawing every tenth discharge, starting with two different starting points (randomly selected between 1 and 10). Having a different starting point for each of the two subsamples guaranteed that they would not overlap. Discharges were sampled so that 10% of each hospital's discharges in each quarter were selected for each of the subsamples. The two samples could be combined to form a single, generalizable 20% subsample of discharges. Beginning with the 2005 NIS, 10% subsamples are no longer provided on the NIS CD-ROMs. However, users may still draw their own subsamples, if desired.

Change to Hospital Sampling Procedure Beginning with the 1998 NIS

Beginning with the 1998 NIS sampling procedures, all frame hospitals within a stratum have an equal probability of selection for the sample, regardless of whether they appeared in prior NIS samples. This deviates from the procedure used for earlier samples, which maximized the longitudinal component of the NIS series.

Further description of the sampling procedures for earlier releases of the NIS can be found in the special report: *Design of the HCUP Nationwide Inpatient Sample, 1997*. This report is available on the HCUP-US Website at http://www.hcup-us.ahrq.gov/db/nation/nis/nisrelatedreports.jsp. For a description of the development of the new sample design for 1998 and subsequent data years, see the special report: *Changes in NIS Sampling and Weighting Strategy for 1998.* This report is available on the HCUP-US Website.

Zero-Weight Hospitals

Beginning with the 1993 NIS, the NIS samples no longer contain zero-weight hospitals. For a description of zero-weight hospitals in the 1988-1992 samples, refer to the special report: *Design of the HCUP Nationwide Inpatient Sample, Release 1*. This report is available on the HCUP-US Website at http://www.hcup-us.ahrq.gov/db/nation/nis/nisrelatedreports.jsp.

Final Hospital Sample

In Appendix I, we present three figures describing the final hospital sample. Figure 5 depicts the numbers of hospitals sampled each year, while Figure 6 presents the numbers of discharges in each year of the NIS. For the 1988-1992 NIS, zero-weight hospitals were maintained to provide a longitudinal sample. Therefore, two figures exist for each of these years: one number for the regular NIS sample and another number for the total sample.

Figure 7 displays the weighted number of discharges sampled each year. Note that this number decreased from 35,408,207 in 1997 to 34,874,001 in 1998, a difference of 534,206 (1.5%). This slight decline is associated with two changes to the 1998 NIS design: the exclusion of community, rehabilitation hospitals from the hospital universe, and a change to the calculation of hospital universe discharges for the weights. Prior to 1998, we calculated discharges as the sum of total facility admissions (AHA data element ADMTOT), which includes long-term care admissions, plus births (AHA data element BIRTHS) reported for each U.S. community hospital in the AHA Annual Survey Database (Health Forum, LLC © 2007).

Beginning in 1998, we calculate discharges as the sum of hospital admissions (AHA data element ADMH) plus births for each U.S. community, non-rehabilitation hospital. This number is more consistent with the number of discharges we receive from the State data sources. We also substitute total facility admissions, if the number of hospital admissions is missing. Without these changes, the weighted number of discharges for 1998 would have been 35,622,743. The exclusion of community, rehabilitation hospitals reduced the number of universe hospitals by 177 and the number of weighted discharges by 214,490. The change in the calculation of discharges reduced the weighted number of discharges by 534,252.

Figure 8 presents a summary of the 2006 NIS hospital sample by geographic region and the number of:

- Universe hospitals (Universe)
- Frame hospitals (Frame)
- Sampled hospitals (Sample)
- Target hospitals (Target = 20% of the universe)
- Surplus hospitals (Surplus = Sample – Target).

Figure 9 summarizes the estimated U.S. population by geographic region. For each region, the figure reveals:

- The estimated U.S. population
- The estimated population of States in the 2006 NIS
- The percentage of estimated U.S. population included in NIS States.

Figure 10 depicts the number of discharges in the 2006 sample for each State.

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

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

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

SAMPLE WEIGHTS

To obtain nationwide estimates, we developed discharge weights using the AHA universe as the standard. These were developed separately for hospital- and discharge-level analyses. Hospital-level weights were developed to extrapolate NIS sample hospitals to the hospital universe. Similarly, discharge-level weights were developed to extrapolate NIS sample discharges to the discharge universe.

Hospital Weights

Hospital weights to the universe were calculated by post-stratification. For each year, hospitals were stratified on the same variables that were used for sampling: geographic region, urban/rural location, teaching status, bed size, and control. The strata that were collapsed for sampling were also collapsed for sample weight calculations. Within each stratum *s*, each NIS sample hospital's universe weight was calculated as:

$W_s(\text{universe}) = N_s(\text{universe}) \div N_s(\text{sample})$

where $W_s(\text{universe})$ was the hospital universe weight, and $N_s(\text{universe})$ and $N_s(\text{sample})$ were the number of community hospitals within stratum *s* in the universe and sample, respectively. Thus, each hospital's universe weight (HOSPWT) is equal to the number of universe hospitals it represents during that year. Because 20% of the hospitals in each stratum were sampled when possible, the hospital weights are usually near five.

Discharge Weights

The calculations for discharge-level sampling weights were similar to the calculations for hospital-level sampling weights. The discharge weights are usually constant for all discharges within a stratum. The only exceptions are for strata with sample hospitals that, according to the AHA files, were open for the entire year but contributed less than a full year of data to the NIS. For those hospitals, we *adjusted* the number of observed discharges by a factor of 4 ÷ Q, where Q was the number of calendar quarters for which 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. This adjustment was performed only for weighting purposes. The NIS data set includes only the actual (unadjusted) number of observed discharges.

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 number of total discharges was available for every hospital in the universe from the AHA files. Each universe hospital's AHA discharge total was calculated as the sum of newborns and hospital discharges.

Discharge weights to the universe were calculated by post-stratification. Hospitals were stratified just as they were for universe hospital weight calculations. Within stratum *s*, for hospital *i*, each NIS sample discharge's universe weight was calculated as:

$DW_{is}(\text{universe}) = [DN_s(\text{universe}) \div ADN_s(\text{sample})] * (4 \div Q_i)$

where DW_{is}(universe) was the discharge weight; DN_s(universe) represented the number of discharges from community hospitals in the universe within stratum *s*; ADN_s(sample) was the number of *adjusted* discharges from sample hospitals selected for the NIS; and Q_i represented the number of quarters of discharge data contributed by hospital *i* to the NIS (usually $Q_i = 4$). Thus, each discharge's weight (DISCWT) is equal to the number of universe discharges it represents in stratum *s* during that year. Because all discharges from 20% of the hospitals in each stratum were sampled when possible, the discharge weights are usually near five.

Appendix I: Tables and Figures

Table 1: 2006 Data Sources

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

State	Data Organization
NY	New York State Department of Health
OH	Ohio Hospital Association
OK	Oklahoma State Department of Health
OR	Oregon Association of Hospitals and Health Systems
RI	Rhode Island Department of Health
SC	South Carolina State Budget & Control Board
SD	South Dakota Association of Healthcare Organizations
TN	Tennessee Hospital Association
TX	Texas Department of State Health Services
UT	Utah Department of Health
VT	Vermont Association of Hospitals and Health Systems
VA	Virginia Health Information
WA	Washington State Department of Health
WI	Wisconsin Department of Health & Family Services
WV	West Virginia Health Care Authority

Return to Introduction

Table 2: Number of NIS States, Hospitals, and Discharges, by Year

Calendar Year	States in the Frame	Number of States	Sample Hospitals	Sample Discharges
1988	California, Colorado, Florida, Iowa, Illinois, Massachusetts, New Jersey, and Washington	8	758	5,265,756
1989	Added Arizona, Pennsylvania, and Wisconsin	11	875	6,110,064
1990	No new additions	11	861	6,268,515
1991	No new additions	11	847	6,156,188
1992	No new additions	11	838	6,195,744
1993	Added Connecticut, Kansas, Maryland, New York, Oregon, and South Carolina	17	913	6,538,976
1994	No new additions	17	904	6,385,011
1995	Added Missouri and Tennessee	19	938	6,714,935
1996	No new additions	19	906	6,542,069
1997	Added Georgia, Hawaii, and Utah	22	1012	7,148,420
1998	No new additions	22	984	6,827,350
1999	Added Maine and Virginia	24	984	7,198,929
2000	Added Kentucky, North Carolina, Texas, and West Virginia	28	994	7,450,992
2001	Added Michigan, Minnesota, Nebraska, Rhode Island, and Vermont	33	986	7,452,727
2002	Added Nevada, Ohio, and South Dakota; Dropped Arizona	35	995	7,853,982
2003	Added Arizona, Indiana, and New Hampshire; Dropped Maine	37	994	7,977,728
2004	Added Arkansas; Dropped Pennsylvania	37	1,004	8,004,571
2005	Added Oklahoma; Dropped Virginia	37	1,054	7,995,048
2006	Added Virginia	38	1,045	8,074,825

Return to Introduction

Table 3. Summary of NIS Releases

Data from	Media/format options	Structure of Releases
1988-1992 • 8 States in 1988 • 11 States in 1989-1992	On CD-ROM, In ASCII format	5 years of data in a 6-CD set, compressed files Two 10% subsamples of discharges for each year
1993 • 17 states 1994 • 17 states 1995 • 19 states 1996 • 19 states 1997 • 22 states 1998 • 22 states 1999 • 24 states 2000 • 28 states 2001 • 33 states	On CD-ROM, In ASCII format	1 year of data in a 2-CD set, compressed files Two 10% subsamples of discharges for each year
2002 • 35 states 2003 • 37 states 2004 • 37 states	On CD-ROM, In ASCII format	1 year of data in a 2-CD set, compressed files Two 10% subsamples of discharges for each year A companion file with four different sets of severity measures
2005 • 37 states 2006 • 38 states	On CD-ROM, In ASCII format	1 year of data in a 2-CD set, compressed files A companion file with four different sets of severity measures, and also diagnosis and procedure groups

Table 4. Summary of NIS Data Sources, Hospitals, and Inpatient Stays, 1988-2006

Year	Data sources	Number of hospitals	Number of discharges in the NIS, unweighted	Number of discharges in the NIS, weighted for national estimates
1988	CA CO FL IL IA MA NJ WA	759	5,265,756	35,171,448
1989	AZ CA CO FL IL IA MA NJ PA WA WI *(Added AZ, PA, WI)*	882	6,110,064	35,104,645
1990	AZ CA CO FL IL IA MA NJ PA WA WI *(No change)*	871	6,268,515	35,215,397
1991	AZ CA CO FL IL IA MA NJ PA WA WI *(No change)*	859	6,156,188	35,036,492
1992	AZ CA CO FL IL IA MA NJ PA WA WI *(No change)*	856	6,195,744	35,011,385
1993	AZ CA CO CT FL IL IA KS MD MA NJ NY OR PA SC WA WI *(Added CT, KS, MD, NY, OR, SC)*	913	6,538,976	34,714,530
1994	AZ CA CO CT FL IL IA KS MD MA NJ NY OR PA SC WA WI *(No change)*	904	6,385,011	34,622,203
1995	AZ CA CO CT FL IL IA KS MD MA MO NJ NY OR PA SC TN WA WI *(Added MO, TN)*	938	6,714,935	34,791,998
1996	AZ CA CO CT FL IL IA KS MD MA MO NJ NY OR PA SC TN WA WI *(No change)*	906	6,542,069	34,874,386
1997	AZ CA CO CT FL GA HI IL IA KS MD MA MO NJ NY OR PA SC TN UT WA WI *(Added GA, HI, UT)*	1,012	7,148,420	35,408,207
1998	AZ CA CO CT FL GA HI IL IA KS MD MA MO NJ NY OR PA SC TN UT WA WI *(No change)*	984	6,827,350	34,874,001
1999	AZ CA CO CT FL GA HI IL IA KS MD MA ME MO NJ NY OR PA SC TN UT VA WA WI *(Added ME, VA)*	984	7,198,929	35,467,673
2000	AZ CA CO CT FL GA HI IL IA KS KY MD MA ME MO NC NJ NY OR PA SC TN TX UT VA WA WI WV *(Added KY, NC, TX, WV)*	994	7,450,992	36,417,565

Year	Data sources	Number of hospitals	Number of discharges in the NIS, unweighted	Number of discharges in the NIS, weighted for national estimates
2001	AZ CA CO CT FL GA HI IL IA KS KY MD MA ME MI MN MO NC NE NJ NY OR PA RI SC TN TX UT VA VT WA WI WV *(Added MI, MN, NE, RI, VT)*	986	7,452,727	37,187,641
2002	CA CO CT FL GA HI IL IA KS KY MD MA ME MI MN MO NC NE NJ NY NV OH OR PA RI SC SD TN TX UT VA VT WA WI WV *(Added NV, OH, SD; AZ data were not available)*	995	7,853,982	37,804,021
2003	AZ CA CO CT FL GA HI IL IN IA KS KY MD MA MI MN MO NC NE NH NJ NY NV OH OR PA RI SC SD TN TX UT VA VT WA WI WV *(Added AZ, IN, NH; ME data were not available)*	994	7,977,728	38,220,659
2004	AR AZ CA CO CT FL GA HI IL IN IA KS KY MD MA MI MN MO NC NE NH NJ NY NV OH OR RI SC SD TN TX UT VA VT WA WI WV *(Added AR; PA data were not available)*	1,004	8,004,571	38,661,786
2005	AR AZ CA CO CT FL GA HI IL IN IA KS KY MD MA MI MN MO NC NE NH NJ NY NV OH OK OR RI SC SD TN TX UT VT WA WI WV *(Added OK; VA data were not available)*	1,054	7,995,048	39,163,834
2006	AR AZ CA CO CT FL GA HI IL IN IA KS KY MD MA MI MN MO NC NE NH NJ NY NV OH OK OR RI SC SD TN TX UT VA VT WA WI WV *(Added VA)*	1,045	8,074,825	39,450,216

Return to Introduction

Table 5. NIS Related Reports and Database Documentation Available on HCUP-US

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

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

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

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

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

Load Programs

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

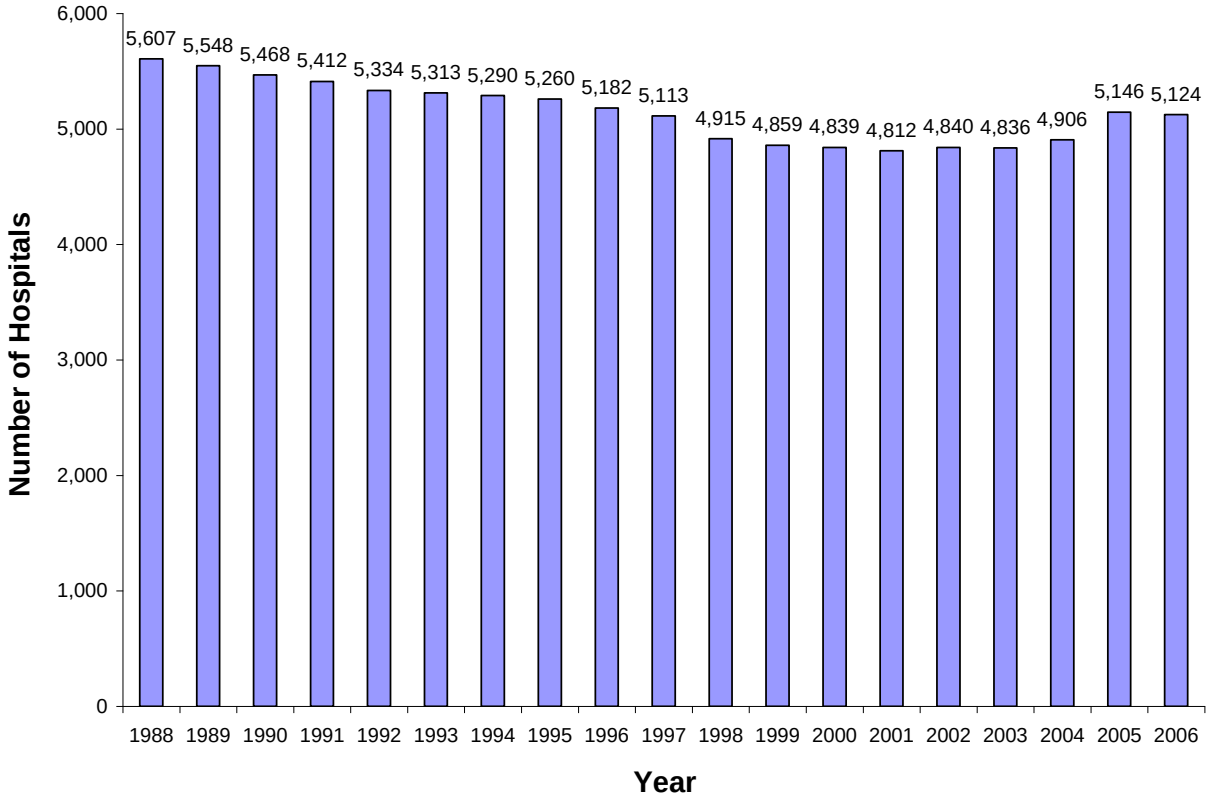
HCUP Tools: Labels and Formats
- Overview of Clinical Classifications Software (CCS), a categorization scheme that groups ICD-9-CM diagnosis and procedure codes into mutually exclusive categories
- Labels file for CCS categories
- Labels file for multiple versions of Diagnosis Related Groups (DRGs) and Major Diagnostic Categories (MDCs)
- NIS SAS format library program to create a value labels
- NIS ICD-9-CM formats to label ICD-9-CM diagnoses and procedures
- NIS Severity formats to label severity data elements

NIS Related Reports

Links to HCUP-US page with various NIS related reports such as the following:
- Design of the Nationwide Inpatient Sample for 1988 to 2005 (*included in this document beginning in 2006*)
- Changes in NIS Sampling and Weighting Strategy for 1998
- Calculating Nationwide Inpatient Sample Variances
- Using the HCUP Nationwide Inpatient Sample to Estimate Trends
- NIS Comparison Reports (available for years in which the NIS sample changed)
- HCUP Data Quality Reports for 1988-2006
- HCUP E-Code Evaluation Report

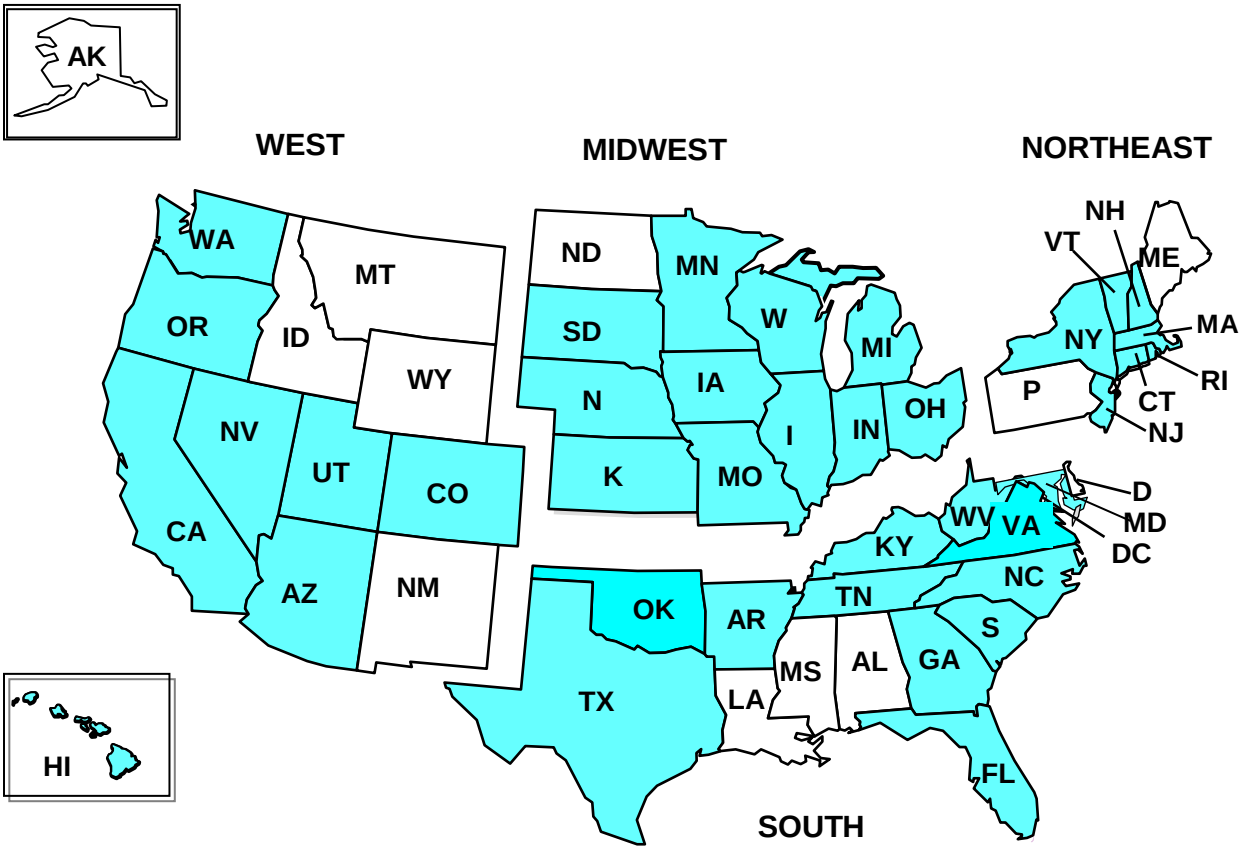
HCUP Supplemental Files
- Cost-to-Charge Ratio files
- Hospital Market Structure (HMS) files
- NIS Trends Supplemental files

Figure 1: Hospital Universe, by Year[4]



Return to Introduction

Figure 2: NIS States, by Region



Return to Introduction

Table 6: All States, by Region

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

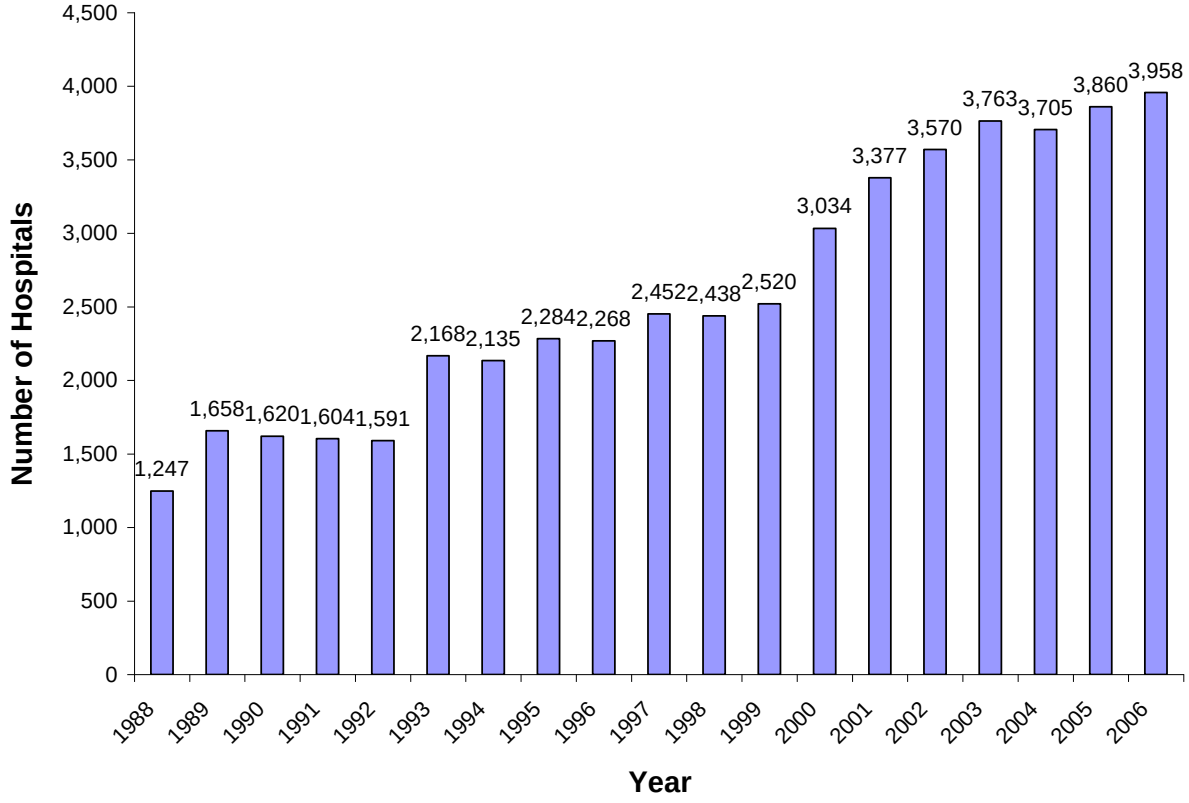
Return to Introduction

Table 7: Bed Size Categories, by Region

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

Return to Introduction

Figure 3: NIS Hospital Sampling Frame, by Year



Return to Introduction

Figure 4: Number of Hospitals in the 2006 Universe, Frame, and Sample for Frame States

Part A: Arkansas – North Carolina

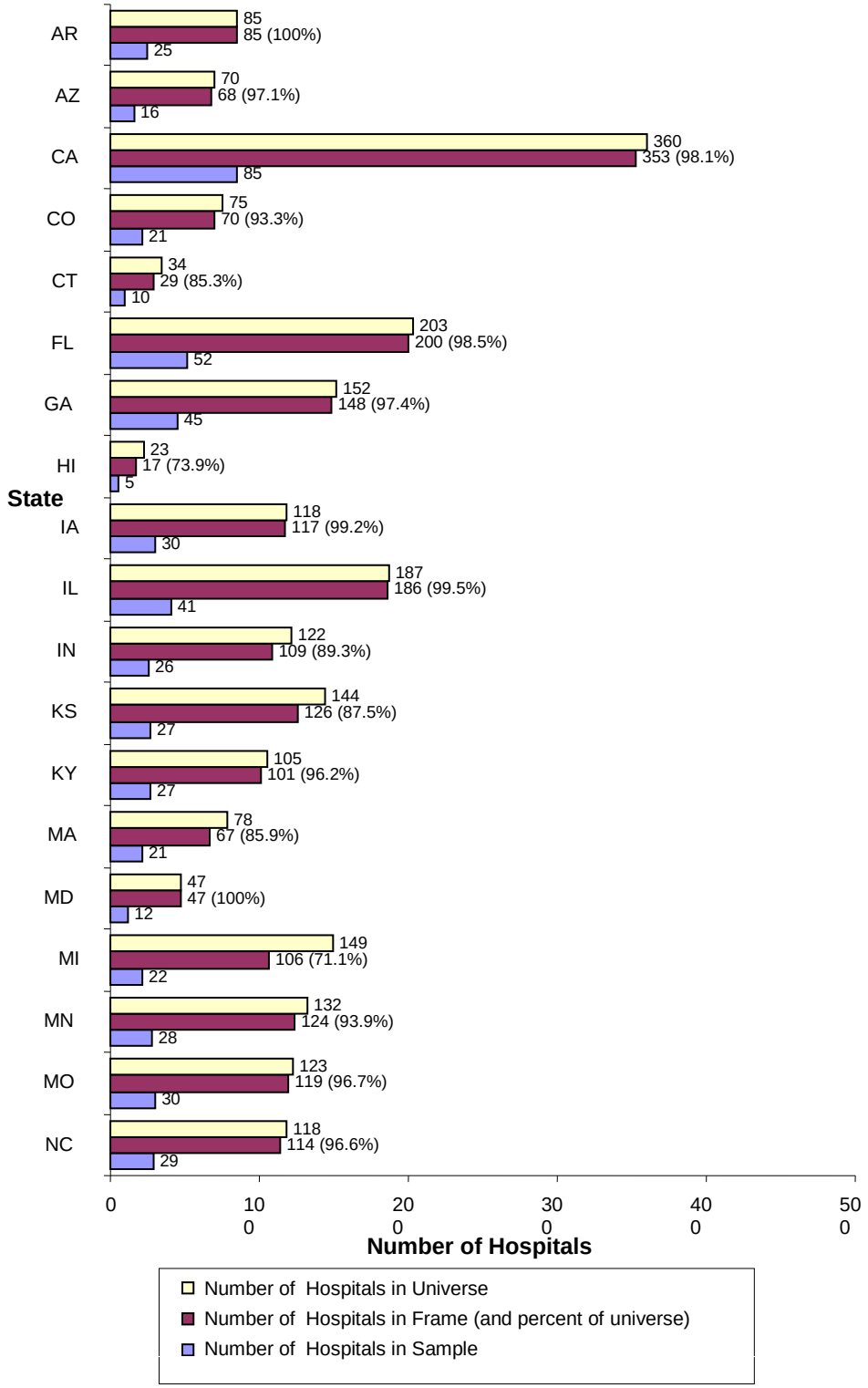


Figure 4: Number of Hospitals in the 2006 Universe, Frame, and Sample for Frame States

Part B: Nebraska – West Virginia

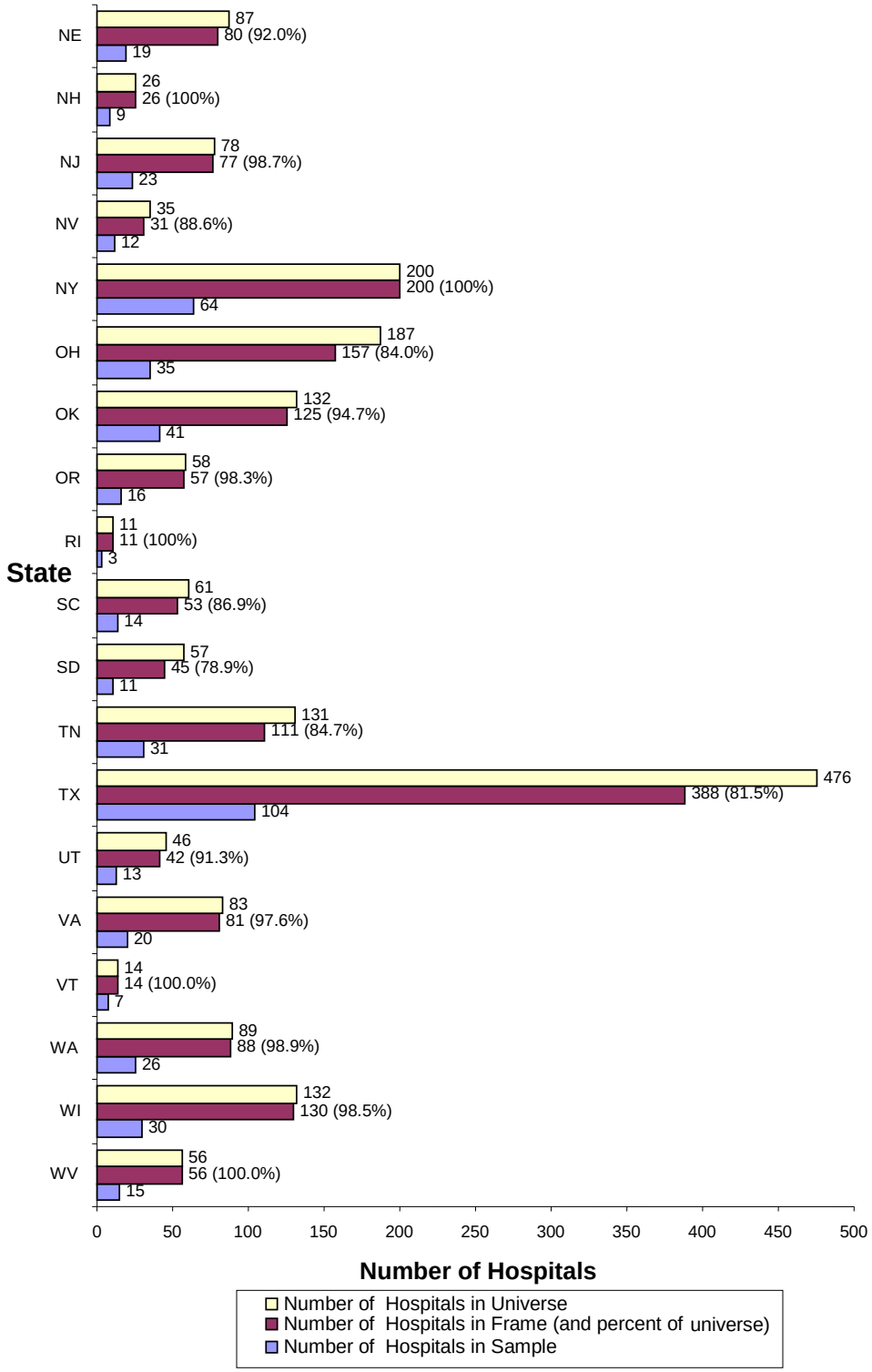
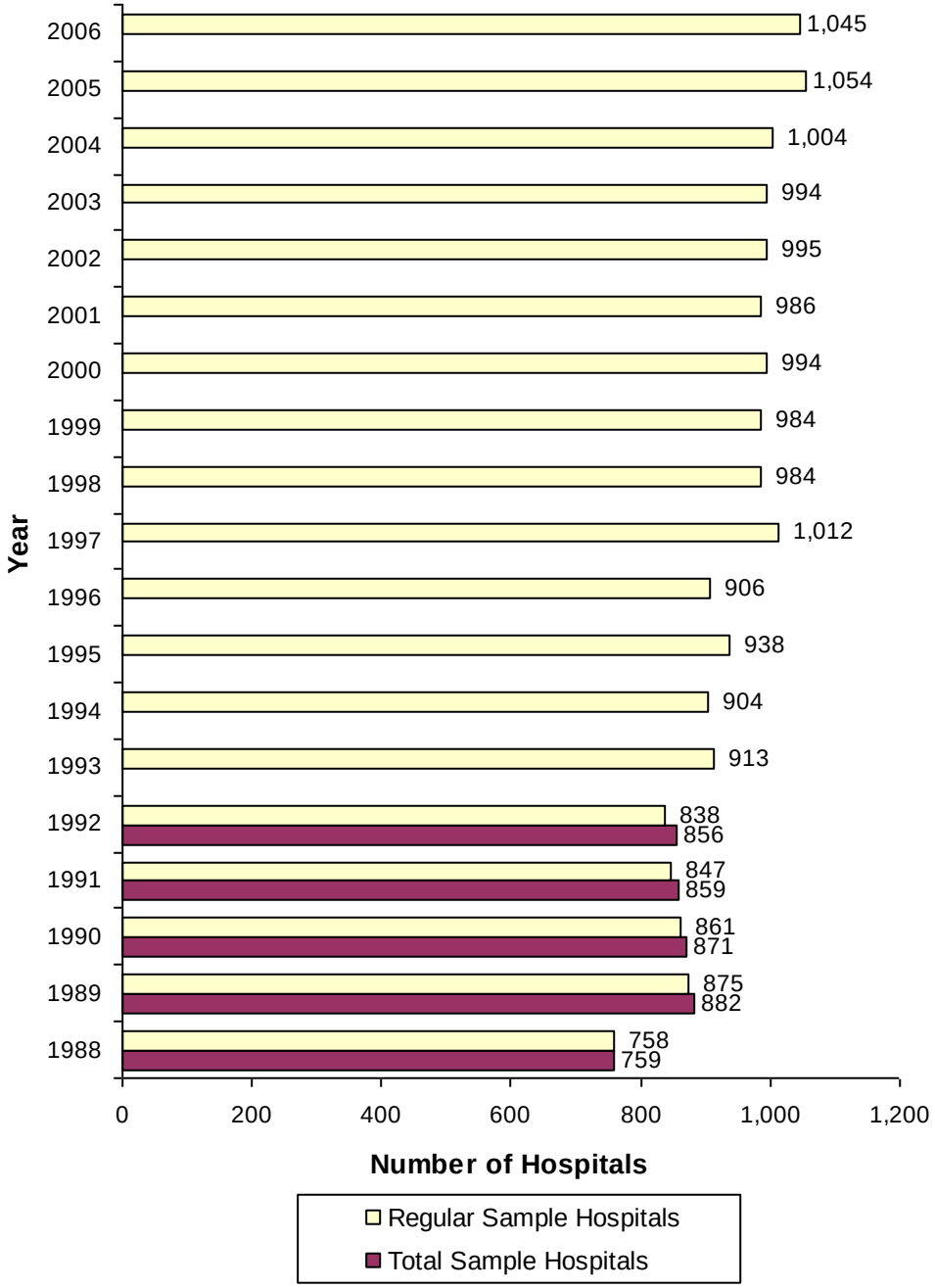


Table 8: Number of Hospitals and Discharges in 2006 AHA Universe, Frame, and NIS, by State

State	Number of Hospitals and Discharges in 2006 AHA Universe, Frame, and NIS, by State					
	AHA		Frame		NIS	
	Hospitals	Discharges	Hospitals	Discharges	Hospitals	Discharges
Non-Frame States	840	4,993,113	0	0	0	0
Arkansas	85	409,270	85	408,458	25	99,813
Arizona	70	764,945	68	749,955	16	172,669
California	360	3,950,635	353	3,840,964	85	856,488
Colorado	75	487,984	70	473,850	21	107,801
Connecticut	34	446,658	29	425,240	10	154,047
Florida	203	2,597,761	200	2,519,588	52	673,665
Georgia	152	1,102,671	148	1,065,886	45	285,592
Hawaii	23	119,748	17	95,319	5	18,503
Iowa	118	394,288	117	366,662	30	101,321
Illinois	187	1,732,464	186	1,679,905	41	348,310
Indiana	122	814,316	109	773,198	26	176,487
Kansas	144	384,182	126	344,456	27	61,153
Kentucky	105	657,228	101	633,534	27	160,788
Massachusetts	78	917,083	67	632,160	21	168,259
Maryland	47	744,032	47	747,184	12	172,690
Michigan	149	1,332,242	106	891,026	22	154,827
Minnesota	132	701,463	124	597,348	28	156,746
Missouri	123	901,593	119	900,522	30	229,201

State	Number of Hospitals and Discharges in 2006 AHA Universe, Frame, and NIS, by State					
	AHA		Frame		NIS	
	Hospitals	Discharges	Hospitals	Discharges	Hospitals	Discharges
North Carolina	118	1,124,678	114	1,106,005	29	296,444
Nebraska	87	241,987	80	163,932	19	39,191
New Hampshire	26	129,542	26	126,364	9	43,189
New Jersey	78	1,199,416	77	1,123,919	23	285,237
Nevada	35	291,116	31	284,742	12	89,061
New York	200	2,799,866	200	2,630,180	64	776,115
Ohio	187	1,688,462	157	1,587,287	35	356,498
Oklahoma	132	526,879	125	485,082	41	191,179
Oregon	58	388,611	57	377,019	16	100,552
Rhode Island	11	139,497	11	139,533	3	25,446
South Carolina	61	563,964	53	479,446	14	145,123
South Dakota	57	107,926	45	82,556	11	11,459
Tennessee	131	990,367	111	857,561	31	260,565
Texas	476	2,994,670	388	2,791,722	104	648,712
Utah	46	274,838	42	262,636	13	79,658
Virginia	83	870,856	81	857,990	20	234,573
Vermont	14	56,044	14	54,335	7	34,898
Washington	89	632,823	88	625,373	26	161,846
Wisconsin	132	681,511	130	650,321	30	144,123
West Virginia	56	295,487	56	283,651	15	52,596
Total	5,124	39,450,216	3,958	32,114,909	1,045	8,074,825

Figure 5: Number of Hospitals Sampled, by Year



[Return to Introduction]

Figure 6: Number of NIS Discharges, Unweighted, by Year

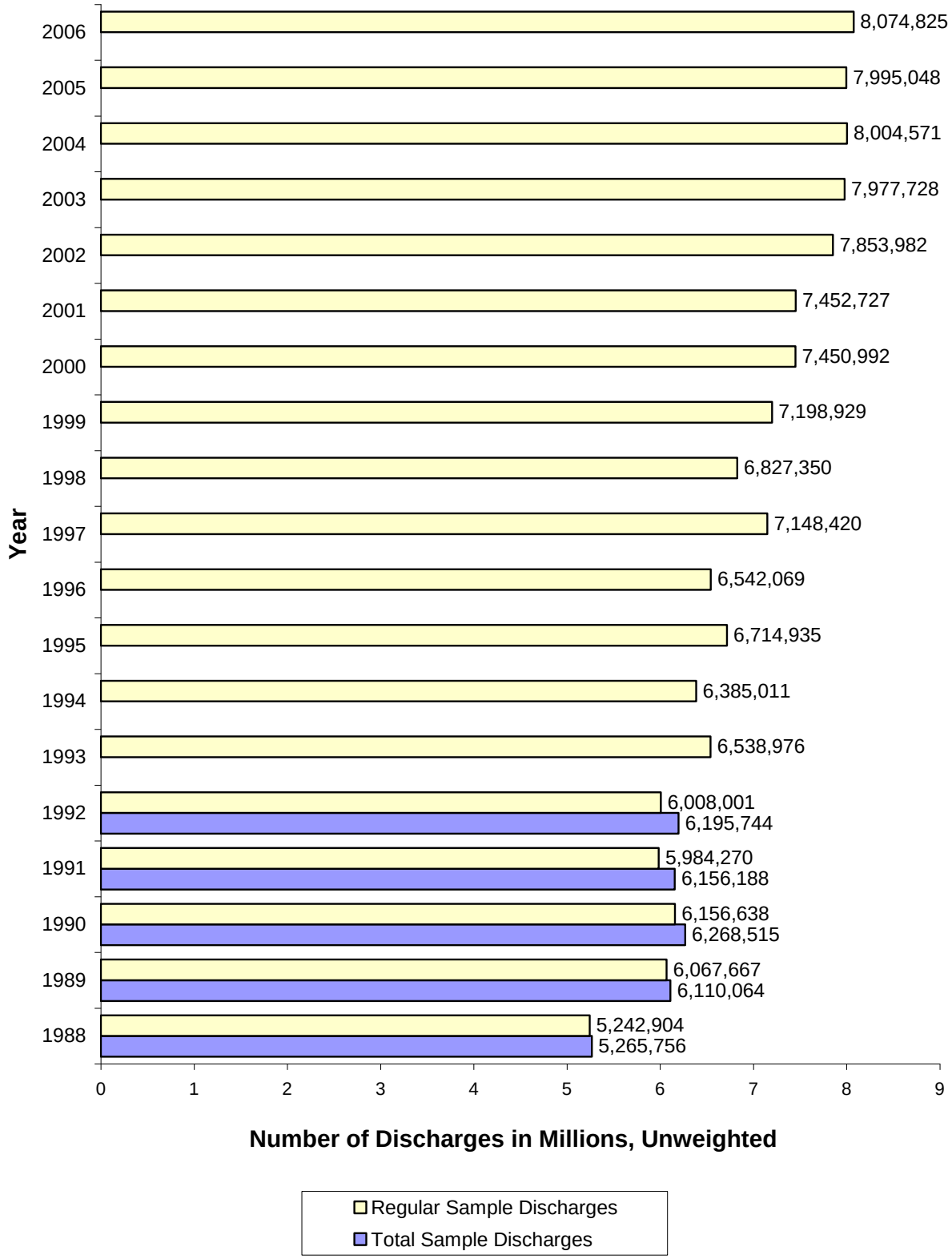


Figure 7: Number of NIS Discharges, Weighted, by Year

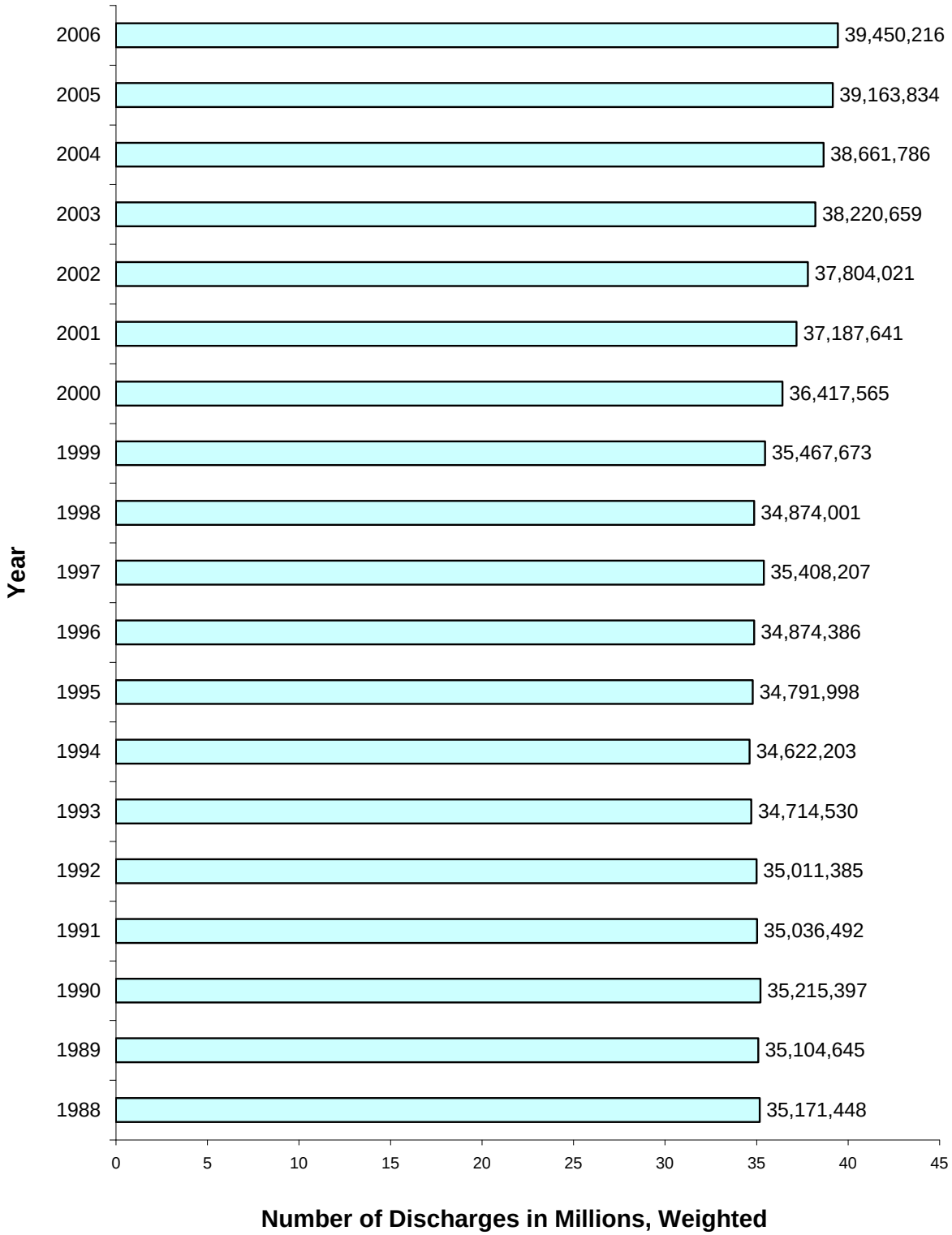


Figure 8: Number of Hospitals in the 2006 Universe, Frame, Sample, Target, and Surplus, by Region

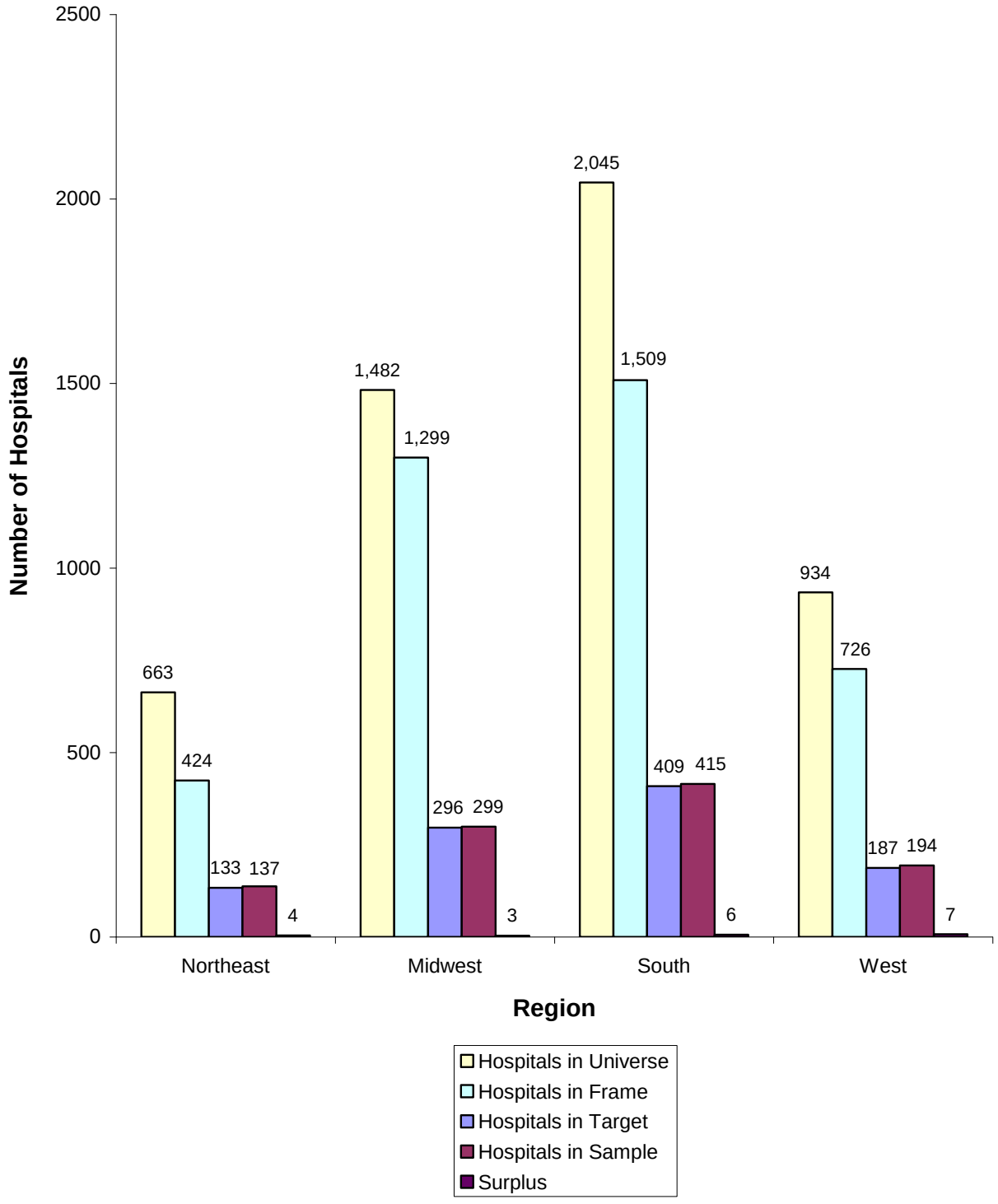
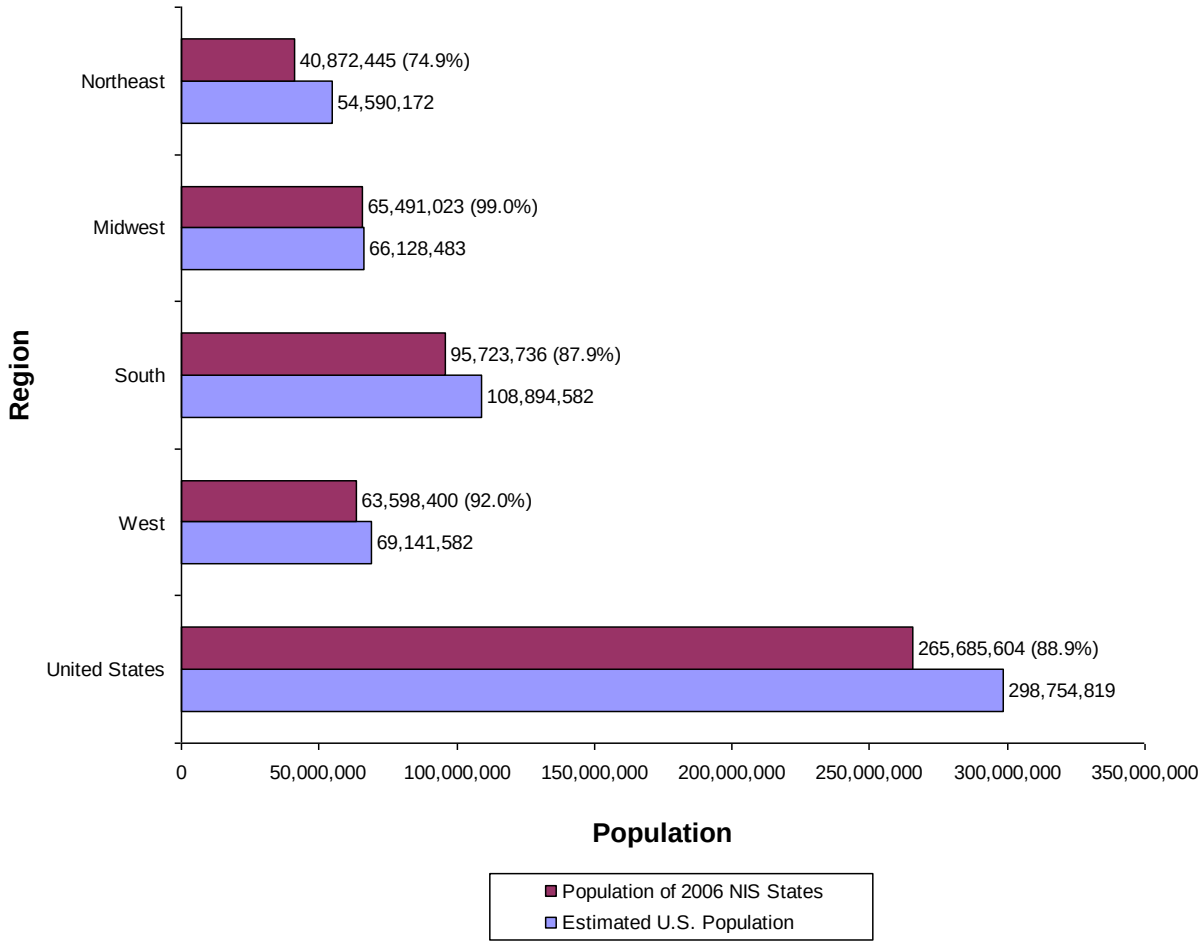


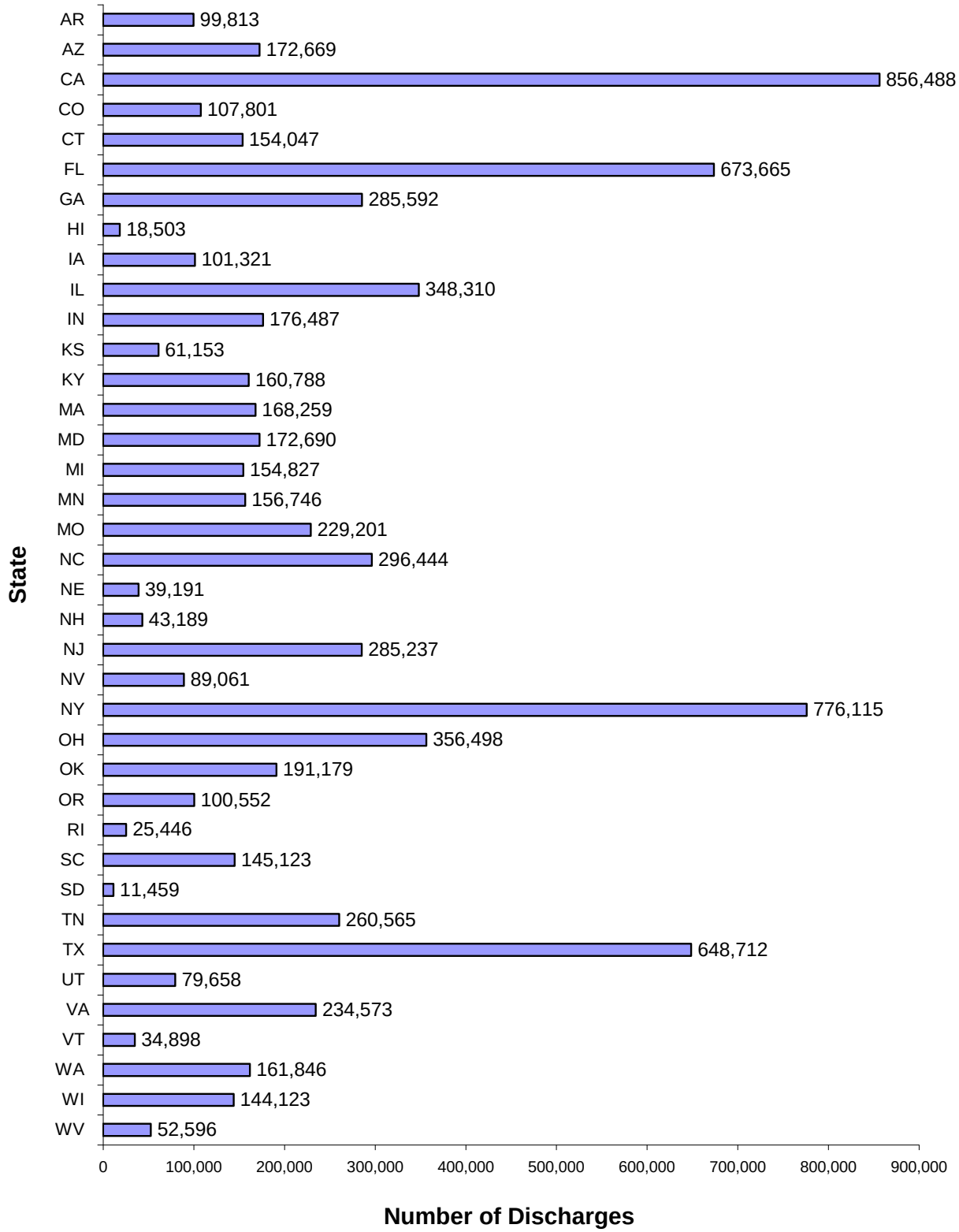
Figure 9: Percentage of U.S. Population in 2006 NIS States, by Region

Calculated using the estimated U.S. population on July 1, 2006.[5]



Return to Introduction

Figure 10: Number of Discharges in the 2006 NIS, by State



Appendix II: State-Specific Restrictions

The table below enumerates the types of restrictions applied to the Nationwide Inpatient Sample. Restrictions include the following types:

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

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

Confidentiality of Hospitals - Restricted Identification of Hospitals

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

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

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

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

*Available in AR.

Confidentiality of Hospitals - Limitation on Sampling

Limitations on sampling were needed for the following data sources:

- CT: CHIME, Inc.
  - CHIME is to be notified if more than 50% of their hospitals appear in any year of the NIS.
  - 2006 NIS – 34% of the hospitals were sampled.

- IL: Illinois Department of Public Health
  - Illinois requested that no more than 40% of Illinois discharges appear in any discharge quarter of NIS data.
  - 2006 NIS – 20% of the discharges in Illinois were sampled. No hospitals were dropped from the sampling frame.

- MI: Michigan Health & Hospital Association
  - Reporting of total charge is limited in the Michigan data. Thirty-three of 140 hospitals were dropped from the sampling frame because they did not report any total charges. These hospitals were fairly evenly distributed by hospital type. There were no sampling strata in the State containing only hospitals without total charges.

- NE: Nebraska Hospital Association
  - Two out of 85 hospitals were dropped from the sampling frame because they had incomplete data and were missing a high percentage of Medicare discharges.

- OH: Ohio Hospital Association
  - Three hospitals were dropped from the sampling frame to meet additional Ohio confidentiality requirements.

- VA: Virginia Health Information
  - VHI is to be notified if more than 50% of their hospitals appear in any year of the NIS.
  - 2006 NIS – 25% of the hospitals were sampled.

Some States limit the hospitals that can be included in the NIS. The following data sources requested that hospitals be dropped from the sampling frame whenever there were fewer than two hospitals in a sampling stratum. For more details about the number of hospitals included in the AHA Universe, Frame, and NIS for each NIS State, refer to Table 8 in Appendix I.

- GA: GHA: An Association of Hospitals & Health Systems

- HI: Hawaii Health Information Corporation

- IN: Indiana Hospital & Health Association

- MI: Michigan Health & Hospital Association

- NE: Nebraska Hospital Association

- OH: Ohio Hospital Association
- OK: Oklahoma State Department of Health
- SC: South Carolina State Budget & Control Board
- SD: South Dakota Association of Healthcare Organizations
- TN: Tennessee Hospital Association

Confidentiality of Hospitals - Restricted Release of Stratifiers

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

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

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

- HOSP_CONTROL, control/ownership of hospital
- HOSP_LOCATION, location (urban/rural) of hospital
- HOSP_TEACH, teaching status of hospital
- HOSP_BEDSIZE, bed size of hospital
- HOSP_LOCTEACH, location/teaching status of hospital

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

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

- FL: Florida Agency for Health Care Administration
  - Age in days (AGEDAY) is set to missing on all records
- MA: Division of Health Care Finance and Policy
  - Age in days (AGEDAY) is set to missing on all records
- NH: New Hampshire Department of Health & Human Services
  - Age in days (AGEDAY) is set to missing on all records
- SC: South Carolina State Budget & Control Board
  - Age in days (AGEDAY) is set to missing on all records
- TX: Texas Department of State Health Services

- Age in days (AGEDAY) is set to missing on all records
- Age in years (AGE) is set to the midpoints of age ranges defined by the data source. There were 22 age groups for the general patient population and 5 age groups for the HIV or alcohol/drug use patients. The age groups are shown below:

- Texas Restriction on AGE for General Patient Population other than HIV or Drug/Alcohol Use Patients

Age Range	New value of AGE
0	0
1-4	2
5-9	7
10-14	12
15-17	16
18-19	19
20-24	22
25-29	27
30-34	32
35-39	37
40-44	42
45-49	47
50-54	52
55-59	57
60-64	62
65-69	67
70-74	72
75-79	77
80-84	82
85-89	87
90 and above	90

Texas Restriction on AGE for HIV or Drug/Alcohol Use Patients

Age Range	New value of AGE
0	0
1-17	8
18-44	31
45-64	54
65-74	69
75 and above	75

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

Confidentiality of Records – Other Restrictions

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

- CA: Office of Statewide Health Planning & Development
  - Admission month (AMONTH), age in days, (AGEDAY), age in years (AGE), gender (FEMALE), and race (RACE), are suppressed for some records. In some cases, AGE is set to the midpoint of the age category.

- FL: Florida Agency for Health Care Administration
  - Admission month (AMONTH) is set to missing on all records

- GA: GHA: An Association of Hospitals & Health Systems
  - Patient race (RACE) is set to missing on all records

Confidentiality of Physicians

The following data sources restrict the release of physician identifiers:

- CT: CHIME, Inc.
- GA: GHA: An Association of Hospitals & Health Systems
- IL: Illinois Department of Public Health
- MA: Division of Health Care Finance and Policy
- NC: North Carolina Department of Health and Human Services
- UT: Utah Department of Health
- VT: Vermont Association of Hospitals and Health Systems
- WV: West Virginia Health Care Authority

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

- MDNUM1_R/MDNUM2_R (beginning in 2003)
- MDNUM1_S/MDNUM2_S (2001 to 2002)
- MDID_S/SURGID_S (prior to 2001)

Missing Discharges

The following data sources may be missing discharge records for specific populations of patients:

- IA: Iowa Hospital Association
  - Beginning in data year 2001, the Iowa Hospital Association prohibits the release of two types of discharges: HIV infections (defined by MDC of 25) and behavioral health including chemical dependency care or psychiatric care (defined by a service code of BHV). These discharges were not included in the source file provided to HCUP and were therefore not included in the NIS.

- NE: Nebraska Hospital Association
  - The Nebraska Hospital Association prohibits the release of discharge records for patients with HIV diagnoses. These discharges were not included in the source file provided to HCUP and were therefore not included in the NIS.

Appendix III: Data Elements

Table 1. Data Elements in the NIS Inpatient Core Files

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

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
Admission day of week or weekend	AWEEKEND	1998-2006	Admission on weekend: (0) admission on Monday-Friday, (1) admission on Saturday-Sunday	
	ADAYWK	*1988-1997*	*Admission day of week: (1) Sunday, (2) Monday, (3) Tuesday, (4) Wednesday, etc.*	
Admission month	AMONTH	1988-2006	Admission month coded from (1) January to (12) December	FL
Admission source	ASOURCE	1988-2006	Admission source, uniform coding: (1) ER, (2) another hospital, (3) another facility including long-term care, (4) court/law enforcement, (5) routine/birth/other	
	ASOURCE_X	1998-2006	Admission source, as received from data source using State-specific coding	
	ASOURCEUB92	2003-2006	Admission source (UB-92 standard coding). For newborn admissions (ATYPE = 4): (1) normal delivery, (2) premature delivery, (3) sick baby, (4) extramural birth; For non-newborn admissions (ATYPE NE 4): (1) physician referral, (2) clinic referral, (3) HMO referral, (4) transfer from a hospital, (5) transfer from a skilled nursing facility, (6) transfer from a another health care facility, (7) emergency room, (8) court/law enforcement, (A) transfer from a critical access hospital	CA, MD, RI
Admission type	ATYPE	1988-2006	Admission type, uniform coding: (1) emergency, (2) urgent, (3) elective, (4) newborn, (5) trauma center beginning in 2003 data, (6) other	CA
	ELECTIVE	2002-2006	Indicates elective admission: (1) elective, (0) non-elective admission	
Age at admission	AGE	1988-2006	Age in years coded 0-124 years	
	AGEDAY	1988-2006	Age in days coded 0-365 only when the age in years is less than 1	FL, MA, NH, SC, TX
Clinical Classifications	DXCCS1 - DXCCS15	1998-2006	CCS category for all diagnoses for NIS beginning in 1998	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
Software (CCS) category	*DCCHPR1*	*1988-1997*	*CCS category for principal diagnosis for NIS prior to 1998. CCS was formerly called the Clinical Classifications for Health Policy Research (CCHPR).*	
	PRCCS1 - PRCCS15	1998-2006	CCS category for all procedures for NIS beginning in 1998	
	PCCHPR1	*1988-1997*	*CCS category for principal procedure for NIS prior to 1998. CCS was formerly called the Clinical Classifications for Health Policy Research (CCHPR).*	
Data source information	*DSNUM*	*1988-1997*	*Data source number*	
	DSTYPE	*1988-1997*	*Data source type: (1) State data organization, (2) Hospital association, (3) Consortia*	
Diagnosis information	DX1 - DX15	1988-2006	Diagnoses, principal and secondary (ICD-9-CM). Beginning in 2003, the diagnosis array does not include any external cause of injury codes. These codes have been stored in a separate array ECODEn.	
	NDX	1988-2006	Number of diagnoses coded on the original record	
	DSNDX	*1988-1997*	*Number of diagnosis fields provided by the data source*	
	DXSYS	*1988-1997*	*Diagnosis system (ICD-9-CM)*	
	DXV1 - DXV15	*1988-1997*	*Diagnosis validity flags*	
Diagnosis Related Group (DRG)	DRG	1988-2006	DRG in use on discharge date	
	DRGVER	1988-2006	Grouper version in use on discharge date	
	DRG10	*1988-1999*	*DRG Version 10 (effective October 1992 - September 1993)*	
	DRG18	*1998-2005*	*DRG Version 18 (effective October 2000 - September 2001)*	
	DRG24	2006	DRG Version 24 (effective October 2006 - September 2007)	
Discharge quarter	DQTR	1988-2006	Coded: (1) Jan - Mar, (2) Apr - Jun, (3) Jul - Sep, (4) Oct - Dec	
	DQTR_X	2006	Discharge quarter, as received from data source	
Discharge weights (Weights for 1988-1993 are on Hospital Weights file)	DISCWT	1998-2006	Discharge weight on Core file and Hospital Weights file for NIS beginning in 1998. In all data years except 2000, this weight is used to create national estimates for all analyses. In 2000 only, this weight is used to create national estimates for all analyses, excluding those that involve total charges.	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	DISCWT_U	*1993-1997*	*Discharge weight on Core file and Hospital Weights file for NIS prior to 1998*	
	DISCWTcharge	*2000*	*Discharge weight for national estimates of total charges. In 2000 only, this weight is used to create national estimates for analyses that involve total charges.*	
	DISCWT10	*1998-2004*	*Discharge weight on 10% subsample Core file for NIS from 1998 to 2004. In all data years except 2000, this weight is used to create national estimates for all analyses. In 2000 only, this weight is used to create national estimates for all analyses, excluding those that involve total charges.*	
	D10CWT_U	*1993-1997*	*Discharge weight on 10% subsample Core file for NIS prior to 1998*	
	DISCWTcharge10	*2000*	*Discharge weight for national estimates of total charges on 10% subsample file. In 2000 only, this weight is used to create national estimates for analyses that involve total charges.*	
Discharge year	YEAR	1988-2006		
Disposition of patient (discharge status)	*DISP*	*1988-1997*	*Disposition of patient, uniform coding used prior to 1998: (1) routine, (2) short-term hospital, (3) skilled nursing facility, (4) intermediate care facility, (5) another type of facility, (6) home health care, (7) against medical advice, (20) died*	
	DIED	1988-2006	Indicates in-hospital death: (0) did not die during hospitalization, (1) died during hospitalization	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	DISPUB92	1998-2006	Disposition of patient, UB-92 coding: (1) routine, (2) short-term hospital, (3) skilled nursing facility, (4) intermediate care, (5) another type of facility, (6) home health care, (7) against medical advice, (8) home IV provider, (20) died in hospital, (40) died at home, (41) died in a medical facility, (42) died, place unknown, (43) alive, Federal health facility, (50) Hospice, home, (51) Hospice, medical facility, (61) hospital-based Medicare approved swing bed , (62) another rehabilitation facility, (63) long-term care hospital, (64) certified nursing facility, (65) psychiatric hospital, (66) critical access hospital (71) another institution for outpatient services, (72) this institution for outpatient services, (99) discharged alive, destination unknown	CA, IN, MD
	DISPUNIFORM	1998-2006	Disposition of patient, uniform coding used beginning in 1998: (1) routine, (2) transfer to short-term hospital, (5) other transfers, including skilled nursing facility, intermediate care, and another type of facility, (6) home health care, (7) against medical advice, (20) died in hospital, (99) discharged alive, destination unknown	
External causes of injury and poisoning	ECODE1 - ECODE4	2003-2006	External cause of injury and poisoning code, primary and secondary (ICD-9-CM). Beginning in 2003, external cause of injury codes are stored in a separate array ECODEn from the diagnosis codes in the array DXn. Prior to 2003, these codes are contained in the diagnosis array (DXn).	
	E_CCS1 - E_CCS4	2003-2006	CCS category for the external cause of injury and poisoning codes	
	NECODE	2003-2006	Number of external cause of injury codes on the original record. A maximum of 4 codes are retained on the NIS.	
Gender of patient	FEMALE	1998-2006	Indicates gender for NIS beginning in 1998: (0) male, (1) female	
	SEX	*1988-1997*	*Indicates gender for NIS prior to 1998: (1) male, (2) female*	
Hospital information	DSHOSPID	1988-2006	Hospital number as received from the data source	GA, HI, IN, KS, MI, NE, OH, OK, SC, SD, TN, TX

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	HOSPID	1988-2006	HCUP hospital number (links to Hospital Weights file)	
	HOSPST	1988-2006	State postal code for the hospital (e.g., AZ for Arizona)	
	HOSPSTCO	*1988-2002*	*Modified Federal Information Processing Standards (FIPS) State/county code for the hospital links to Area Resource File (available from the Bureau of Health Professions, Health Resources and Services Administration). Beginning in 2003, this data element is available only on the Hospital Weights file.*	
	NIS_STRATUM	2000-2006	Stratum used to sample hospitals, based on geographic region, control, location/teaching status, and bed size. Stratum information is also contained in the Hospital Weights file.	
Indicates in-hospital birth	HOSPBRTH	2006	Indicator that discharge record includes diagnosis of birth that occurred in the hospital	
Length of stay	LOS	1988-2006	Length of stay, edited	
	LOS_X	1988-2006	Length of stay, as received from data source	
Location of the patient	PL_UR_CAT4	2003-2006	Urban–rural designation for patient's county of residence: (1) large metropolitan, (2) small metropolitan, (3) micropolitan, (4) non-core	
Major Diagnosis Category (MDC)	MDC	1988-2006	MDC in use on discharge date	
	MDC10	*1988-1999*	*MDC Version 10 (effective October 1992 - September 1993)*	
	MDC18	*1998-2005*	*MDC Version 18 (effective October 2000 - September 2001)*	
	MDC24	2006	MDC Version 24 (effective October 2006 - September 2007)	
Median household income for patient's ZIP Code	ZIPINC_QRTL	2003-2006	Median household income quartiles for patient's ZIP Code. For 2005, the median income quartiles are defined as: (1) $1 - $35,999; (2) $36,000 - $44,999; (3) $45,000 - $58,999; and (4) $59,000 or more.	
	ZIPINC	*1998-2002*	*Median household income category in files beginning in 1998: (1) $1-$24,999, (2) $25,000-$34,999, (3) $35,000-$44,999, (4) $45,000 and above*	
	ZIPINC4	*1988-1997*	*Median household income category in files prior to 1998: (1) $1-$25,000, (2) $25,001-$30,000, (3) $30,001-$35,000, (4) $35,001 and above*	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	ZIPINC8	*1988-1997*	*Median household income category in files prior to 1998: (1) $1-$15,000, (2) $15,001-$20,000, (3) $20,001-$25,000, (4) $25,001-$30,000, (5) $30,001-$35,000, (6) $35,001-$40,000, (7) $40,001-$45,000, (8) $45,001 or more*	
Neonatal/ maternal flag	NEOMAT	1988-2006	Assigned from diagnoses and procedure codes: (0) not maternal or neonatal, (1) maternal diagnosis or procedure, (2) neonatal diagnosis, (3) maternal and neonatal on same record	
Payer information	PAY1	1988-2006	Expected primary payer, uniform: (1) Medicare, (2) Medicaid, (3) private including HMO, (4) self-pay, (5) no charge, (6) other	
	PAY1_N	*1988-1997*	*Expected primary payer, nonuniform: (1) Medicare, (2) Medicaid, (3) Blue Cross, Blue Cross PPO, (4) commercial, PPO, (5) HMO, PHP, etc., (6) self-pay, (7) no charge, (8) Title V, (9) Worker's Compensation, (10) CHAMPUS, CHAMPVA, (11) other government, (12) other*	
	PAY1_X	1998-2006	Expected primary payer, as received from the data source	
	PAY2	1988-2006	Expected secondary payer, uniform: (1) Medicare, (2) Medicaid, (3) private including HMO, (4) self-pay, (5) no charge, (6) other	AZ, CA, CO, FL, HI, IA, NH, OH, OK, RI, SD, VA
	PAY2_N	*1988-1997*	*Expected secondary payer, nonuniform: (1) Medicare, (2) Medicaid, (3) Blue Cross, Blue Cross PPO, (4) commercial, PPO, (5) HMO, PHP, etc., (6) self-pay, (7) no charge, (8) Title V, (9) Worker's Compensation, (10) CHAMPUS, CHAMPVA, (11) other government, (12) other*	
	PAY2_X	1998-2006	Expected secondary payer, as received from the data source	AZ, CA, CO, FL, HI, IA, NH, OH, OK, RI, SD, VA
Physician identifiers,	*MDID_S*	*1988-2000*	*Synthetic attending physician number in files prior to 2001*	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
synthetic	MDNUM1_R	2003-2006	Re-identified attending physician number in files starting in 2003	CA, CT, GA, HI, IL, IN, MA, NC, OH, OK, UT, VT, WI, WV
	MDNUM1_S	*2001-2002*	*Synthetic attending physician number in files beginning in 2001 and discontinued in 2003*	
	SURGID_S	*1988-2000*	*Synthetic secondary physician number in files prior to 2001*	
	MDNUM2_R	2003-2006	Re-identified secondary physician number in files starting in 2003	CA, CT, GA, HI, IL, IN, MA, NC, OH, OK, UT, VT, WI, WV
	MDNUM2_S	*2001-2002*	*Synthetic secondary physician number in files beginning in 2001 and discontinued in 2003*	
Procedure information	PR1 - PR15	1988-2006	Procedures, principal and secondary (ICD-9-CM)	
	NPR	1988-2006	Number of procedures coded on the original record	
	DSNPR	*1988-1997*	*Number of procedure fields in this data source*	
	PRSYS	*1988-1997*	*Procedure system (ICD-9-CM)*	
	PRV1 -PRV15	*1988-1997*	*Procedure validity flag*	
	PRDAY1	1988-2006	Number of days from admission to principal procedure.	IL, OH, OK, UT, WA, WV
	PRDAY2 - PRDAY15	1998-2006	Number of days from admission to secondary procedures	AZ, CO, IL, IN, MI, NV, OH, OK, UT, VA, WA, WI, WV
Race of patient	RACE	1988-2006	Race, uniform coding: (1) white, (2) black, (3) Hispanic, (4) Asian or Pacific Islander, (5) Native American, (6) other	GA, IL, KY, MN, NV, OH, OR, WA, WV
Record identifier, synthetic	KEY	1998-2006	Unique record number for file beginning in 1998	
	SEQ	*1988-1997*	*Unique record number for NIS prior to 1998*	
	SEQ_SID	*1988-1997*	*Unique record number for NIS prior to*	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
			1998	
	PROCESS	*1988-1997*	*Processing number for NIS prior to 1998*	
Total charges	TOTCHG	1988-2006	Total charges, edited	
	TOTCHG_X	1988-2006	Total charges, as received from data source	

Return to Introduction

Table 2. Data Elements in the NIS Hospital Weights Files

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

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
Discharge counts	N_DISC_U	1988-2006	Number of AHA universe discharges in the stratum	
	S_DISC_U	1988-2006	Number of sampled discharges in the sampling stratum (NIS_STRATUM or STRATUM)	
	S_DISC_S	*1988-1997*	*Number of sampled discharges in the stratum STRAT_ST*	
	N_DISC_F	*1988-1997*	*Number of frame discharges in the stratum*	
	N_DISC_S	*1988-1997*	*Number of State's discharges in the stratum*	
	TOTAL_DISC	1998-2006	Total number of discharges from this hospital in the NIS	
	TOTDSCHG	*1988-1997*	*Total number of discharges from this hospital in the NIS*	
Discharge weights	DISCWT	1998-2006	Discharge weight used in the NIS beginning in 1998. In all data years except 2000, this weight is used to create national estimates for all analyses. In 2000 only, this weight is used to create national estimates for all analyses, excluding those that involve total charges.	
	DISCWT_U	*1988-1997*	*Discharge weights used in the NIS prior to 1998.*	
	DISCWT_F	*1988-1997*	*Discharge weights to the sample frame are available only in 1988-1997*	
	DISCWT_S	*1988-1997*	*Discharge weights to the State are available only in 1988-1997*	
	DISCWTcharge	*2000*	*Discharge weight for national estimates of total charges for 2000 only.*	
Discharge Year	YEAR	1988-2006	Discharge year	
Hospital counts	*N_HOSP_F*	*1988-1997*	*Number of frame hospitals in the stratum*	
	N_HOSP_S	*1988-1997*	*Number of State's hospitals in the stratum*	
	N_HOSP_U	1988-2006	Number of AHA universe hospitals in the stratum	
	S_HOSP_S	*1988-1997*	*Number of sampled hospitals in STRAT_ST*	
	S_HOSP_U	1988-2006	Number of sampled hospitals in the stratum (NIS_STRATUM or STRATUM)	
Hospital identifiers	HOSPID	1988-2006	HCUP hospital number (links to Inpatient Core files)	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	AHAID	1988-2006	AHA hospital identifier that matches AHA Annual Survey Database (not available for all States)	GA, HI, IN, KS, MI, NE, OH, OK, SC, SD, TN, TX
	IDNUMBER	1988-2006	AHA hospital identifier without the leading 6 (not available for all States)	GA, HI, IN, KS, MI, NE, OH, OK, SC, SD, TN, TX
	HOSPNAME	1993-2006	Hospital name from AHA Annual Survey Database (not available for all States)	AR, GA, HI, IN, KS, MI, NE, OH, OK, SC, SD, TN, TX
Hospital location	HOSPADDR	1993-2006	Hospital address from AHA Annual Survey Database (not available for all States)	AR, GA, HI, IN, KS, MI, NE, OH, OK, SC, SD, TN, TX
	HOSPCITY	1993-2006	Hospital city from AHA Annual Survey Database (not available for all States)	AR, GA, HI, IN, KS, MI, NE, OH, OK, SC, SD, TN, TX
	HOSPST	1988-2006	Hospital State postal code for hospital (e.g., AZ for Arizona)	
	HOSPSTCO	2002-2006	Modified Federal Information Processing Standards (FIPS) State/county code	GA, HI, IN, KS, MI, NE, OH, OK, SC, SD, TN, TX
	HFIPSSTCO	2005-2006	Unmodified Federal Information Processing Standards (FIPS) State/county code for the hospital. Links to the Area Resource File (available from the Bureau of Health Professions, Health Resources and Services Administration)	GA, HI, IN, KS, MI, NE, OH, OK, SC, SD, TN, TX
	HOSPZIP	1993-2006	Hospital ZIP Code from AHA Annual Survey Database (not available for all States)	AR, GA, HI, IN, KS, MI, NE, OH, OK, SC, SD, TN, TX
Hospital characteristics	HOSP_BEDSIZE	1998-2006	Bed size of hospital: (1) small, (2) medium, (3) large	
	H_BEDSZ	*1993-1997*	*Bed size of hospital: (1) small, (2) medium, (3) large*	
	ST_BEDSZ	*1988-1992*	*Bed size of hospital: (1) small, (2) medium, (3) large*	

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes	Unavailable in 2006 for:
	HOSP_CONTROL	1998-2006	Control/ownership of hospital: (0) government or private, collapsed category, (1) government, nonfederal, public, (2) private, non-profit, voluntary, (3) private, invest-own, (4) private, collapsed category	
	H_CONTRL	*1993-1997*	*Control/ownership of hospital: (1) government, nonfederal (2) private, non-profit (3) private, investor-own*	
	ST_OWNER	*1988-1992*	*Control/ownership of hospital: (1) public (2) private, non-profit (3) private for profit*	
	HOSP_ LOCATION	1998-2006	Location: (0) rural, (1) urban	
	H_LOC	*1993-1997*	*Location: (0) rural, (1) urban*	
	HOSP_ LOCTEACH	1998-2006	Location/teaching status of hospital: (1) rural, (2) urban non-teaching, (3) urban teaching	
	H_LOCTCH	*1993-1997*	*Location/teaching status of hospital: (1) rural, (2) urban non-teaching, (3) urban teaching*	
	LOCTEACH	*1988-1992*	*Location/teaching status of hospital: (1) rural, (2) urban non-teaching, (3) urban teaching*	
	HOSP_REGION	1998-2006	Region of hospital: (1) Northeast, (2) Midwest, (3) South, (4) West	
	H_REGION	*1993-1997*	*Region of hospital: (1) Northeast, (2) Midwest, (3) South, (4) West*	
	ST_REG	*1988-1992*	*Region of hospital: (1) Northeast, (2) Midwest, (3) South, (4) West*	
	HOSP_TEACH	1998-2006	Teaching status of hospital: (0) non-teaching, (1) teaching	
	H_TCH	*1993-1997*	*Teaching status of hospital: (0) non-teaching, (1) teaching*	
	NIS_STRATUM	1998-2006	Stratum used to sample hospitals beginning in 1998; includes geographic region, control, location/teaching status, and bed size	
	STRATUM	*1988-1997*	*Stratum used to sample hospitals prior to 1998; includes geographic region, control, location/teaching status, and bed size*	
	STRAT_ST	*1988-1997*	*Stratum for State-specific weights*	
Hospital weights	HOSPWT	1998-2006	Weight to hospitals in AHA universe (i.e., total U.S.) beginning in 1998	
	HOSPWT_U	*1988-1997*	*Weight to hospitals in AHA universe (i.e., total U.S.) prior to 1998*	
	HOSPWT_F	*1988-1997*	*Weight to hospitals in the sample frame*	
	HOSPWT_S	*1988-1997*	*Weight to hospitals in the State*	

Return to Introduction

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

All data elements listed below are available for all States in the 2005 NIS Disease Severity Measures files.

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes
AHRQ Comorbidity Software (AHRQ)	CM_AIDS	2002-2006	AHRQ comorbidity measure: Acquired immune deficiency syndrome
	CM_ALCOHOL	2002-2006	AHRQ comorbidity measure: Alcohol abuse
	CM_ANEMDEF	2002-2006	AHRQ comorbidity measure: Deficiency anemias
	CM_ARTH	2002-2006	AHRQ comorbidity measure: Rheumatoid arthritis/collagen vascular diseases
	CM_BLDLOSS	2002-2006	AHRQ comorbidity measure: Chronic blood loss anemia
	CM_CHF	2002-2006	AHRQ comorbidity measure: Congestive heart failure
	CM_CHRNLUNG	2002-2006	AHRQ comorbidity measure: Chronic pulmonary disease
	CM_COAG	2002-2006	AHRQ comorbidity measure: Coagulopathy
	CM_DEPRESS	2002-2006	AHRQ comorbidity measure: Depression
	CM_DM	2002-2006	AHRQ comorbidity measure: Diabetes, uncomplicated
	CM_DMCX	2002-2006	AHRQ comorbidity measure: Diabetes with chronic complications
	CM_DRUG	2002-2006	AHRQ comorbidity measure: Drug abuse
	CM_HTN_C	2002-2006	AHRQ comorbidity measure: Hypertension, uncomplicated and complicated
	CM_HYPOTHY	2002-2006	AHRQ comorbidity measure: Hypothyroidism
	CM_LIVER	2002-2006	AHRQ comorbidity measure: Liver disease
	CM_LYMPH	2002-2006	AHRQ comorbidity measure: Lymphoma
	CM_LYTES	2002-2006	AHRQ comorbidity measure: Fluid and electrolyte disorders
	CM_METS	2002-2006	AHRQ comorbidity measure: Metastatic cancer
	CM_NEURO	2002-2006	AHRQ comorbidity measure: Other neurological disorders
	CM_OBESE	2002-2006	AHRQ comorbidity measure: Obesity
	CM_PARA	2002-2006	AHRQ comorbidity measure: Paralysis
	CM_PERIVASC	2002-2006	AHRQ comorbidity measure: Peripheral vascular disorders
	CM_PSYCH	2002-2006	AHRQ comorbidity measure: Psychoses
	CM_PULMCIRC	2002-2006	AHRQ comorbidity measure: Pulmonary circulation disorders
	CM_RENLFAIL	2002-2006	AHRQ comorbidity measure: Renal failure
	CM_TUMOR	2002-2006	AHRQ comorbidity measure: Solid tumor without metastasis
	CM_ULCER	2002-2006	AHRQ comorbidity measure: Peptic ulcer disease excluding bleeding
	CM_VALVE	2002-2006	AHRQ comorbidity measure: Valvular disease
	CM_WGHTLOSS	2002-2006	AHRQ comorbidity measure: Weight loss

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes
All Patient Refined DRG (3M)	APRDRG	2002-2006	All Patient Refined DRG
	APRDRG_Risk_ Mortality	2002-2006	All Patient Refined DRG: Risk of Mortality Subclass
	APRDRG_Severity	2002-2006	All Patient Refined DRG: Severity of Illness Subclass
All-Payer Severity-adjusted DRG (HSS, Inc.)	APSDRG	2002-2006	All-Payer Severity-adjusted DRG
	APSDRG_ Mortality_Weight	2002-2006	All-Payer Severity-adjusted DRG: Mortality Weight
	APSDRG_LOS_ Weight	2002-2006	All-Payer Severity-adjusted DRG: Length of Stay Weight
	APSDRG_Charge _Weight	2002-2006	All-Payer Severity-adjusted DRG: Charge Weight
Disease Staging (Medstat)	DS_DX_ Category1	2002-2006	Disease Staging: Principal Disease Category
	DS_Stage1	2002-2006	Disease Staging: Stage of Principal Disease Category
	DS_LOS_Level	2002-2006	Disease Staging: Length of Stay Level
	DS_LOS_Scale	2002-2006	Disease Staging: Length of Stay Scale
	DS_Mrt_Level	2002-2006	Disease Staging: Mortality Level
	DS_Mrt_Scale	2002-2006	Disease Staging: Mortality Scale
	DS_RD_Level	2002-2006	Disease Staging: Resource Demand Level
	DS_RD_Scale	2002-2006	Disease Staging: Resource Demand Scale
Linkage Variables	HOSPID	2002-2006	HCUP hospital identification number
	KEY	2002-2006	HCUP record identifier

Return to Introduction

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

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

Type of Data Element	HCUP Variable Name	Years Available	Coding Notes
Clinical Classifications Software category for Mental Health and Substance Abuse (CCS-MHSA)	CCSMGN1 – CCSMGN15	2005-2006	CCS-MHSA general category for all diagnoses
	CCSMSP1 – CCSMSP15	2005-2006	CCS-MHSA specific category for all diagnoses
	ECCSMGN1 – ECCSMGN4	2005-2006	CCS-MHSA general category for all external cause of injury codes
Chronic Condition Indicator	CHRON1 – CHRON15	2005-2006	Chronic condition indicator for all diagnoses: (0) non-chronic condition, (1) chronic condition
	CHRONB1 – CHRONB15	2005-2006	Chronic condition indicator body system for all diagnoses: (1) Infectious and parasitic disease, (2) Neoplasms, (3) Endocrine, nutritional, and metabolic diseases and immunity disorders, (4) Diseases of blood and blood-forming organs, (5) Mental disorders, (6) Diseases of the nervous system and sense organs, (7) Diseases of the circulatory system, (8) Diseases of the respiratory system, (9) Diseases of the digestive system, (10) Diseases of the genitourinary system, (11) Complications of pregnancy, childbirth, and the puerperium, (12) Diseases of the skin and subcutaneous tissue, (13) Diseases of the musculoskeletal system, (14) Congenital anomalies, (15) Certain conditions originating in the perinatal period, (16) Symptoms, signs, and ill-defined conditions, (17) Injury and poisoning, (18) Factors influencing health status and contact with health services
Procedure Class	PCLASS1 – PCLASS15	2005-2006	Procedure Class for all procedures: (1) Minor Diagnostic, (2) Minor Therapeutic, (3) Major Diagnostic, (4) Major Therapeutic
Linkage Variables	HOSPID	2002-2006	HCUP hospital identification number
	KEY	2002-2006	HCUP record identifier

Return to Introduction

ENDNOTES

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

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

[3] We used the following American Hospital Association Annual Survey Database (Health Forum, LLC © 2007) data elements to assign the NIS Teaching Hospital Indicator:

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

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

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

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

Beginning with the 1998 NIS, we used the following SAS code to assign the teaching hospital status indicator, HOSP_TEACH:

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

[4] Most AHA Annual Survey Database files do not cover a January-to-December period for every hospital. The numbers of hospitals for 1988-1991 are based on adjusted versions of the files which we created by apportioning the data from adjacent survey files across calendar years. The numbers of hospitals for later years are based on the unadjusted AHA Annual Survey Database files.

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