

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)

2009

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 2011

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

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

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 2009 NATIONWIDE INPATIENT SAMPLE (NIS)?	3
UNDERSTANDING THE NIS	4
ABSTRACT	5
INTRODUCTION TO THE HCUP NATIONWIDE INPATIENT SAMPLE (NIS)	6
Overview of NIS Data	6
<i>NIS Data Sources, Hospitals, and Inpatient Stays</i>	7
<i>State-Specific Restrictions</i>	7
<i>Contents of DVD</i>	8
<i>NIS Data Elements</i>	8
Getting Started	9
<i>Copying and Decompressing the ASCII Files</i>	9
<i>Downloading and Running the Load Programs</i>	9
<i>NIS Documentation</i>	9
<i>HCUP On-Line Tutorials</i>	10
HOW TO USE THE NIS FOR DATA ANALYSIS	10
Calculating National Estimates	10
Studying Trends	12
Choosing Data Elements for Analysis	12
Hospital-Level Data Elements	13
Constructing Patient Population Characteristics	14
ICD-9-CM Diagnosis and Procedure Codes	14
Missing Values	15
Variance Calculations	15
Computer Software for Variance Calculations	16
Longitudinal Analyses	17
Discharge Subsamples	18
SAMPLING OF HOSPITALS	18
Sampling of Hospitals Included in the NIS	18
<i>The NIS Hospital Universe</i>	18
<i>Hospital Merges, Splits, and Closures</i>	19
<i>Stratification Variables</i>	19
Hospital Sampling Frame	21
Hospital Sample Design	22
<i>Design Considerations</i>	22
<i>Overview of the Sampling Procedure</i>	23
<i>Subsamples</i>	24
<i>Change to Hospital Sampling Procedure Beginning with the 1998 NIS</i>	24

<i>Zero-Weight Hospitals</i>	24
Final Hospital Sample	24
SAMPLE WEIGHTS	27
Hospital Weights	27
Discharge Weights	27
APPENDIX I: TABLES AND FIGURES	I-1
Table 1: 2009 Data Sources	I-1
Table 2: Number of NIS States, Hospitals, and Discharges, by Year	I-3
Table 3: Summary of NIS Releases	I-4
Table 4: Summary of NIS Data Sources, Hospitals, and Inpatient Stays, 1988-2009	I-6
Table 5: NIS Related Reports and Database Documentation Available on HCUP-US ...	I-9
Figure 1: Hospital Universe, by Year	I-10
Figure 2: NIS States, by Region	I-11
Table 6: All States, by Region	I-11
Table 7: Bed Size Categories, by Region	I-12
Figure 3: NIS Hospital Sampling Frame, by Year	I-13
Figure 4: Number of Hospitals in the 2009 Universe, Frame, and Sample for Frame States	I-14
<i>Part A: Arkansas – Indiana</i>	I-14
<i>Part B: Kansas – North Carolina</i>	I-15
<i>Part C: Nebraska – Rhode Island</i>	I-16
<i>Part D: South Carolina – Wyoming</i>	I-17
Table 8: Number of Hospitals and Discharges in 2009 AHA Universe, Frame, and NIS, by State	I-18
Figure 5: Number of Hospitals Sampled, by Year	I-20
Figure 6: Number of NIS Discharges, Unweighted, by Year	I-21
Figure 7: Number of NIS Discharges, Weighted, by Year	I-22
Figure 8: Number of Hospitals in the 2009 Universe, Frame, Sample, Target, and Surplus, by Region	I-23
Figure 9: Percentage of U.S. Population in the 2009 NIS States, by Region	I-24
Figure 10: Number of Discharges in the 2009 NIS, by State	I-25
APPENDIX II: STATE-SPECIFIC RESTRICTIONS	II-1
APPENDIX III: DATA ELEMENTS	III-1
Table 1. Data Elements in the NIS Inpatient Core Files	III-1
Table 2. Data Elements in the NIS Hospital Weights Files	III-12
Table 3. Data Elements in the NIS Disease Severity Measures Files	III-16
Table 4. Data Elements in the NIS Diagnosis and Procedure Groups Files	III-19
ENDNOTES	E-1

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

***** REMINDER *****

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

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

- Will not use the data for any purpose other than research or aggregate statistical reporting.
- Will not re-release any data to unauthorized users.
- Will not identify or attempt to identify any individual. **Will not report any statistics where the number of observations (i.e., individual discharge records) in any given cell of tabulated data is less than or equal to 10.**
- Will not link HCUP data to data from another source that identifies individuals.
- Will not report information that could identify individual establishments (e.g., hospitals).
- Will not use the data concerning individual establishments for commercial or competitive purposes involving those establishments.
- Will not use the data to determine rights, benefits, or privileges of individual establishments.
- Will not identify or attempt to identify any establishment when its identity has been concealed on the database.
- Will not contact establishments included in the data.
- Will not attribute to data contributors any conclusions drawn from the data.
- Will not use data elements from the proprietary severity adjustment software packages (3M APR-DRGs, HSS APS-DRGs, and Thomson Reuters Disease Staging) for any commercial purpose or to disassemble, decompile, or otherwise reverse engineer the proprietary software.
- Must acknowledge the "Healthcare Cost and Utilization Project (HCUP)," as described in the Data Use Agreement, in reports.

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

[†] The on-line Data Use Agreement training session and the Data Use Agreement are available on the HCUP User Support (HCUP-US) Website at <http://www.hcup-us.ahrq.gov>.

[‡] Specific provisions are detailed in the Data Use Agreement for Nationwide Inpatient Sample.

HCUP CONTACT INFORMATION

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

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

After completing the on-line training tool, please submit signed data use agreements to HCUP at:

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

Phone: (866) 290-HCUP (4287)

Fax: (301) 427-1430

Website: <http://www.ahrq.gov/data/hcup/>

For technical assistance:

Visit the HCUP-US Website at

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

Or for issues related to reading and loading the files contact the HCUP Central Distributor at

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

Fax: (866) 792-5313

E-mail: HCUPDistributor@ahrq.gov

For other issues send an e-mail to HCUP User Support at

hcup@ahrq.gov

WHAT'S NEW IN THE 2009 NATIONWIDE INPATIENT SAMPLE (NIS)?

- The 2009 NIS contains two additional states: New Mexico and Montana.
- The following data elements were added to the Core File beginning with the 2009 NIS:
 - Ten additional secondary diagnoses for a total of 25 diagnoses
 - Ten additional secondary Clinical Classifications Software (CCS) diagnosis categories
 - Major Diagnostic Category (MDC) in use on discharge date, calculated without Present on Admission (POA) indicators (MDC_NoPOA). MDC_NoPOA and the Diagnosis Related Group (DRG) calculated without POA indicators (DRG_NoPOA) are useful because the lack of POA flags from many states prevents the assignment of the standard MDC and DRG for a few DRGs involving Hospital Acquired Conditions (HAC).
- The following data elements were added to the Diagnosis and Procedure Groups (DX_PR_GRPs) File beginning with the 2009 NIS:
 - Multi-level CCS categories for the principal diagnosis and the first listed E-code and procedure.
 - Ten additional Chronic Condition Indicators
 - Ten additional Chronic Condition Body System Indicators
- Georgia and Illinois provide race/ethnicity data for the first time for data year 2009. In addition, Iowa improved their race/ethnicity data in 2009 by capturing information about the Hispanic population.
- The 2009 NIS is distributed on a single DVD-ROM instead of two CD-ROMs.

UNDERSTANDING THE NIS

This document, Introduction to the NIS, 2009, 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. Important considerations for data analysis are provided along with references to detailed reports. In-depth documentation for the NIS is available on the HCUP User Support (HCUP-US) Website (www.hcup-us.ahrq.gov).

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

**The Agency for Healthcare Research and Quality and
the staff of the Healthcare Cost and Utilization Project (HCUP) thank you for purchasing
the HCUP 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; for 2009, these states comprise 96 percent of the U.S. population. 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. For NIS data 1997 and earlier, revised weights should be used to make estimates comparable to later data. Refer to [NIS Trends Weights Files](#) and 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. For Medicare, the NIS includes Medicare Advantage patients, a population that is missing from Medicare claims data but that comprises as much as 20 percent of Medicare beneficiaries. The NIS' large sample size enables analyses of rare conditions, such as specific types of cancer; 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 © 2010) 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. In 2009, 18 of the 44 states do not include the hospital name and 17 of these do not include the AHA hospital identifier. Thus 43% of hospitals in the NIS do not include AHA hospital identifiers and cannot be linked to other data sources at the hospital level.

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*, available on the HCUP-US Website, which describes these changes.) 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 © 2010) 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, therefore such facilities are included in the NIS sampling frame. 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.

The NIS is available yearly, beginning with 1988, allowing analysis of trends over time. (Analyses of time trends are recommended from 1993 forward. For NIS data 1997 and earlier, revised weights should be used to make estimates comparable to later data. Refer to *NIS Trends Weights Files* and the report, *Using the HCUP Nationwide Inpatient Sample to Estimate Trends*, available on the HCUP User Support (HCUP-US) Website, for details.)

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 © 2010)
- NIS Documentation and tools – including file specifications, programs for loading the 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 and procedures, specifically HIV/AIDS, behavioral health, and abortion. Detailed information on these State-specific restrictions is available in *Appendix II*.

Contents of DVD

The NIS is distributed as fixed-width ASCII formatted data files compressed with WinZip®. Previously it was distributed on two CD ROMs, but beginning with the 2009 NIS, it is distributed on a single DVD. It includes the following files:

Inpatient Core File: This inpatient discharge-level file contains data for all 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.

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

Diagnosis and Procedure Groups File: This discharge-level file contains data elements from AHRQ software tools designed to facilitate the use of the ICD-9-CM diagnostic and procedure information in the HCUP databases. The unit of observation is an *inpatient stay record*. The HCUP unique record identifier (KEY) provides the linkage between the Core file and the Diagnosis and Procedure Groups file. Table 4 in Appendix III contains a list of data elements in the Diagnosis and Procedure Groups file. This file is 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 to calculate national estimates. 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

In order to load and analyze the NIS data on a computer, you will need the following:

- A DVD drive
- A hard drive with 15 gigabytes of space available
- SAS, SPSS, Stata or similar analysis software.

Copying and Decompressing the ASCII Files

To copy and decompress the data from the DVD, follow these steps:

- 1) Create a directory for the NIS on your hard drive.
- 2) Unzip each ASCII file from the DVD, saving it into the new directory using WinZip® or a similar utility. (Evaluation versions of WinZip may be downloaded from the WinZip Website at www.winzip.com.)

Downloading and Running the Load Programs

Programs to load the data into SAS, SPSS, or Stata, are available on the HCUP User Support Website (HCUP-US). To download and run the load programs, follow these steps:

- 1) Go to the NIS Database Documentation page on HCUP-US at <http://www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp>.
- 2) Go to the "Load Programs" section on this page.
- 3) Click on "SAS Load Programs", "SPSS Load Programs", or "STATA Load Programs" to go to the corresponding Load Programs page.
- 4) Select and download the load programs you need. The load programs are specific to the data year. For example, the load program for the 2009 NIS Core file is linked to "Core File" under "2009 NIS". Save the load programs into the same directory as the NIS ASCII files on your computer.
- 5) Edit and run the load programs as appropriate for your environment to load and save the analysis files. For example, add directory paths for the input and output files if needed.

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.

HCUP On-Line Tutorials

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

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

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

The [Producing National HCUP Estimates](#) tutorial is designed to help users understand how the three nationwide databases – the NIS, NEDS, and KID – can be used to produce national and regional estimates.

The [Calculating Standard Errors](#) tutorial shows how to accurately determine the precision of the estimates produced from the HCUP nationwide databases. Users will learn two methods for calculating standard errors for estimates produced from the HCUP nationwide databases.

New tutorial are added periodically. The Online Tutorial Series is located on the HCUP-US Website at http://hcup-us.ahrq.gov/tech_assist/tutorials.jsp.

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. Researchers wishing to examine readmissions should use State Inpatient Databases (SID) (<http://www.hcup-us.ahrq.gov/sidoverview.jsp>) and accompanying Revisit Files which allow identification of readmissions for individual patients (<http://www.hcup-us.ahrq.gov/toolssoftware/revisit/revisit.jsp>).

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	• <i>The 10% Subsample Core File was discontinued with the 2005 NIS.</i>
2001-2004	• DISCWT for all analyses	• DISCWT10 for all analyses
2000	<ul style="list-style-type: none"> • DISCWT to create nationwide estimates for all analyses <u>except</u> those that involve total charges. • DISCWTCHARGE to create nationwide estimates of total charges. 	<ul style="list-style-type: none"> • DISCWT10 to create nationwide estimates for all analyses, <u>except</u> 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/nhds.htm>) 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, check your estimates against 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. HCUPnet generates statistics using the NIS, KID, and SID for those States that have agreed to participate. In addition, HCUPnet provides Quick Statistics – ready-to-use tables on commonly requested information – as well as national statistics based on the AHRQ Quality Indicators.

Studying Trends

- When studying trends over time using the 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 alternate 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. 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.

Hospital-Level Data Elements

- Note that specific hospital identifiers, including the AHA identifier, are not available for 43% of hospitals in the NIS because certain states do not release hospital identifiers. This means that for those hospitals, you will not be able to link the NIS to outside data sources that require hospital-specific identifiers. However, there are hospital-level data elements for nearly all hospitals in the NIS that allow you to study certain hospital characteristics including ownership/control, teaching status, rural/urban location, bedsize, and Census region of the country.
- The sampling strategy for the NIS was redesigned in 1998 and one stratifier (hospital ownership) was redefined to reduce the number of ownership strata in some regions. This redefinition resulted in collapsed ownership categories that were useful in addressing the requirements of some HCUP Partners to protect the confidentiality of hospitals, but they were not as informative as the original variable. Thus beginning with the 1998 NIS, the variable H_CTRL was replaced by the HOSP_CTRL (collapsed) data element.

To facilitate analyses by hospital ownership, the data element containing hospital ownership categories without any collapsing (H_CTRL) was restored to the Hospital File beginning with the 2008 NIS. The H_CTRL data element contains the following three hospital ownership categories:

- 1 - Government, nonfederal
- 2 - Private, non-profit
- 3 - Private, invest-own

Note, however, that H_CTRL is set to missing in a few instances to satisfy HCUP Partner requirements.

The H_CTRL values for 1998-2007 NIS hospitals are now available in the supplemental NIS Hospital Ownership File on HCUP-US at <http://www.hcup-us.ahrq.gov/db/nation/nis/nisownership.jsp>.

- New hospital-level data elements were added to the NIS hospital file beginning in 2007. These data elements, which are listed in Table 2 of Appendix III, are derived from the AHA Annual Survey Database (Health Forum, LLC © 2010). While these data elements enable a greater breadth of analysis, certain limitations apply.

1. Some of the new data elements pertain to nurse staffing at hospitals, which is reported as a total for the hospital/facility. Therefore, it is possible that some hospitals may also have included counts from nursing homes.
2. The adjusted patient days are based on inpatient days with an adjustment made for outpatient activities, calculated as: $\text{Inpatient days} * (1 + \text{Outpatient Revenue} / \text{Inpatient Revenue})$.
3. Data for hospitals that do not complete the AHA Annual Survey are missing from the new data elements.

A detailed description of the data elements is available on HCUP-US. Note that some HCUP states do not allow the release of this information.

Constructing Patient Population Characteristics

- Summary variables that provide insight into the patient population can be constructed at the hospital level using HOSPID and the variables already present in the NIS. Suggested summary variables include:
 - Percentage of Hispanic discharges
 - Percentage of black discharges
 - Percentage of non-white discharges
 - Percentage of pediatric discharges
 - Percentage of discharges for adults age 65 and older
 - Percentage of discharges with an expected primary payer of Medicare
 - Percentage of discharges with an expected primary payer of Medicaid
 - Percentage of discharges with an expected primary payer of private insurance
 - Percentage of uninsured discharges (expected primary payer of self-pay and no charge)
 - Percentage of discharges with no comorbidities
 - Percentage of discharges with one or two comorbidities
 - Percentage of discharges with three or more comorbidities

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/icd/icd9cm.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 25 diagnoses (15 prior to the 2009 NIS) and 15 procedures, the number of diagnoses and procedures varies by State. Some States provide as many as 31 diagnoses and procedures or more, while other States provide as few as 9 diagnoses and 6 procedures. Because very few cases have more than 25 diagnoses or 15 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). See the notes on diagnoses (<http://www.hcup-us.ahrq.gov/db/vars/dxn/nisnote.jsp>) and procedures (<http://www.hcup-us.ahrq.gov/db/vars/prn/nisnote.jsp>) on the HCUP-US Web site to view the number of diagnosis and procedure fields provided by each state.

- 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. For example, race is missing on 15% of discharges in the 2009 NIS because some hospitals and HCUP State Partners do not supply it. (The percentage of missing race values was higher in previous years.) Therefore race-specific estimates may be biased. This is especially true for estimates of discharge totals by race. Another set of data elements that are missing are hospital identifiers, which allow you to link to other datasets with the AHA hospital identifier. In 2009, about 43% of hospitals were missing specific identifiers.

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 © 2010). 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 © 2010).

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/metroareas/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 2009, 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 2009. In most cases, the difference between the universe and the frame represents the difference in the number of community, non-rehabilitation hospitals in the 2009 AHA Annual Survey Database (Health Forum, LLC © 2010) 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 over 96% of inpatient discharges from Texas universe hospitals because excluded hospitals tend to have relatively few discharges.

Similar to Texas, because smaller Louisiana hospitals are not required to submit data to the Louisiana Department of Health and Hospitals, a significant portion of Louisiana hospitals are omitted from the sampling frame. However, because excluded hospitals tend to have relatively few discharges, those available for the NIS include over 91% of inpatient discharges from Louisiana universe hospitals.

Refer to [Table 8 of Appendix I](#) for a full list of the number of hospitals, and discharges included in the 2009 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 1 lists the number of NIS States, hospitals, and discharges by year. For example, the National Hospital Discharge Survey (<http://www.cdc.gov/nchs/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. The number of discharges in the sample may vary slightly depending on the hospitals selected for the NIS in any given year. 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 © 2010).

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.

The small decline in both the number of discharges in the sample and the weighted number of discharges for 2009 is not related to any change in the sampling or weighting strategy. The reduction in the number of discharges is consistent with the information from the AHA Annual Survey of Hospitals.

Figure 8 presents a summary of the 2009 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 2009 NIS
- The percentage of estimated U.S. population included in NIS States.

Figure 10 depicts the number of discharges in the 2009 sample for each State.

Special consideration was needed to handle the Massachusetts data in the 2006 and the 2007 NIS. Fourth quarter data from sampled hospitals in Massachusetts were unavailable for inclusion in the 2006 and the 2007 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 $4 \div Q$, 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 $DW_{is}(\text{universe})$ was the discharge weight; $DN_s(\text{universe})$ represented the number of discharges from community hospitals in the universe within stratum s ; $ADN_s(\text{sample})$ 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.