

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 NATIONAL INPATIENT SAMPLE (NIS)
2012

THE 2012 NIS HAS BEEN REDESIGNED.

The new NIS is a **sample of discharges** from all hospitals participating in HCUP. For prior years, the NIS was a **sample of hospitals**.

Please read all documentation carefully.

These pages provide only an introduction to the NIS 2012 package.

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

Issued June 2014
Updated November 2015

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

Phone: (866) 290-HCUP (4287)
Email: hcup@ahrq.gov
Web site: <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
Email: HCUPDistributor@ahrq.gov

Table of Contents

SUMMARY OF DATA USE RESTRICTIONS.....	1
HCUP CONTACT INFORMATION.....	2
WHAT'S NEW IN THE 2012 NATIONAL INPATIENT SAMPLE (NIS)?	3
UNDERSTANDING THE NIS	4
ABSTRACT	5
INTRODUCTION TO THE HCUP NATIONAL INPATIENT SAMPLE (NIS)	7
Overview of NIS Data.....	7
<i>NIS 2012 Redesign</i>	7
<i>Impact of New Design on Estimates</i>	7
<i>Types of Hospitals Included in the NIS</i>	8
<i>Sample Design for 2012 NIS</i>	8
<i>Weighted Estimates</i>	9
<i>Trends</i>	9
<i>NIS Data Sources, Hospitals, and Inpatient Stays</i>	9
<i>Partner Restrictions</i>	9
<i>Contents of NIS</i>	10
<i>NIS Data Elements</i>	11
Getting Started.....	11
<i>Copying and Decompressing the ASCII Files</i>	11
<i>Downloading and Running the Load Programs</i>	12
<i>NIS Documentation</i>	12
<i>HCUP On-Line Tutorials</i>	12
HOW TO USE THE NIS FOR DATA ANALYSIS	13
Calculating National Estimates.....	13
Studying Trends.....	14
Choosing Data Elements for Analysis	15
Hospital-Level Data Elements	15
ICD-9-CM Diagnosis and Procedure Codes.....	15
Missing Values.....	16
Variance Calculations	16
Computer Software for Variance Calculations	17
Longitudinal Analyses	18
SAMPLING PROCEDURE	18
The NIS Hospital Universe	18
Stratification Data Elements	19
Design Considerations	21
Overview of the Sampling Procedure	22
WEIGHTS	23

APPENDIX I: TABLES AND FIGURES	I-1
Table 1: The 2012 National Inpatient Sample (NIS) Design Changes	I-1
Table 2: States Participating in the 2012 NIS	I-2
Table 3: Summary of NIS States, Hospitals, and Inpatient Stays, 1988-2012.....	I-4
Table 4: NIS Related Reports and Database Documentation Available on HCUP-US.....	I-7
Figure 1: NIS States, by Census Division	I-8
Table 5: Hospital Size Categories (in Number of Beds), by Region	I-9
Table 6: Number of Hospitals and Discharges in 2012 Universe, Frame, and NIS, by Census Division	I-10
Figure 2: Percentage of U.S. Population Covered in the 2012 NIS by Census Division and Region, Calculated using the Estimated U.S. Population on July 1, 2012.....	I-11
Figure 3: Number of Discharges (in Thousands) in 2012 NIS by Census Division and Census Region	I-12
APPENDIX II: STATE-SPECIFIC RESTRICTIONS.....	II-1
APPENDIX III: DATA ELEMENTS	III-1
Table 1: Data Elements in the 2012 NIS Inpatient Core Files	III-1
Table 2: Data Elements in the NIS Hospital Weights Files	III-5
Table 3: Data Elements in the NIS Disease Severity Measures Files	III-6
Table 4: Data Elements in the NIS Diagnosis and Procedure Groups Files.....	III-8
APPENDIX IV: TEACHING HOSPITAL INDICATOR ASSIGNMENT	IV-1
APPENDIX V: 2012 NIS REDESIGN REPORT EXECUTIVE SUMMARY.....	V-1
Previous Study Results	V-1
Finalizing the New Design.....	V-2
<i>Enlisted HCUP Partner Support</i>	V-3
<i>Removed Long-Term Acute Care Hospitals</i>	V-3
<i>Improved Estimates of the Total Number of Discharges in the Universe</i>	V-4
<i>Used State Hospital Identifiers Rather than AHA Hospital Identifiers.....</i>	V-5
<i>Estimated the Effects of Design Changes on Sample Estimates</i>	V-5
Conclusions	V-7
Table 1: Impact of Incremental Modifications to the Universe on Universe Statistics.....	V-8
Table 2: Impact of Incremental Modifications to the Universe on the Margin of Error for Sample Statistics	V-9

**HCUP NATIONAL INPATIENT SAMPLE (NIS)
SUMMARY OF DATA USE RESTRICTIONS**

******* 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) Web site 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: hcp@ahrq.gov

WHAT'S NEW IN THE 2012 NATIONAL INPATIENT SAMPLE (NIS)?

Beginning in 2012, the NIS was redesigned. It was formerly a **sample of hospitals**, and all discharges from those hospitals were retained.

The new NIS starting with 2012 data is a **sample of discharges** from all hospitals participating in HCUP.

The 2012 National Inpatient Sample (NIS) was redesigned to improve national estimates. To highlight the design change, beginning with 2012 data, AHRQ renamed the NIS from the “**Nationwide** Inpatient Sample” to the “**National** Inpatient Sample.” The redesign incorporates three major types of changes:

- Revisions to the sample design—the NIS is now a sample of discharge records from all HCUP-participating hospitals, rather than a sample of hospitals from which all discharges were retained.
- Revisions to how hospitals are defined—the NIS now uses the definitions of hospitals and discharges supplied by the statewide data organizations that contribute to HCUP, rather than the definitions used by the AHA Annual Survey.
- Revisions to enhance confidentiality—the NIS now eliminates State and hospital identifiers and other data elements that are not uniformly available across States.

The new sample strategy is expected to result in more precise estimates than the previous NIS design by reducing sampling error. For many estimates, confidence intervals under the new design are about half the length of confidence intervals under the previous design.

As a result of the changes implemented in the redesign, users should expect one-time disruptions to historical trends for counts, rates, and means estimated from the NIS, beginning with data year 2012. For trends analysis using NIS data 2011 and earlier, revised weights should be used to make estimates comparable to the new design beginning with 2012 data. Refer to [1993-2011 NIS Trend Weights Files](#) on the HCUP User Support (HCUP-US) Web site for details.

See [Table 1](#) in [Appendix I](#), The National Inpatient Sample (NIS) Design Changes, for a summary of changes to the NIS. For a detailed description of the NIS redesign and the effects of the design changes on sample estimates, please see the NIS Redesign Report Executive Summary in [Appendix V](#) or the full [NIS Redesign Report](#) available on the HCUP User Support (HCUP-US) Web site.

Information on previous years of the NIS may be found in the *Introduction to the NIS, 2011* ([PDF](#) file, 1.3 MB; [HTML](#)).

UNDERSTANDING THE NIS

This document, Introduction to the NIS, 2012, summarizes the content of the NIS and describes the development of the NIS sample and weights. 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) Web site (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 National Inpatient Sample (NIS).**

HCUP National Inpatient Sample (NIS)

ABSTRACT

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

The NIS is a database of hospital inpatient stays derived from billing data submitted by hospitals to statewide data organizations across the U.S. These inpatient data include clinical and resource use information typically available from discharge abstracts. Researchers and policy makers use the NIS to make national estimates of health care utilization, access, charges, quality, and outcomes.

The NIS covers all patients, including individuals covered by Medicare, Medicaid, or private insurance, as well as those who are 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, uncommon treatments, and special patient populations.

The NIS is sampled from the [State Inpatient Databases \(SID\)](#), all inpatient data that are currently contributed to HCUP. As displayed in [Figure 2](#) in [Appendix I](#), the 2012 NIS sampling frame covers more than 95 percent of the U.S. population; and as shown in [Table 6](#) in [Appendix I](#), it includes more than 94 percent of discharges from U.S. community hospitals. Weights are provided to calculate national estimates. [Table 2](#) in [Appendix I](#) lists the statewide data organizations participating in the NIS, and [Table 3](#) in [Appendix I](#) contains a summary of NIS data sources, hospitals, and discharges, by year.

The 2012 National Inpatient Sample (NIS) is redesigned to improve national estimates. To highlight the design change, beginning with 2012 data, AHRQ renamed the NIS from the “**Nationwide** Inpatient Sample” to the “**National** Inpatient Sample.” The redesign incorporates three major types of changes:

- Revisions to the sample design—the NIS is now a sample of discharge records from all HCUP-participating hospitals, rather than a sample of hospitals from which all discharges were retained.
- Revisions to how hospitals and discharges are defined—the NIS now uses the definitions of hospitals and discharges supplied by the statewide data organizations that contribute to HCUP, rather than the definitions used by the AHA Annual Survey.

- Revisions to enhance confidentiality—the NIS now eliminates State and hospital identifiers and other data elements that are not uniformly available across States.

The new sampling strategy is expected to result in more precise estimates than the previous NIS design by reducing sampling error. For many estimates, confidence intervals under the new design are about half the length of confidence intervals under the previous design.

Key features of the most recent NIS (2012) include:

- The NIS is drawn from all States participating in HCUP, covering more than 95 percent of the U.S. population.
- The NIS approximates a 20-percent **stratified sample of discharges** from U.S. community hospitals, excluding rehabilitation and long-term acute care hospitals.
- The self-weighting design of the new NIS reduces the margin of error for estimates and delivers more stable and precise estimates than previous versions of the NIS.
- The NIS protects patient confidentiality because State and hospital identifiers are no longer provided.
- The new NIS retains a large sample size, which enables analyses of rare conditions, uncommon treatments, and special patient populations.

Changes to the NIS may impact some types of analyses. For example, the elimination of hospital identifiers means that hospital linkages can no longer be done with the NIS and the sampling of discharges means that analyses relying on a census of discharges from sampled hospitals can no longer be performed. Because inpatient data are available for many individual States through the HCUP Central Distributor, state inpatient data can be used for many analyses no longer possible with the NIS.

See [Table 1](#) in [Appendix I](#), The National Inpatient Sample (NIS) Design Changes, for a summary of design changes. For a detailed description of the NIS redesign, please see the NIS Redesign Report Executive Summary in [Appendix V](#) or the full [NIS Redesign Report](#) available on the HCUP User Support (HCUP-US) Web site.

The NIS is available yearly, beginning with 1988, allowing analysis of trends over time. Analyses of time trends are recommended from 1993 forward. For trends analysis using NIS data 2011 and earlier, revised weights should be used to make estimates comparable to the new 2012 design. Refer to [1993-2011 NIS Trend Weights Files](#) and the report, [Using the HCUP Nationwide Inpatient Sample to Estimate Trends](#),¹ available on the HCUP User Support (HCUP-US) Web site, for details.

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

¹ As of June, 2014, this report had not yet been updated for the new 2012 NIS design. However, the methods described in the report are still valid.

INTRODUCTION TO THE HCUP NATIONAL INPATIENT SAMPLE (NIS)

Overview of NIS Data

The National 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 includes over 7 million inpatient stays.

The NIS contains clinical and resource use information included in a typical discharge abstract. The NIS is a database of hospital inpatient stays derived from billing data submitted by hospitals to statewide data organizations across the U.S.

NIS 2012 Redesign

The 2012 National Inpatient Sample (NIS) is redesigned to improve national estimates. To highlight the design change, beginning with 2012 data, AHRQ renamed the NIS from the “**Nationwide** Inpatient Sample” to the “**National** Inpatient Sample.” The redesign incorporates three major types of changes.

- First, the sample design was revised to create a sample of discharge records from all HCUP-participating hospitals, rather than all discharge records from a sample of hospitals.
- Second, the definition of discharges and hospitals was changed in several ways. The discharge universe was revised to exclude long-term acute care hospitals. In addition, the count of discharges in the universe is now based on the observed number of HCUP State Inpatient Database discharges, when available, rather than on American Hospital Association (AHA) admissions plus births. Finally, the definition of hospital entities is now based on State-supplied hospital identifiers rather than on AHA identification numbers.
- Third, confidentiality is enhanced by eliminating State identifiers and data elements that are not uniformly available across States, such as AHA hospital identifiers, secondary payer, and data elements with State-specific coding.

Impact of New Design on Estimates

The new NIS is now stratified by nine Census Divisions rather than four Census Regions, which will allow more refined analyses of geographic variation in U.S. hospitalizations. The new sampling strategy is expected to result in more precise estimates than the previous NIS design by reducing sampling error. For national-level estimates, the 2012 NIS systematic design reduces the margin of error by 42 to 48 percent over the previous NIS design for the outcomes studied (total discharges, average length-of-stay, average charges, and mortality rates); thus **the new NIS design generates estimates that are about twice as precise as those from the old design**. The margin of error is commonly used by the popular press to describe the reliability of sample statistics. Technically, it is the half-width of a confidence interval around a sample statistic, such as a rate or a mean. The systematic design also consistently reduced the margin of error for estimates at the DRG level.

As a result of the changes implemented in the 2012 redesign, users should expect one-time disruptions to historical trends for counts, rates, and means estimated from the NIS, beginning with data year 2012. For 2012 we expect overall trends in discharge counts to decline by about 4.3 percent, overall trends in average length-of-stay to decline by about 1.5 percent, overall

trends in total charges to decline by about 0.5 percent, and overall trends in hospital mortality to decline by about 2.0 percent. New weights for prior years of the NIS to make prior year estimates comparable to the new design implemented in 2012 are available for download under [1993-2011 NIS Trend Weights Files](#) from the [NIS Database Documentation](#) page on the HCUP-US Web site.

See [Table 1](#) in [Appendix I](#), The National Inpatient Sample (NIS) Design Changes, for a summary of design changes. For a detailed description of the NIS redesign and the effects on sample estimates, please see the NIS Redesign Report Executive Summary in [Appendix V](#) or the full [NIS Redesign Report](#) available on the HCUP User Support (HCUP-US) Web site.

The NIS sampling and weighting strategy was also revised in 1998. The full description of this revision 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 User Support (HCUP-US) Web site at <http://www.hcup-us.ahrq.gov>.

Types of Hospitals Included in the NIS

The NIS is a sample of discharges from U.S. community hospitals, defined as “all non-Federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions.”² Included among community hospitals are specialty hospitals such as obstetrics-gynecology, ear-nose-throat, orthopedic, and pediatric institutions. Also included are public hospitals and academic medical centers. Starting in 2005, the AHA included long term acute care facilities with average lengths-of-stay less than 30 days in the definition of community hospitals, and such facilities were included in the NIS sampling frame. However, because long-term acute care hospital data was not uniformly available from all States participating in HCUP, and their average length of stay (ALOS) was over 25 days (unlike other community hospitals with an ALOS of about 4.5 days), long-term acute care hospitals were excluded in the 2012 NIS redesign. Exclusion of long-term acute care hospitals mainly affects statistics related to the elderly – estimates of discharge counts, ALOS, charges, and mortality are reduced for the older age groups because of the demographics of patients in long-term acute care hospitals.

Sample Design for 2012 NIS

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 the nine U.S. census divisions (the four census regions were used prior to the 2012 NIS).

Prior to 2012, the NIS was 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 included all hospitals in the SID, and thus was limited by the availability of inpatient data from the data sources currently participating in HCUP. Starting with the 2012 NIS, a systematic sampling design was used to construct the database. Rather than first drawing a sample of hospitals and then keeping all discharges from that sample, in the 2012 NIS redesign a ***sample of discharges*** was drawn from ***all hospitals*** in the hospital frame. The new systematic sample is a self-weighted sample design similar to simple random sampling, but it is more efficient. It ensures that the sample is representative of the population on the following critical factors:

² See the AHA “community hospital designation” at <http://www.ahadataviewer.com/glossary>.

- hospital – unidentified
- census division of hospital
- hospital ownership
- urban-rural location of hospital
- hospital teaching status
- number of beds in the hospital
- diagnosis-related group (DRG) for the hospital stay
- admission month of the hospital stay

Weighted Estimates

To facilitate the production of national estimates, 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 reports on *Design of the Nationwide Inpatient Sample* found on the [NIS Related Reports](#) page on the HCUP-US Web site. Detailed information on the design of the NIS from 2006-2011 is available in the NIS Introduction for each year on the [NIS Database Documentation - Archive](#) page on the HCUP-US Web site.

Trends

The NIS is available yearly, beginning with 1988, allowing analysis of trends over time. Analyses of time trends are recommended from 1993 forward. For trends analysis using NIS data 2011 and earlier, revised weights should be used to make estimates comparable to the new design beginning with 2012 data. Refer to [1993-2011 NIS Trend Weights Files](#) and the report, [Using the HCUP Nationwide Inpatient Sample to Estimate Trends](#),³ available on the HCUP User Support (HCUP-US) Web site, for details.

NIS Data Sources, Hospitals, and Inpatient Stays

The NIS is sampled from the [State Inpatient Databases \(SID\)](#), all inpatient data that are currently contributed to HCUP. As displayed in [Figure 2](#) in [Appendix I](#), the 2012 NIS sampling frame covers more than 95 percent of the U.S. population; and as shown in [Table 6](#) in [Appendix I](#), it includes more than 94 percent of discharges from U.S. community hospitals. Weights are provided to calculate national estimates. [Table 2](#) in [Appendix I](#) lists the statewide data organizations participating in the NIS, and [Table 3](#) in [Appendix I](#) contains a summary of NIS data sources, hospitals, and discharges, by year.

Partner Restrictions

Some HCUP Partners 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](#).

³ As of June, 2014, this report had not yet been updated for the new 2012 NIS design. However, the methods described in the report are still valid.

Contents of NIS

Each release of the NIS includes:

- Data in fixed-width ASCII format on DVD or CD-ROM (prior to 2009).
- Over 7 million inpatient records per year.
- Discharge-level weights to calculate national estimates for discharges.
- NIS Documentation and tools – including file specifications, programs for loading the ASCII data into SAS and SPSS, and value labels. Beginning with 2005, code is also provided for loading the NIS ASCII files into Stata.

The NIS is distributed as fixed-width ASCII formatted data files compressed with SecureZIP® from PKWARE. **Beginning with the 2010 NIS, the files are encrypted.** Previously it was distributed on two CD-ROMs, but beginning with the 2009 NIS, it is distributed on a single DVD. The 2012 NIS includes the following compressed files:

Inpatient Core File: This inpatient discharge-level file contains a sample of hospital discharge records from 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*. Prior to the 2012 NIS, the HCUP hospital identifier (HOSPID) provided the linkage between the NIS Inpatient Core files and the Hospital Weights file. **Beginning with the 2012 NIS, the NIS hospital number (HOSP_NIS) provides the linkage between the NIS Inpatient Core files and the Hospital Weights file. The HOSP_NIS values are reassigned each year, so they cannot be used to link hospitals across years.** 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 two 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*. Prior to the 2012 NIS, the HCUP unique record identifier (KEY) provided the linkage between the Core files and the Disease Severity Measures file. **Beginning with the 2012 NIS, the unique NIS record number (KEY_NIS) 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 derived from AHRQ software tools based on the ICD-9-CM diagnostic and procedure information in the HCUP databases. The unit of observation is an *inpatient stay record*. Prior to the 2012 NIS, the HCUP unique record identifier (KEY) provided the linkage between the Core file and the Diagnosis and Procedure Groups file. **Beginning with the 2012 NIS, the unique NIS record number (KEY_NIS) provides the linkage between the Core files 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](#) Web site, NIS purchasers can access complete file documentation, including data element notes, file layouts, summary statistics, and related technical reports. Similarly, purchasers can also download SAS, SPSS, and Stata load programs from this Web site. Available online documentation and supporting files are detailed in [Appendix I](#), [Table 4](#).

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 Web site (<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 at least 15 gigabytes of space available
- A third-party zip utility such as ZIP Reader, 7-Zip, or WinZip®
- 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 a third-party zip utility such as ZIP Reader, WinZip, or 7-Zip. Beginning with the 2010 NIS, you will be prompted to enter the encryption password (sent separately by email) to decrypt each file.

NOTE: Attempts to unzip encrypted files using Windows will produce an error message similar to the following: "Compressed (Zipped) Folders Error. The password you typed is not correct. Try typing it again." The solution is to use a third-party zip utility such as ZIP Reader, 7-Zip, or WinZip rather than the built-in Windows Explorer function to open the archive.

ZIP Reader may be downloaded for free at <http://comm.pkware.com/download-zipreader-business.html>.

7-Zip can be downloaded for free at <http://www.7-zip.org/>.

Evaluation versions of WinZip may be downloaded from the WinZip Web site 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 Web site (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 2012 NIS Core file is linked to “Core File” under “2012 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

Year-specific NIS documentation files on the HCUP-US Web site (<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 4](#) in [Appendix I](#) details both the NIS related reports and the comprehensive NIS database documentation available on HCUP-US.

HCUP On-Line Tutorials⁴

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 national (nationwide) databases.

⁴ As of June, 2014, the HCUP Online Tutorials had not yet been updated for the new 2012 NIS design. However, the same statistical techniques should be used to calculate standard errors and confidence intervals. There is one change in example programs: HOSPID (the encrypted hospital identifier) should be replaced by HOSP_NIS.

The [Producing National HCUP Estimates](#) tutorial is designed to help users understand how the three national (nationwide) databases – the NIS, Nationwide Emergency Department Sample (NEDS), and Kids' Inpatient Database (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 national (nationwide) databases.

The [HCUP Multi-year Analysis](#) tutorial presents solutions that may be necessary when conducting analyses that span multiple years of HCUP data.

New tutorials are added periodically and existing tutorials are updated when necessary. The Online Tutorial Series is located on the HCUP-US Web site 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 Web site (<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 Web site (<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

- **In order to produce national estimates, you MUST use discharge weights.** Use the discharge weight (DISCWT⁵) to project discharges in the NIS Core files to the discharges from all U.S. community, non-rehabilitation hospitals. (For trends analysis using NIS data 2011 and earlier, see the next section of this report regarding trend weights.)

⁵ Prior to 1998, the discharge weight was named DISCWT_U. For 2000 only, use DISCWT to create national estimates for all analyses except those that involve total charges; and use DISCWTCHARGE to create national estimates of total charges.

- 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](#)⁶ on the [HCUP-US Web site](#).
- When creating national estimates, it is recommended to check your estimates against other data sources, if available.
- 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.
- For the 1998 redesign, rehabilitation hospitals—although classified as community hospitals by the AHA—were excluded from the NIS universe because (1) the State data did not always include discharges from those hospitals, and (2) patient characteristics and outcomes differ from short-term acute care hospitals. Likewise, for the 2012 redesign we also excluded long term acute care (LTAC) hospitals from the NIS universe for the same reasons.
- As a result of the changes implemented in the 2012 redesign, users should expect one-time disruptions to historical trends for counts, rates, and means estimated from the NIS, beginning with data year 2012. For 2012 we expect overall trends in discharge counts to decline by about 4.3 percent, overall trends in average length-of-stay to decline by about 1.5 percent, overall trends in total charges to decline by about 0.5 percent, and overall trends in hospital mortality to decline by about 2.0 percent. For a detailed description of the effects of the design changes on sample estimates, please see the NIS Redesign Report Executive Summary in [Appendix V](#) or the full [NIS Redesign Report](#) available on the HCUP User Support (HCUP-US) Web site.
- In order to facilitate analysis of trends using multiple years of NIS data, AHRQ developed new discharge trend weights for the 1993-2011 HCUP NIS. These weights were calculated in the same way as the weights for the redesigned 2012 NIS, and are designed to be used instead of the original NIS discharge weights for trends analysis. The new discharge trend weights replace the earlier NIS Trend Weights that were developed for the 1988-1997 NIS

⁶ As of June, 2014, this report had not yet been updated for the new 2012 NIS design. However, the same statistical techniques should be used to calculate standard errors and confidence intervals. There is one change in example programs: HOSPID (the encrypted hospital identifier) should be replaced by HOSP_NIS.

following the 1998 NIS redesign. The new trend weights are available for download under [1993-2011 NIS Trend Weights Files](#) from the [NIS Database Documentation](#) page on the HCUP-US Web site. The report [Using the HCUP Nationwide Inpatient Sample to Estimate Trends](#),⁷ available on the HCUP-US Web site under [Methods Series](#), includes recommendations for trends analysis.

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 Web site. 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 hospital. 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 some hospitals; thus, national estimates using RACE should be interpreted and reported with caveats.
- 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. For example, the most reliable way to identify ED admissions in the HCUP databases is to use the data element HCUP_ED, which considers all possible evidence of ED services. Unfortunately, this information is not always complete.

Hospital-Level Data Elements

- Beginning with the 2012 NIS, specific hospital identifiers (the AHA identifier), are no longer available. This means that 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 all hospitals in the NIS that allow you to study certain hospital characteristics including ownership/control, teaching status, rural/urban location, bed size, and the nine census divisions (as well as the four regions).

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 across hospitals. Some hospitals provide as many as 31 diagnoses and procedures or more, while others provide as few as 9 diagnoses and 6 procedures. Because very few cases have more than 25 diagnoses or

⁷ As of June, 2014, this report had not yet been updated for the new 2012 NIS design. However, the methods described in the report are still valid.

more than 15 procedures, the diagnosis and procedure vectors were truncated when necessary to save space in the NIS data files. Two data elements are provided which tell you exactly how many diagnoses and procedures were on the original records (NDX and NPR). See the notes on [diagnoses](#) and [procedures](#) on the HCUP-US Web site.

- The collection and reporting of external cause of injury (E codes) also varies across hospitals depending on the presence of State laws or mandates for the collection of E codes. 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.

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 about 5% of discharges in the 2012 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.

There are several techniques available to help assess and overcome this missing data bias.⁸ Descriptions of such data preparation and adjustment are outside the scope of this report; however, it is recommended that researchers evaluate and adjust for missing data, if necessary.

Variance Calculations

It may be important for researchers to calculate a measure of precision for some estimates based on the NIS sample data. Variance estimates must take into account both the sampling design and the form of the statistic. A stratified systematic sample of discharges was drawn from a sorted list of discharges comprising *all* discharges in the sampling frame. **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](#),⁹ available on the HCUP-US Web site.

If discharges inside the sampling frame are similar to discharges outside the frame, the sample of discharges can be treated as if they were randomly selected from the entire universe of discharges within each stratum. Although the NIS is no longer a cluster sample, discharges are still clustered by hospitals, so hospitals (HOSP_NIS) should be treated as clusters when calculating statistics. Standard formulas for a stratified, single-stage cluster sample without replacement should still 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-

⁸ See, for example, van Buuren, S. (2012). *Flexible Imputation of Missing Data*. CRC Press, Boca Raton, FL.

⁹ As of June, 2014, this report had not yet been updated for the new 2012 NIS design. However, the same statistical techniques should be used to calculate standard errors and confidence intervals. The one change for example programs is that HOSPID should be replaced by HOSP_NIS.

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, national estimates of outcomes. According to finite-sample theory, the intent of the estimation process is to obtain estimates that are precise representations of the national 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.

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 national population in a given year is only a snapshot in time of the possible interrelationships among hospital 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 fixed at that point in time, and because the estimate is for a fixed 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 national discharge population in a particular year is viewed as a random sample that resulted from a specific set of random events drawn from an underlying superpopulation of similar random events that might have occurred. For example, the outcome of a particular hospitalization might differ depending admission timing, hospital staffing during the stay, and so on. Different methods are used for calculating variances under the two sample theories. The choice of an appropriate method for calculating variances for nationwide estimates depends on the type of measure and the intent of the estimation process.

Computer Software for Variance Calculations

The discharge weights 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 national population statistics.

In most cases, computer programs are readily available to perform these calculations. Several statistical programming packages allow weighted analyses.¹⁰ 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,

¹⁰ 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.

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 Web site 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 Web site: <http://www.hcp.med.harvard.edu/statistics/survey-soft/>.

The NIS database includes a Hospital Weights file with data elements required by these programs to calculate finite population statistics. The file includes hospital identifiers (Primary Sampling Units or PSUs), stratification data elements, 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.

Longitudinal Analyses

Beginning with the 2012 data, the NIS includes a sample of discharges from all HCUP hospitals. However, the NIS hospital number (HOSP_NIS) values are reassigned each year, so they cannot be used to link hospitals across years. Thus longitudinal analyses of specific hospitals is not supported by the NIS.

SAMPLING PROCEDURE

The NIS Hospital Universe

Each year, the AHA's Health Forum administers the AHA Annual Survey of Hospitals. The purpose of the survey is to collect utilization, financial, service, and personnel information on each of the nation's hospitals. The survey's overall response rate averages approximately 85 percent each year, which is high for a voluntary survey given its length and the size of the universe (about 6,000 hospitals). For hospitals that do not respond, the AHA imputes items based on prior-year information, so that data are available for all hospitals in the universe.

The hospital universe is defined by all hospitals that were open during any part of the calendar year and were designated as community hospitals in the AHA Annual Survey. For purposes of the NIS, the definition of a *community hospital* is that used by the AHA: "all nonfederal short-term general and other specialty hospitals, excluding hospital units of institutions." Consequently, Veterans Affairs hospitals and other Federal hospitals are excluded. Beginning with the 1998 redesign, rehabilitation hospitals are excluded. Beginning with the 2012 redesign, long-term acute care hospitals are also excluded.

Long-term acute care hospitals are classified as community hospitals by the AHA if they have an average length-of-stay (ALOS) less than 30 days. However, long-term acute care hospital data was not uniformly available from all States participating in HCUP, and ALOS data from these facilities was over 25 days (unlike other community hospitals with an ALOS of about 4.5 days). Thus, long term acute care facilities were eliminated from the 2012 NIS.

Prior to the 2012 NIS, NIS sample weights were calculated by dividing the number of universe discharges by the number of sampled discharges within each hospital stratum. The number of universe discharges had been estimated using data from the AHA annual hospital survey. In particular, the total number of discharges in the universe was estimated by the sum of births and admissions contained in the AHA annual survey for all hospitals in the universe.

Given that HCUP Partners supply more than 95 percent of discharges nationwide, beginning with the 2012 NIS, we estimate the universe count of discharges within each stratum, using the actual count of discharges contained in HCUP data. We use the AHA counts only for hospitals in the universe that do not appear in HCUP data coming from the statewide data organizations.

This option was not considered for the previous 1998 redesign because HCUP data included a much smaller percentage of discharges in the United States, and the differences between HCUP counts and AHA counts would tend to adversely affect trends as the mix of HCUP States changed from year to year. In 2011, for hospitals in both the AHA and the SID, in 43 of 46 States, the AHA survey data estimated State discharge totals that were between 1 percent and 17 percent higher than the observed SID discharge totals. Overall, the AHA survey estimated about a 4 percent higher count of discharges than the observed SID count.

In the 2012 redesign, a logical corollary of switching from AHA discharge estimates to SID discharge counts was to distinguish unique hospitals using the SID hospital identifiers rather than the AHA hospital identifiers. For the vast majority of hospitals, the SID hospital identifiers are in one-to-one correspondence with the AHA hospital identifiers. However, about 10 percent of the AHA identifiers actually correspond to two or more hospitals in the SID that have common ownership within a hospital system. For these “combined” AHA identifiers, the number of estimated discharges and the number of hospital beds in the AHA data reflect the sum of estimated discharges and the sum of beds, respectively, from the constituent hospitals. As a result, these combined hospitals could have been allocated to the wrong bed size stratum in the sample design. Also, the between-hospital variance was combined with the within-hospital variance for these combined hospitals. Therefore, use of the SID hospital identifiers in the 2012 NIS disaggregates the previously combined hospitals in many States, which is likely to improve the classification of hospitals and improve variance estimates.

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](#).¹¹ For a list of all data sources, refer to [Table 2](#) in [Appendix I](#).

Stratification Data Elements

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 data elements, the exclusion of

¹¹ As of June 2014, this report had not been updated for the new NIS design; however the methods described are still valid.

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 Web site at <http://www.hcup-us.ahrq.gov/db/nation/nis/nisrelatedreports.jsp>. (A description of the sampling procedures and definitions of strata data elements 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 Web site.)

Again in 2012, the NIS team evaluated and revised the sampling strategy for 2012 and subsequent data years, in order to best represent the U.S. See [Table 1](#) in [Appendix I](#), The National Inpatient Sample (NIS) Design Changes, for a summary of design changes. For a detailed description of the NIS redesign, please see the NIS Redesign Report Executive Summary in [Appendix V](#) or the full [NIS Redesign Report](#) available on the HCUP User Support (HCUP-US) Web site.

Prior to 2012, the NIS sampling strata were defined based on five hospital characteristics contained in the AHA hospital files. Beginning with the 2012 NIS, the only hospital-level stratification factor that changes is that we stratify hospitals by census *division* rather than census *region*;¹² and the stratification data elements were defined as follows:

1. *Census Division – New England, Middle Atlantic, East North Central, West North Central, South Atlantic, South Central, Mountain, and Pacific.* This is an important stratification data element because practice patterns have been shown to vary substantially by region. For example, lengths of stay tend to be longer in East Coast hospitals than in West Coast hospitals. [Figure 1](#) in [Appendix I](#) shows the NIS States by census division.
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. We stratified hospitals as public, voluntary, and proprietary. When necessary, we combined strata so that a minimum of two hospitals are included in each stratum.
3. *Location – urban or rural. Government payment policies for hospital services often differ according to this designation. Also, rural hospitals are generally smaller and offer fewer services than urban hospitals. We categorized hospitals with a CBSA type of Metropolitan or Division as urban, while we designated hospitals with a CBSA type of Micropolitan or Rural as rural.*
4. *Teaching Status – teaching or non-teaching.* The missions of teaching hospitals differ from non-teaching hospitals. In addition, financial considerations differ between these two hospital groups. Currently, the Medicare Diagnosis Related Group (DRG) payments are uniformly higher to *teaching hospitals*. We considered a hospital to be a teaching hospital if it met any one of the following three criteria: (See [Appendix IV](#) for details.)
 - Residency training approval by the Accreditation Council for Graduate Medical Education (ACGME)

¹² However, researchers will still be able to make estimates for census regions by aggregating census divisions.

- Membership in the Council of Teaching Hospitals (COTH)
 - A ratio of full-time equivalent interns and residents to beds of .25 or higher
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 5](#) 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 2012, 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 of the sample, implicit stratification data elements included de-identified hospital number, Diagnosis Related Group (DRG) and admission month. The discharges were sorted according to these data elements prior to systematic random sampling.

Design Considerations

Prior to 2012, the NIS was a stratified probability sample of hospitals in the frame, with sampling probabilities proportional to the number of U.S. community hospitals in each stratum: sampling probabilities were 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.

Starting with the 2012 NIS, a systematic sampling design is used to construct the database. Rather than first drawing a sample of hospitals and then keeping all discharges from that sample, in the 2012 NIS redesign a sample of discharges was drawn from *all* hospitals in the hospital frame. Both designs selected approximately 20 percent of the target universe of discharges from United States community hospitals, excluding rehabilitation and long-term acute care hospitals.

The new systematic sample is a self-weighted sample design similar to simple random sampling, but it is more efficient. It ensures that the sample is representative of the population on the following critical factors: hospital factors (hospital – unidentified, census division,

ownership, urban-rural location, teaching status, number of beds) and patient factors (diagnosis-related group, admission month). Within each stratum all discharges are sorted in the following order on patient-level “control” variables: encrypted hospital ID, DRG, admission month, and a random number.

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, especially for specific patient populations (e.g. organ transplant recipients).

The [NIS Redesign Report](#) assessed the accuracy of NIS estimates and considered alternative stratified sampling allocation schemes. However, systematic sampling design was preferred for several reasons:

- It significantly reduced the margin of error for estimates and delivered improved accuracy more consistently across diagnosis-related groups (DRGs) compared with the other alternatives.
- It is easy for analysts to use because of its self-weighting design.
- It is straightforward to estimate accurate confidence intervals using standard statistical software.
- There is little researcher demand for 100 percent of discharges from a sample of hospitals, and researchers who require complete discharge data from every hospital can use the SID data.
- Its implementation on an annual basis will be efficient and can be accomplished using readily available software and accepted methods.

Overview of the Sampling Procedure

The strata for the 2012 NIS systematic sampling design are the same as those for the previous NIS sample design except that the four census regions are replaced by the nine census divisions—New England, Middle Atlantic, East North Central, West North Central, South Atlantic, South Central, Mountain, and Pacific. Within each stratum, dischargers are sorted by re-identified hospital number. Then, within each hospital, discharges are sorted by their DRG and their admission month. This sorting ensures that the NIS sample will be representative on these factors.

Next, within each stratum, a number of discharges proportionate to the number of discharges in the universe are selected systematically from the sorted list. For example, if the sampling frame was equal to the universe and 20 percent of the universe was required, then every fifth discharge would be selected from the sorted list of discharges, beginning with a randomly selected start at discharge number 1, 2, 3, 4, or 5 on the list.

To ensure a self-weighted sample that has 20 percent of the universe within each stratum represented, sampling rates would vary within each stratum, depending on the proportion of the population of discharges covered by the discharges in the sampling frame. Thus, the sampling rate would not always be 20 percent within each stratum. For strata that were missing more discharges, the sampling rate would be higher to ensure that the number of sampled discharges would equal 20 percent of the universe.

WEIGHTS

To obtain nationwide estimates, we developed discharge weights to extrapolate NIS sample discharges to the discharge universe. NIS discharge weights are calculated by dividing the number of universe discharges by the number of sampled discharges within each NIS stratum. Historically, the number of universe discharges had been estimated using data from the AHA annual hospital survey. In particular, the total number of discharges in the universe was estimated by the sum of births and admissions contained in the AHA annual survey for all hospitals in the universe.

Given that HCUP Partners supply more than 95 percent of discharges nationwide, beginning with the 2012 NIS, we now estimate the universe count of discharges within each stratum using the actual count of discharges contained in HCUP data. The only exceptions are for strata with HCUP hospitals that, according to the AHA files, were open for the entire year but contributed less than a full year of data to HCUP. For those hospitals, we *adjust* the number of observed discharges by a factor of $12 \div M$, where M is the number of months for which the hospital contributed discharges to HCUP. For example, when a hospital contributed only six months of discharge data to HCUP, the *adjusted* number of discharges is double the observed number.

For non-HCUP hospitals in the universe we use adjusted AHA discharge estimates. To adjust the AHA discharge estimates we multiplied them by the overall ratio of HCUP discharges to estimated AHA discharges for HCUP hospitals in the census division.

The discharge weights are constant for all discharges within a stratum, where the stratum is defined by hospital characteristics: census division, rural/urban location, bed size, teaching status, and ownership. The previous design provided discharge weights that reflected the universe of discharges in each of the *four census regions*. The 2012 NIS design provides discharge weights that reflect the universe of discharges in each of the nine *census divisions*.

Each discharge weight is essentially equal to the number of target universe discharges that each sampled discharge represents in its stratum. Discharge weights to the universe were calculated as follows: Within stratum *s*, each NIS sample discharge's universe weight was calculated as:

$$DW_s(\text{universe}) = DN_s(\text{universe}) \div DN_s(\text{sample})$$

where $DW_s(\text{universe})$ was the discharge weight; $DN_s(\text{universe})$ represented the number of discharges from community hospitals in the universe within stratum *s*; and $DN_s(\text{sample})$ was the number of discharges selected for the NIS. Thus, each discharge's weight (DISCWT) is equal to the number of universe discharges it represents in stratum *s* during that year.¹³ Because 20% of the universe discharges in each stratum were sampled, the discharge weights are near five.

Prior to the 2012 NIS redesign, the NIS included weights to project NIS hospitals to the number of hospital in the target universe. However, with the 2012 NIS redesign the hospital weights are discontinued because the NIS is now a sample of discharges from all available HCUP SID community hospitals, excluding rehabilitation and long-term acute care hospitals.

¹³ Although discharge characteristics (DRG and admission month) are implicit stratifiers for sampling, they do not play a role in weighting.

APPENDIX I: TABLES AND FIGURES

Table 1: The 2012 National Inpatient Sample (NIS) Design Changes

Feature	Previous Design (1998-2011)	New 2012 Design
Universe	Included long-term acute care hospitals	Removed long-term acute care hospitals
	Discharge estimates based on AHA admissions plus births	Discharge estimates based on SID discharges when available (for about 90% of all hospitals); otherwise, based on adjusted AHA counts
	Hospitals defined based on AHA IDs	Hospitals defined based on State-supplied hospital identifiers for HCUP states
Sample design	Sample hospitals and then retain all discharges from each sampled hospital	Systematic sample of discharges from all frame hospitals
	Stratified by: <ul style="list-style-type: none"> • hospital census <i>region</i>,¹⁴ • ownership, • urban/rural location, • teaching status, and • number of beds (bed size categories) 	Stratified by <ul style="list-style-type: none"> • hospital census <i>division</i>,¹⁵ • ownership, • urban/rural location, • teaching status, and • number of beds (bed size categories)
	Sorted by three-digit hospital ZIP Code within strata before sampling	Sorted by hospital and by DRG and admission month within strata before sampling
	Sample without self-weighting requires weights for all estimates	Self-weighting sample requires weights for estimating totals, but not for means and rates
Data elements	Includes State and hospital identifiers and data elements with State-specific coding	<ul style="list-style-type: none"> • Drops State identifiers and data elements that were not available uniformly across the States, such as hospital identifiers, secondary payer, and data elements with State-specific coding • Drop hospital weights • Retain certain high value State-specific data elements (See Appendix B)

Abbreviations: AHA, American Hospital Association; DRG, diagnosis-related group; ID, identification numbers; SID, State Inpatient Databases

[Return to Introduction](#)

¹⁴ Census region: Northeast, Midwest, South, West

¹⁵ Census division: New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific

Table 2: States Participating in the 2012 NIS

State	Data Organization
AK	Alaska State Hospital and Nursing Home Association
AR	Arkansas Department of Health
AZ	Arizona Department of Health Services
CA	Office of Statewide Health Planning & Development
CO	Colorado Hospital Association
CT	Connecticut Hospital Association
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
LA	Louisiana Department of Health and Hospitals
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
MT	MHA - An Association of Montana Health Care Providers
NC	North Carolina Department of Health and Human Services
ND	North Dakota (data provided by the Minnesota Hospital Association)
NE	Nebraska Hospital Association
NJ	New Jersey Department of Health

State	Data Organization
NM	New Mexico Department of Health
NV	Nevada Department of Health and Human Services
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
PA	Pennsylvania Health Care Cost Containment Council
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 Services
WV	West Virginia Health Care Authority
WY	Wyoming Hospital Association

[Return to Introduction](#)

Table 3: Summary of NIS States, Hospitals, and Inpatient Stays, 1988-2012

Year	States	Number of States	Number of Hospitals	Number of Discharges in the NIS, Unweighted	Number of Discharges in the NIS, Weighted	Number of Discharges in the NIS, Weighted with Trend Weight
1988	CA CO FL IA IL MA NJ WA	8	759	5,265,756	35,171,448	--
1989	AZ CA CO FL IA IL MA NJ PA WA WI	11	882	6,110,064	35,104,645	--
1990	AZ CA CO FL IA IL MA NJ PA WA WI	11	871	6,268,515	35,215,397	--
1991	AZ CA CO FL IA IL MA NJ PA WA WI	11	859	6,156,188	35,036,492	--
1992	AZ CA CO FL IA IL MA NJ PA WA WI	11	856	6,195,744	35,011,385	--
1993	AZ CA CO CT FL IA IL KS MA MD NJ NY OR PA SC WA WI	17	913	6,538,976	34,715,985	33,736,753
1994	AZ CA CO CT FL IA IL KS MA MD NJ NY OR PA SC WA WI	17	904	6,385,011	34,622,203	33,149,768
1995	AZ CA CO CT FL IA IL KS MA MD MO NJ NY OR PA SC TN WA WI	19	938	6,714,935	34,791,998	33,647,121
1996	AZ CA CO CT FL IA IL KS MA MD MO NJ NY OR PA SC TN WA WI	19	906	6,542,069	34,874,386	33,386,097
1997	AZ CA CO CT FL GA HI IA IL KS MA MD MO NJ NY OR PA SC TN UT WA WI	22	1,012	7,148,420	35,408,207	33,232,257
1998	AZ CA CO CT FL GA HI IA IL KS MA MD MO NJ NY OR PA SC TN UT WA WI	22	984	6,827,350	34,874,001	33,923,632
1999	AZ CA CO CT FL GA HI IA IL KS MA MD ME MO NJ NY OR PA SC TN UT VA WA WI	24	984	7,198,929	35,467,673	34,440,994
2000	AZ CA CO CT FL GA HI IA IL KS KY MA MD ME MO NC NJ NY OR PA SC TN TX UT VA WA WI WV	28	994	7,450,992	36,417,565	35,300,425
2001	AZ CA CO CT FL GA HI IA IL KS KY MA MD ME MI MN MO NC NE NJ NY OR PA RI SC TN TX UT VA VT WA WI WV	33	986	7,452,727	37,187,641	36,093,550
2002	CA CO CT FL GA HI IA IL KS KY MA MD ME MI MN MO NC NE NJ NV NY OH OR PA RI SC SD TN TX UT VA VT WA WI WV	35	995	7,853,982	37,804,021	36,523,831

Year	States	Number of States	Number of Hospitals	Number of Discharges in the NIS, Unweighted	Number of Discharges in the NIS, Weighted	Number of Discharges in the NIS, Weighted with Trend Weight
2003	AZ CA CO CT FL GA HI IA IL IN KS KY MA MD MI MN MO NC NE NH NJ NV NY OH OR PA RI SC SD TN TX UT VA VT WA WI WV	37	994	7,977,728	38,220,659	37,074,605
2004	AR AZ CA CO CT FL GA HI IA IL IN KS KY MA MD MI MN MO NC NE NH NJ NV NY OH OR RI SC SD TN TX UT VA VT WA WI WV	37	1,004	8,004,571	38,661,786	37,496,978
2005	AR AZ CA CO CT FL GA HI IA IL IN KS KY MA MD MI MN MO NC NE NH NJ NV NY OH OK OR RI SC SD TN TX UT VT WA WI WV	37	1,054	7,995,048	39,163,834	37,843,039
2006	AR AZ CA CO CT FL GA HI IA IL IN KS KY MA MD MI MN MO NC NE NH NJ NV NY OH OK OR RI SC SD TN TX UT VA VT WA WI WV	38	1,045	8,074,825	39,450,216	38,076,556
2007	AR AZ CA CO CT FL GA HI IA IL IN KS KY MA MD ME MI MN MO NC NE NH NJ NV NY OH OK OR RI SC SD TN TX UT VA VT WA WI WV WY	40	1,044	8,043,415	39,541,948	38,155,908
2008	AR AZ CA CO CT FL GA HI IA IL IN KS KY LA MA MD ME MI MN MO NC NE NH NJ NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY	42	1,056	8,158,381	39,885,120	38,210,889
2009	AR AZ CA CO CT FL GA HI IA IL IN KS KY LA MA MD ME MI MN MO MT NC NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY	44	1,050	7,810,762	39,434,956	37,734,584
2010	AK AR AZ CA CO CT FL GA HI IA IL IN KS KY LA MA MD ME MI MN MO MS MT NC NE NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY	45	1,051	7,800,441	39,008,298	37,352,013
2011	AK AR AZ CA CO CT FL GA HI IA IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY	46	1,049	8,023,590	38,590,733	36,962,415

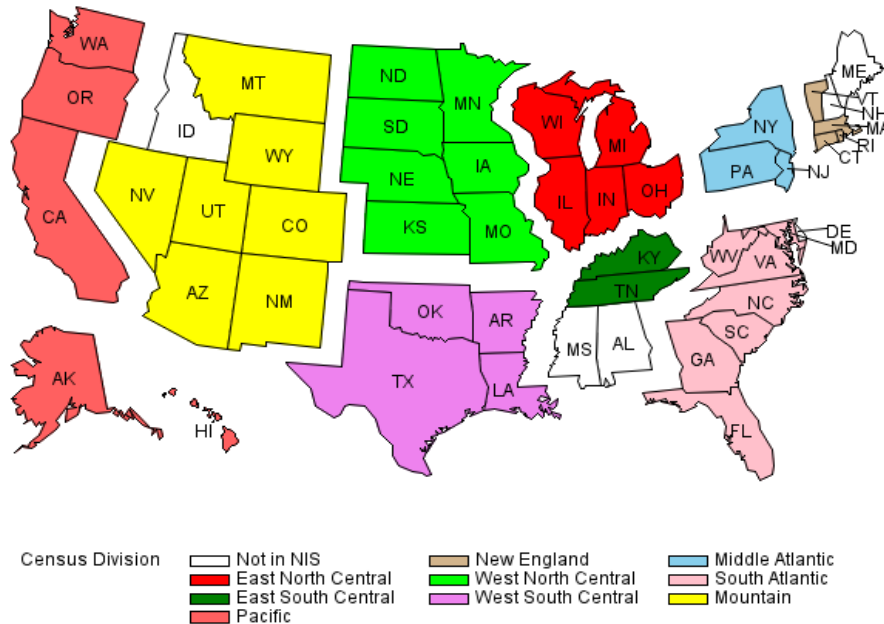
Year	States	Number of States	Number of Hospitals	Number of Discharges in the NIS, Unweighted	Number of Discharges in the NIS, Weighted	Number of Discharges in the NIS, Weighted with Trend Weight
2012	AK AR AZ CA CO CT FL GA HI IA IL IN KS KY LA MA MD MI MN MO MT NC ND NE NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY	44	4,378	7,296,968	36,484,846	36,484,846

[Return to Introduction](#)

Table 4: NIS Related Reports and Database Documentation Available on HCUP-US

<p>Restrictions on the Use of the NIS</p> <ul style="list-style-type: none"> • Data Use Agreement for the NIS 	<p>Load Programs Programs to load the ASCII data files into statistical software:</p> <ul style="list-style-type: none"> • SAS • SPSS • Stata
<p>Description of the NIS Files</p> <ul style="list-style-type: none"> • Introduction to the NIS, 2012 – <i>this document</i> • 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, NIS Data Elements, and State-Specific Restrictions (<i>included in this document beginning in 2006</i>) – identifies the NIS data sources and restrictions on sampling and the release of data elements 	<p>HCUP Tools: Labels and Formats</p> <ul style="list-style-type: none"> • 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 value labels • NIS ICD-9-CM formats to label ICD-9-CM diagnoses and procedures • NIS Severity formats to label severity data elements
<p>Availability of Data Elements</p> <ul style="list-style-type: none"> • Availability of NIS data elements from 1988-2012 	<p>NIS Related Reports Links to HCUP-US page with various NIS related reports such as the following:</p> <ul style="list-style-type: none"> • Design of the Nationwide Inpatient Sample for 1988 to 2005 (<i>included in this document beginning in 2006</i>) • Changes in NIS Sampling and Weighting Strategy for 1998 • Calculating National Inpatient Sample Variances • Using the HCUP National Inpatient Sample to Estimate Trends • NIS Comparison Reports (available for years in which the NIS sample changed) • HCUP Data Quality Reports for 1988-2012 • HCUP E-Code Evaluation Report
<p>Description of Data Elements in the NIS</p> <ul style="list-style-type: none"> • 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 	<p>HCUP Supplemental Files</p> <ul style="list-style-type: none"> • Cost-to-Charge Ratio files • Hospital Market Structure (HMS) files • 1993-2002 NIS Supplemental Discharge-Level Files
<p>Known Data Issues</p> <ul style="list-style-type: none"> • Information on corrections to the NIS data sets • Link to 1993-2011 NIS Trend Weights Files 	<p>SAS File Information</p> <ul style="list-style-type: none"> • File Information for all States and years

Figure 1: NIS States, by Census Division¹⁶



All States, by U.S Census Bureau¹⁷ Region and Census Division¹⁸

- Region 1 ([Northeast](#))
 - Division 1 ([New England](#)) [Maine](#), [New Hampshire](#), [Vermont](#), [Massachusetts](#), [Rhode Island](#), [Connecticut](#)
 - Division 2 ([Mid-Atlantic](#)) [New York](#), [Pennsylvania](#), [New Jersey](#)
- Region 2 ([Midwest](#)) (Prior to June 1984, the Midwest Region was designated as the North Central Region.)
 - Division 3 ([East North Central](#)) [Wisconsin](#), [Michigan](#), [Illinois](#), [Indiana](#), [Ohio](#)
 - Division 4 ([West North Central](#)) [Missouri](#), [North Dakota](#), [South Dakota](#), [Nebraska](#), [Kansas](#), [Minnesota](#), [Iowa](#)
- Region 3 ([South](#))
 - Division 5 ([South Atlantic](#)) [Delaware](#), [Maryland](#), [District of Columbia](#), [Virginia](#), [West Virginia](#), [North Carolina](#), [South Carolina](#), [Georgia](#), [Florida](#)
 - Division 6 ([East South Central](#)) [Kentucky](#), [Tennessee](#), [Mississippi](#), [Alabama](#)
 - Division 7 ([West South Central](#)) [Oklahoma](#), [Texas](#), [Arkansas](#), [Louisiana](#)
- Region 4 ([West](#))
 - Division 8 ([Mountain](#)) [Idaho](#), [Montana](#), [Wyoming](#), [Nevada](#), [Utah](#), [Colorado](#), [Arizona](#), [New Mexico](#)
 - Division 9 ([Pacific](#)) [Alaska](#), [Washington](#), [Oregon](#), [California](#), [Hawaii](#)

[Return to Introduction](#)

¹⁶ New Hampshire participates in HCUP, but did not provide data in time for the 2010-2012 NIS. Maine and Mississippi participate in HCUP, but did not provide data in time for the 2012 NIS.

¹⁷ U.S. Census Bureau. Census Bureau Regions and Divisions with State FIPS Codes. http://www.census.gov/geo/maps-data/maps/pdfs/reference/us_regdiv.pdf. Accessed November 5, 2013.

¹⁸ States and areas in italics do not participate in HCUP.

Table 5: Hospital Size Categories (in Number of Beds), by Region

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

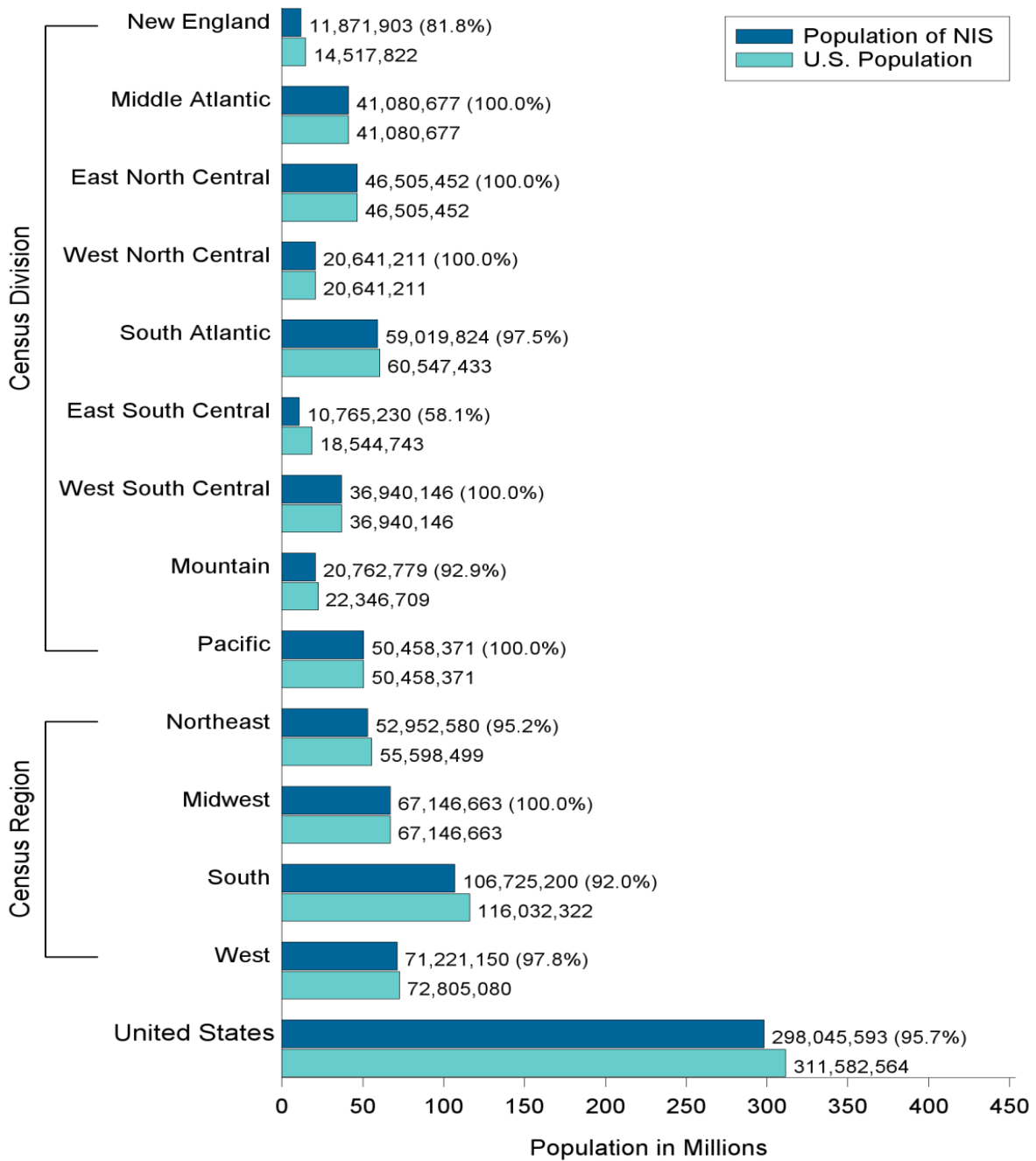
[Return to Introduction](#)

Table 6: Number of Hospitals and Discharges in 2012 Universe, Frame, and NIS, by Census Division

Number of Hospitals and Discharges in 2012 Universe, Frame, and NIS, by Census Division								
Census Region	Census Division	Universe		Frame		NIS		
		Hospitals	Discharges	Hospitals	Discharges	Hospitals	Discharges	Weighted Discharges
Northeast	New England	193	1,712,458	125	1,405,491	124	342,492	1,712,458
Northeast	Middle Atlantic	448	5,269,187	444	5,244,767	444	1,053,839	5,269,187
Northeast	Subtotal	641	6,981,645	569	6,650,258	568	1,396,331	6,981,645
Midwest	East North Central	752	5,732,724	732	5,708,444	731	1,146,546	5,732,724
Midwest	West North Central	693	2,505,496	640	2,469,711	639	501,101	2,505,496
Midwest	Subtotal	1,445	8,238,220	1,372	8,178,155	1,370	1,647,647	8,238,220
South	South Atlantic	725	7,315,085	706	7,054,085	706	1,463,015	7,315,085
South	East South Central	414	2,524,375	214	1,320,440	214	504,873	2,524,375
South	West South Central	785	4,273,641	666	4,149,181	666	854,729	4,273,641
South	Subtotal	1,924	14,113,101	1,586	12,523,706	1,586	2,822,617	14,113,101
West	Mountain	385	2,230,898	329	2,078,985	329	446,178	2,230,898
West	Pacific	539	4,920,982	527	4,879,220	525	984,195	4,920,982
West	Subtotal	924	7,151,880	856	6,958,205	854	1,430,373	7,151,880
Total	Total	4,934	36,484,846	4,383	34,310,324	4,378	7,296,968	36,484,846

[Return to Introduction](#)

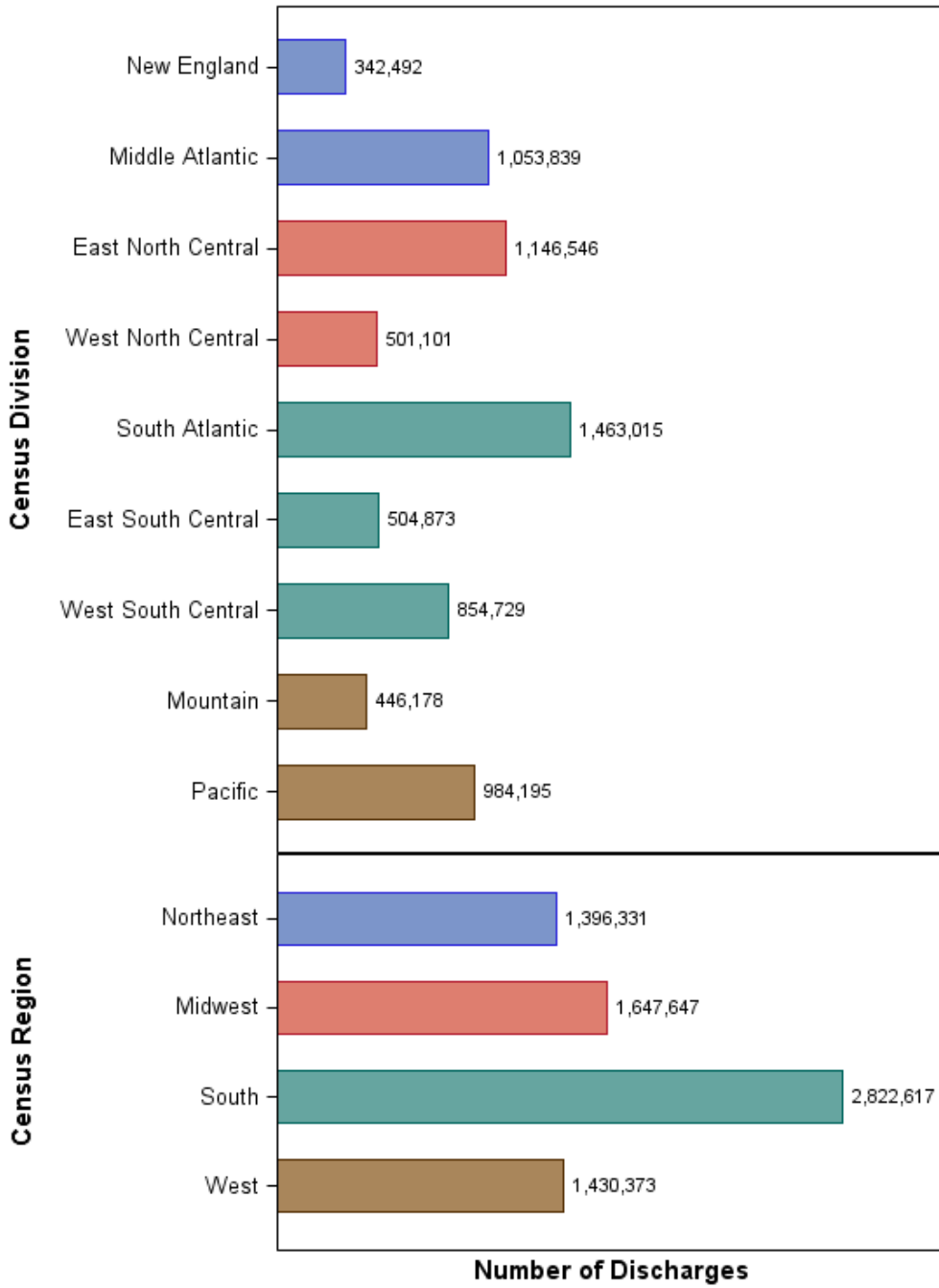
Figure 2: Percentage of U.S. Population Covered in the 2012 NIS by Census Division and Region, Calculated using the Estimated U.S. Population on July 1, 2012.¹⁹



[Return to Introduction](#)

¹⁹ Table 1. Annual Estimates of the Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2013 (NST-EST2013-01). Source: U.S. Census Bureau, Population Division. Release Date: December 2013.

Figure 3: Number of Discharges (in Thousands) in 2012 NIS by Census Division and Census Region



[Return to Introduction](#)

APPENDIX II: STATE-SPECIFIC RESTRICTIONS

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

- [Confidentiality of records](#)
 - Restricted release of age in years
- [Missing discharges](#).

Confidentiality of Records - Restricted Release of Age in Years

- Age (AGE) values greater than 90 are set to 90 for all NIS records to de-identify the data.
- At least one Partner required ages in years (AGE) to be set to the midpoints of age ranges.

Missing Discharges

- At least one Partner prohibits the release of discharge records for patients with HIV diagnoses.
- At least one Partner prohibits the release of behavioral health including chemical dependency care or psychiatric care discharges.
- At least one Partner prohibits the release of Alternate Level of Care (SNF / Swing Bed Skilled) discharges.
- At least one Partner prohibits the release of abortion discharges.

APPENDIX III: DATA ELEMENTS

Table 1: Data Elements in the 2012 NIS Inpatient Core Files

For prior years, refer to documentation on HCUP-US (e.g. the table of data element availability by years <https://www.hcup-us.ahrq.gov/db/nation/nis/NISvariables1988-2012forHCUP-US.pdf> or previous versions of the NIS Introduction).

Type of Data Element	HCUP Name	Coding Notes
Admission information		
Admission day	AWEEKEND	Admission on weekend: (0) admission on Monday-Friday, (1) admission on Saturday-Sunday
Admission month	AMONTH	Admission month coded from (1) January to (12) December
Transferred into hospital	TRAN_IN	Transfer In Indicator: (0) not a transfer, (1) transferred in from a different acute care hospital [ATYPE NE 4 & (ASOURCE=2 or POO=4)], (2) transferred in from another type of health facility [ATYPE NE 4 & (ASOURCE=3 or POO=5, 6)]
Indicator of emergency department service	HCUP_ED	Indicator that discharge record includes evidence of emergency department (ED) services: (0) Record does not meet any HCUP Emergency Department criteria, (1) Emergency Department revenue code on record, (2) Positive Emergency Department charge (when revenue center codes are not available), (3) Emergency Department CPT procedure code on record, (4) Admission source of ED, (5) State-defined ED record; no ED charges available
Admission type	ELECTIVE	Indicates elective admission: (1) elective, (0) non-elective admission
Patient demographic and location information		
Age at admission	AGE	Age in years coded 0-124 years
	AGE_NEONATE	Neonatal age (first 28 days after birth) indicator: (0) non-neonatal age (1) neonatal age
Sex of patient	FEMALE	Indicates gender for NIS beginning in 1998: (0) male, (1) female
Race of patient	RACE ²⁰	Race, uniform coding: (1) white, (2) black, (3) Hispanic, (4) Asian or Pacific Islander, (5) Native American, (6) other

²⁰ Race contains missing values on more than 5% of the records.

Type of Data Element	HCUP Name	Coding Notes
Location of patient's residence	PL_NCHS2006	Patient Location: NCHS Urban-Rural Code (V2006). This is a six-category urban-rural classification scheme for U.S. counties: (1) "Central" counties of metro areas of >=1 million population, (2) "Fringe" counties of metro areas of >=1 million population, (3) Counties in metro areas of 250,000-999,999 population, (4) Counties in metro areas of 50,000-249,999 population, (5) Micropolitan counties, (6) Not metropolitan or micropolitan counties
Median household income for patient's ZIP Code	ZIPINC_QRTL	Median household income quartiles for patient's ZIP Code. For 2008, the median income quartiles are defined as: (1) \$1 - \$38,999; (2) \$39,000 - \$47,999; (3) \$48,000 - 62,999; and (4) \$63,000 or more.

Payer information

Primary expected payer	PAY1	Expected primary payer, uniform: (1) Medicare, (2) Medicaid, (3) private including HMO, (4) self-pay, (5) no charge, (6) other
------------------------	------	--

Diagnosis and procedure information

ICD-9-CM diagnoses	DX1 – DX25	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. Beginning in 2009, the diagnosis array was increased from 15 to 25.
	NDX	Number of diagnoses coded on the original record
External causes of injury and poisoning	ECODE1 - ECODE4	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).
	NECODE	Number of external cause of injury codes on the original record. A maximum of 4 codes are retained on the NIS.
ICD-9-CM procedures	PR1 - PR15	Procedures, principal and secondary (ICD-9-CM)
	NPR	Number of procedures coded on the original record
	PRDAY1	Number of days from admission to principal procedure.
	PRDAY2 - PRDAY15	Number of days from admission to secondary procedures

DRG information

Diagnosis Related Group (DRG)	DRG	DRG in use on discharge date
	DRG_NoPOA	DRG in use on discharge date, calculated without Present On Admission (POA) indicators

Type of Data Element	HCUP Name	Coding Notes
	DRGVER	Grouper version in use on discharge date
	DRG24	DRG Version 24 (effective October 2006 - September 2007)
Major Diagnosis Category (MDC)	MDC	MDC in use on discharge date
	MDC_noPOA	MDC in use on discharge date, calculated without Present on Admission (POA) indicators
	MDC24	MDC Version 24 (effective October 2006 - September 2007)

Other data elements derived from ICD-9-CM codes

see also:

Table 3, Data Elements in the NIS Disease Severity Measures File and

Table 4, Data Elements in the NIS Diagnosis and Procedures Groups File

Clinical Classifications Software (CCS) category	DXCCS1 – DXCCS25	Clinical Classifications Software (CCS) category for all diagnoses for NIS beginning in 1998. Beginning in 2009, the diagnosis array was increased from 15 to 25.
	E_CCS1 - E_CCS4	CCS category for the external cause of injury and poisoning codes
	PRCCS1 - PRCCS15	CCS category for all procedures for NIS beginning in 1998
Number of chronic conditions	NCHRONIC	Count of chronic conditions in the diagnosis vector
Operating room procedure indicator	ORPROC	Major operating room procedure indicator for the record: (0) no major operating room procedure, (1) major operating room procedure
Neonatal/maternal flag	NEOMAT	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
Indicates in-hospital birth	HOSPBIRTH	Indicator that discharge record includes diagnosis of birth that occurred in the hospital: (0) Not an in-hospital birth, (1) In-hospital birth
Total charges	TOTCHG	Total charges, edited
Length of stay	LOS	Length of stay, edited

Discharge information

Discharge quarter	DQTR	Coded: (1) First quarter, Jan - Mar, (2) Second quarter, Apr - Jun, (3) Third quarter, Jul - Sep, (4) Fourth quarter, Oct - Dec
Discharge year	YEAR	Calendar year
Disposition of patient (discharge)	DIED	Indicates in-hospital death: (0) did not die during hospitalization, (1) died during hospitalization

Type of Data Element	HCUP Name	Coding Notes
status)	DISPUNIFORM	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
	TRAN_OUT	Transfer Out Indicator: (0) not a transfer, (1) transferred out to a different acute care hospital, (2) transferred out to another type of health facility

Weights (to calculate national estimates)

Discharge weights (weights for 1988-1993 are on Hospital Weights file)	DISCWT	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.
--	--------	---

Hospital information

Hospital identifiers (encrypted)	HOSP_NIS	NIS hospital number (links to Hospital Weights file; does not link to previous years)
Hospital location	HOSP_DIVISION	Census Division of hospital (STRATA): (1) New England, (2) Middle Atlantic, (3) East North Central, (4) West North Central, (5) South Atlantic, (6) East South Central, (7) West South Central, (8) Mountain, (9) Pacific
Hospital stratifier	NIS_STRATUM	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.
Record identifier, synthetic	KEY_NIS	Unique record number for file beginning in 2012

[Return to Introduction](#)

Table 2: Data Elements in the 2012 NIS Hospital Weights Files

For prior years, refer to documentation on HCUP-US (e.g. the table of data element availability by years <https://www.hcup-us.ahrq.gov/db/nation/nis/NISvariables1988-2012forHCUP-US.pdf> or previous versions of the NIS Introduction).

Type of Data Element	HCUP Name	Coding Notes
Discharge counts	N_DISC_U	Number of universe discharges in the stratum
	S_DISC_U	Number of sampled discharges in the sampling stratum (NIS_STRATUM or STRATUM)
	TOTAL_DISC	Total number of discharges from this hospital in the NIS
Discharge weights	DISCWT	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.
Discharge Year	YEAR	Discharge year
	N_HOSP_U	Number of universe hospitals in the stratum
	S_HOSP_U	Number of sampled hospitals in the stratum (NIS_STRATUM or STRATUM)
Hospital identifiers	HOSP_NIS	NIS hospital number (links to Hospital Weights file; does not link to previous years)
Hospital characteristics	HOSP_BEDSIZE	Bed size of hospital (STRATA): (1) small, (2) medium, (3) large
	H_CONTRL	Control/ownership of hospital: (1) government, nonfederal, (2) private, non-profit, (3) private, investor-own
	HOSP_LOCTEACH	Location/teaching status of hospital (STRATA): (1) rural, (2) urban non-teaching, (3) urban teaching
	HOSP_REGION	Region of hospital: (1) Northeast, (2) Midwest, (3) South, (4) West
	HOSP_DIVISION	Census Division of hospital (STRATA): (1) New England, (2) Middle Atlantic, (3) East North Central, (4) West North Central, (5) South Atlantic, (6) East South Central, (7) West South Central, (8) Mountain, (9) Pacific
	NIS_STRATUM	Stratum used to sample hospitals beginning in 1998; includes geographic region, control, location/teaching status, and bed size

[Return to Introduction](#)

Table 3: Data Elements in the 2012 NIS Disease Severity Measures Files

For prior years, refer to documentation on HCUP-US (e.g. the table of data element availability by years <https://www.hcup-us.ahrq.gov/db/nation/nis/NISvariables1988-2012forHCUP-US.pdf> or previous versions of the NIS Introduction).

Type of Data Element	HCUP Name	Coding Notes
AHRQ Comorbidity Software (AHRQ)	CM_AIDS	AHRQ comorbidity measure: Acquired immune deficiency syndrome : (0) Comorbidity is not present, (1) Comorbidity is present
	CM_ALCOHOL	AHRQ comorbidity measure: Alcohol abuse: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_ANEMDEF	AHRQ comorbidity measure: Deficiency anemias : (0) Comorbidity is not present, (1) Comorbidity is present
	CM_ARTH	AHRQ comorbidity measure: Rheumatoid arthritis/ collagen vascular diseases : (0) Comorbidity is not present, (1) Comorbidity is present
	CM_BLDLOSS	AHRQ comorbidity measure: Chronic blood loss anemia: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_CHF	AHRQ comorbidity measure: Congestive heart failure: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_CHRNLUNG	AHRQ comorbidity measure: Chronic pulmonary disease: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_COAG	AHRQ comorbidity measure: Coagulopathy: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_DEPRESS	AHRQ comorbidity measure: Depression: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_DM	AHRQ comorbidity measure: Diabetes, uncomplicated: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_DM CX	AHRQ comorbidity measure: Diabetes with chronic complications: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_DRUG	AHRQ comorbidity measure: Drug abuse: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_HTN_C	AHRQ comorbidity measure: Hypertension, (combine uncomplicated and complicated): (0) Comorbidity is not present, (1) Comorbidity is present
	CM_HYPOTHY	AHRQ comorbidity measure: Hypothyroidism: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_LIVER	AHRQ comorbidity measure: Liver disease: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_LYMPH	AHRQ comorbidity measure: Lymphoma : (0) Comorbidity is not present, (1) Comorbidity is present
CM_LYTES	AHRQ comorbidity measure: Fluid and electrolyte disorders: (0) Comorbidity is not present, (1) Comorbidity is present	

Type of Data Element	HCUP Name	Coding Notes
	CM_METS	AHRQ comorbidity measure: Metastatic cancer: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_NEURO	AHRQ comorbidity measure: Other neurological disorders: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_OBESE	AHRQ comorbidity measure: Obesity: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_PARA	AHRQ comorbidity measure: Paralysis: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_PERIVASC	AHRQ comorbidity measure: Peripheral vascular disorders: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_PSYCH	AHRQ comorbidity measure: Psychoses: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_PULMCIRC	AHRQ comorbidity measure: Pulmonary circulation disorders: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_RENLFAIL	AHRQ comorbidity measure: Renal failure: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_TUMOR	AHRQ comorbidity measure: Solid tumor without metastasis : (0) Comorbidity is not present, (1) Comorbidity is present
	CM_ULCER	AHRQ comorbidity measure: Peptic ulcer disease excluding bleeding: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_VALVE	AHRQ comorbidity measure: Valvular disease: (0) Comorbidity is not present, (1) Comorbidity is present
	CM_WGHTLOSS	AHRQ comorbidity measure: Weight loss: (0) Comorbidity is not present, (1) Comorbidity is present
All Patient Refined DRG (3M)	APRDRG	All Patient Refined DRG
	APRDRG_Risk_Mortality	All Patient Refined DRG: Risk of Mortality Subclass: (0) No class specified, (1) Minor likelihood of dying, (2) Moderate likelihood of dying, (3) Major likelihood of dying, (4) Extreme likelihood of dying
	APRDRG_Severity	All Patient Refined DRG: Severity of Illness Subclass: (0) No class specified, (1) Minor loss of function (includes cases with no comorbidity or complications), (2) Moderate loss of function, (3) Major loss of function, (4) Extreme loss of function
Linkage Data Elements	HOSP_NIS	NIS hospital number (links to Hospital Weights file; does not link to previous years)
	KEY_NIS	Unique record number for file beginning in 2012

[Return to Introduction](#)

Table 4: Data Elements in the 2012 NIS Diagnosis and Procedure Groups Files

For prior years, refer to documentation on HCUP-US (e.g. the table of data element availability by years <https://www.hcup-us.ahrq.gov/db/nation/nis/NISvariables1988-2012forHCUP-US.pdf> or previous versions of the NIS Introduction).

Type of Data Element	HCUP Name	Coding Notes
Chronic Condition Indicator	CHRON1 – CHRON25	Chronic condition indicator for all diagnoses: (0) non-chronic condition, (1) chronic condition.
	CHRONB1 – CHRONB25	Body system indicator 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.
Multi-Level Clinical Classifications Software (CCS) Category	DXMCCS1	Multi-level clinical classification software (CCS) for principal diagnosis. Four levels for diagnoses presenting both the general groupings and very specific conditions
	E_MCCS1	Multi-level clinical classification software (CCS) for first listed E Code. Four levels for E codes presenting both the general groupings and very specific conditions
	PRMCCS1	Multi-level clinical classification software (CCS) for principal procedure. Three levels for procedures presenting both the general groupings and very specific conditions
Procedure Class	PCLASS1 – PCLASS15	Procedure Class for all procedures: (1) Minor Diagnostic, (2) Minor Therapeutic, (3) Major Diagnostic, (4) Major Therapeutic
Linkage Data Elements	HOSP_NIS	NIS hospital number (links to Hospital Weights file; does not link to previous years)
	KEY_NIS	Unique record number for file beginning in 2012

[Return to Introduction](#)

APPENDIX IV: TEACHING HOSPITAL INDICATOR ASSIGNMENT

We used the following American Hospital Association Annual Survey Database (Health Forum, LLC © 2013) data elements to assign the NIS Teaching Hospital Indicator:

<u>AHA Data Element Name</u>	<u>= Description [HCUP Data Element Name].</u>
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].

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 DATA ELEMENT */
/* ***** */
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 ;
```


APPENDIX V: 2012 NIS REDESIGN REPORT EXECUTIVE SUMMARY

Many health researchers across the United States rely upon the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample¹ (NIS)—a database of hospital inpatient stays and discharges that is sponsored by the Agency for Healthcare Research and Quality (AHRQ). Studies based on the NIS help policymakers understand cost, access, quality, utilization, and health outcomes of hospital services. It is critical that the NIS be designed to optimize its capacity for national estimates.

The NIS sampling frame has grown from 8 States in 1988, to 22 States in 1998, to 46 States in 2011—currently covering 97 percent of the U.S. population. Because the sampling frame for the NIS contains nearly the entire universe of discharges, in 2012 we evaluated the sampling approach to determine whether a different strategy could improve the accuracy of national estimates from the NIS. As a result of the 2012 evaluation study, a new NIS sample design was recommended. This evaluation:

- Identified challenges associated with the current NIS design.
- Suggested ways in which alternative sample designs might address these challenges.
- Assessed the impact that three alternative designs might have on estimates of the hospital outcomes most commonly studied by NIS users.
- Recommended a design that optimized the use of all available data to produce the most precise and stable national estimates.

AHRQ has elected to deploy the systematic sampling design that was recommended, effective with the 2012 NIS that is planned for public release in June, 2014. This report lays out the implementation of the new design.

Previous Study Results

For a previous evaluation performed during 2012,² the project team considered and compared three alternative sampling designs to the present NIS design: (1) a slight modification to the present NIS design that stratified hospitals into nine census divisions instead of four census regions, (2) a Neyman allocation design that optimized the estimates of average length of stay (ALOS), and (3) a self-weighting systematic design that took into account patient characteristics such as diagnoses, age, and admission date.

The team recommended the systematic design because:

- It is self-weighting, making the sample easy for analysts to use.
- It delivers improved accuracy more consistently across diagnosis-related groups (DRGs) compared with the other alternatives.
- It is straightforward to estimate accurate confidence intervals using standard statistical software.

¹ With the redesign, beginning with 2012 data AHRQ is changing the name from the “Nationwide Inpatient Sample” to the “National Inpatient Sample.”

² Houchens, RL, Ross, DN, Setodji, CM, Uscher-Pines, L, and Roderick J.A. Little. Nationwide Inpatient Sample Redesign Final Report. September 14, 2012. Deliverable #1823.03. Agency for Healthcare Quality and Research, Rockville, MD.

- Its implementation on an annual basis will be efficient and can be accomplished using readily available software and accepted methods.

The present NIS design draws 100 percent of discharges from a sample of approximately 1,000 hospitals, whereas the proposed systematic design samples a fraction of discharges from across all HCUP hospitals (over 4,500 in 2011). The systematic sample is a self-weighted sample design that is similar to simple random sampling, but it is more efficient and it ensures that the sample is representative of the population on the following critical factors—

- Hospital factors
 - Hospital (unidentified)
 - Census division³
 - Ownership
 - Urban-rural location
 - Teaching status
 - Number of beds (bed size category)
- Patient factors
 - DRG
 - Admission month

The superior performance of the systematic design that samples discharges across all hospitals is not surprising, because patient characteristics and mean outcomes vary significantly among hospitals. Variation in mean outcomes such as ALOS, charges, and mortality rates for discharges among hospitals causes a net loss of information under the present NIS design, which draws a sample of hospitals. This is compared with the systematic design, which draws the same total number of discharges across the entire spectrum of hospitals participating in HCUP. Even though the present NIS design stratifies the hospital sample by hospital characteristics, there can be considerable variation in mean outcomes estimated from one hospital sample to the next, depending on which hospitals are selected for the sample. In contrast, the systematic sampling strategy selects a sample of discharges from all hospitals, which better represents the entire universe of hospitals and increases the information in the total sample of discharges.

For national-level estimates, the systematic design reduced the margin of error by 42 to 48 percent over the present NIS design for the outcomes studied (ALOS, average charges, and mortality rates), thus the new NIS design will be about twice as precise as the old design. The margin of error is commonly used by the popular press to describe the reliability of sample statistics. Technically, it is the half-width of a confidence interval around a sample statistic, such as a rate or a mean. The systematic design also consistently reduced the margin of error for estimates at the DRG level.

Finalizing the New Design

In preparation for implementing the systematic sampling design for the 2012 NIS, we:

- Enlisted HCUP Partner support for the new design.

³The nine census divisions (New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific) will be the smallest geographic areas that can be represented using the new NIS rather than the four census regions of the original NIS (Northeast, South, Midwest, West).

- Removed long-term acute care (LTAC) hospitals from the hospital universe.
- Improved estimates of the total number of discharges in the universe used to derive sample weights by using actual SID discharge counts rather than estimates from the American Hospital Association (AHA) Annual Survey, whenever possible.
- Tied the definition of a hospital to the State-supplied hospital identifier rather than to the American Hospital Association (AHA) hospital identifier whenever possible.
- Estimated the effect of all design changes on sample estimates.

We summarize the results of these activities in the following sections.

Enlisted HCUP Partner Support

It is important that HCUP Partners who contribute data approve the new design. Consequently, AHRQ and Truven Health Analytics researchers jointly presented the new design to HCUP Partners and requested feedback. Along with the sample design changes, AHRQ proposed the following changes to enhance confidentiality and focus the NIS on national estimates:

- Drop State identifiers to prevent State-level estimates.⁴
- Drop data elements that were not available uniformly across the States, such as secondary payer, admission type, and data elements with State-specific coding.
- Replace hospital identifiers with a pseudo-identifier.

Partners who attended the presentation indicated their support. The NIS is not designed for State-level analyses, so little is lost analytically by omitting the State names from the NIS record. Users may turn to the State Inpatient Databases (SID) for analyses requiring State identification or State-specific data elements. The use of hospital pseudo-identifiers will help protect hospital identities while preserving the analyst's ability to estimate hospital-level variation.

Removed Long-Term Acute Care Hospitals⁵

The most recent NIS redesign was implemented for the 1998 data year. For the 1998 redesign, rehabilitation hospitals—although classified as community hospitals by the AHA—were excluded from the NIS universe because (1) the State data did not always include discharges from those hospitals, and (2) outcomes for discharges from rehabilitation hospitals were different from discharges from short-term acute care hospitals. Similarly long-term acute care hospitals are classified as community hospitals by the AHA if they have an average length-of-stay (ALOS) less than 30 days. However, during the most recent analyses we determined that they were not uniformly available from all States participating in HCUP, and their ALOS was over 25 days (unlike other community hospitals with an ALOS of about 4.5 days). Thus, we decided to eliminate long-term acute care hospitals from future editions of the NIS. The effects of this change were relatively minor, as we report later.

⁴ Because the NIS was not stratified by State, State-level estimates were not reliable in the original NIS. Dropping State identifiers also facilitated masking of hospital identifiers.

⁵ LTAC hospitals are certified as acute care hospitals, but have an ALOS greater than 25 days. Patients in LTAC hospitals are often transferred from an intensive or critical care unit, generally have more than one serious condition, and are expected to improve and return home. LTAC hospitals typically provide comprehensive rehabilitation, respiratory therapy, head trauma treatment, and pain management services.

Improved Estimates of the Total Number of Discharges in the Universe

Historically, NIS sample weights were calculated by dividing the number of universe discharges by the number of sampled discharges within each hospital stratum. The number of universe discharges had been estimated using data from the AHA annual hospital survey. In particular, the total number of discharges in the universe was estimated by the sum of births and admissions contained in the AHA annual survey for all hospitals in the universe. Given that HCUP Partners supply over 95 percent of discharges nationwide, for future editions of the NIS, we will estimate the universe count of discharges within each stratum using the actual count of discharges contained in HCUP data. We will use the AHA counts only for non-HCUP hospitals in the universe.

This option was not considered for the previous redesign because HCUP data included a much smaller percentage of discharges in the United States, and the differences between HCUP counts and AHA counts would tend to adversely affect trends as the mix of HCUP States changed from year to year. In 2011, for hospitals in both the AHA and the SID, in 43 of 46 States, the AHA survey data estimated State discharge totals that were between 1 percent and 17 percent higher than the observed SID discharge totals. Overall, the AHA survey estimated about a 4 percent higher count of discharges than the observed SID count. Although the current high HCUP State participation rate is an important factor, there are several other reasons for switching to the HCUP count of discharges:

- Beginning in 1994, the AHA survey asks for the count of *admissions* rather than the count of *discharges*. The SID yields *discharge* counts, and NIS inferences are intended for the universe of *discharges*.
- The estimate of discharges from the AHA survey is the sum of reported admissions and reported births. Summing these two values results in a double count of newborns in neonatal intensive care units (NICUs), who appear in births and in admissions. We estimated that this resulted in an over count of about 0.5 percent for the universe of discharges.
- Compared with the SID data, the AHA survey admission counts may more frequently include long-term care and swing bed admissions. Based on supplemental 2010 data from the Illinois Department of Public Health, we confirmed that the AHA survey count for Illinois included long-term and swing bed admissions, whereas the HCUP SID data for Illinois did not. The SID discharge count agreed with counts from the Health Department data after eliminating the double counts of newborns in NICUs and after eliminating long-term care and swing bed admissions.
- More than 50 percent of AHA respondents report statistics for fiscal years, which often end in June or September, rather than for calendar years. The SID files used for the NIS always span a calendar year for each hospital. While the totals would likely be similar, there would be some difference between fiscal year and calendar year discharge totals.
- The AHA survey admission counts are missing about 17 percent of the time, and the AHA imputes admission counts from an undocumented regression equation for these missing values.
- In a survey of HCUP Partners, most of the respondents stated that they believe that the SID data provide more reliable discharge counts than those contained in the AHA survey data.

The effects of this change were significant for estimates of discharge counts, but not for estimates of means and rates, as we report below.

Used State Hospital Identifiers Rather than AHA Hospital Identifiers

A logical corollary of switching from AHA discharge estimates to SID discharge counts was to distinguish unique hospitals using the SID hospital identifiers rather than the AHA hospital identifiers. For the vast majority of hospitals, the SID hospital identifiers are in one-to-one correspondence with the AHA hospital identifiers. However, about 10 percent of the AHA identifiers actually correspond to two or more hospitals in the SID that have common ownership within a hospital system. For these “combined” AHA identifiers, the number of estimated discharges and the number of hospital beds in the AHA data reflect the sum of estimated discharges and the sum of beds, respectively, from the constituent hospitals. As a result, these combined hospitals could have been allocated to the wrong bed size stratum in the sample design. Also, the between-hospital variance was combined with the within-hospital variance for these combined hospitals.

In some States, the SID hospital identifiers demonstrate the same weakness as the AHA hospital identifiers, and those hospitals remain combined in the new design even though we are switching to the SID hospital identifier. However, use of the SID hospital identifiers disaggregates the previously combined hospitals in many other States, which is likely to improve the classification of hospitals and improve variance estimates.⁶ The marginal effect of this change on outcome estimates was very small, as we report next.

Estimated the Effects of Design Changes on Sample Estimates

The switch from drawing *all discharges from a sample of hospitals* to drawing a *sample of discharges from all hospitals* improved the precision and stability of NIS sample estimates. However, the other modifications listed above affected the values of universe statistics (i.e., the values that sample statistics try to estimate). In particular, these modifications had an effect on the numbers and types of discharges in the universe. Using HCUP and AHA annual survey data for 2011, we estimated the effects of these changes:

1. Switching to the *systematic sample design* from the present NIS sample design⁷
2. Eliminating *long-term acute care* hospitals
3. Using observed *SID discharge counts* in place of estimated AHA discharge counts for estimating the total number of discharges in the universe
4. Using *SID hospital identifiers* in place of AHA hospital identifiers to disaggregate hospitals combined by the AHA hospital identifier.

[Table 1](#) summarizes the effects of these modifications on four universe statistics—discharges, ALOS, average charges, and hospital mortality—obtained from HCUP discharge data and AHA survey data for 2011. The columns are numbered for easy reference. Columns 1 and 2 provide the baseline statistics and describe the universe without any modifications.

⁶ This difference in hospital identifiers renders the NIS hospital-level weights inaccurate. Consequently, hospital-level weights will no longer be provided with the NIS.

⁷ This includes a revision of the hospital sampling strata to stratify hospitals by the nine census divisions rather than by the four census regions used in the existing NIS design. Switching to the systematic design had no effect on the universe and, therefore, no effect on values of universe statistics.

Columns 3 and 4 show the effect of excluding LTAC hospitals from the universe. The total number of discharges declined from 38,590,733 (column 1) to 38,338,545 (column 3), which represents a 0.7 percent overall decline. This decline was mostly in the older age groups (not shown). The removal of LTAC hospitals also decreased ALOS by 1.5 percent, average charges by 0.7 percent, and hospital mortality by 2.0 percent (from a mortality rate of 1.91 percent to 1.87 percent). These changes are all to be expected given the characteristics of patients in LTAC hospitals.

Columns 5 and 6 show the effect of replacing AHA discharge counts with SID discharge counts to estimate discharges in the universe (in addition to excluding LTAC hospitals). This action had a significant impact on the universe discharge count. The total number of discharges in the universe fell from 38,338,545 (column 3) to 36,935,306 for a further decrease of 3.6 percent and an overall decrease of 4.3 percent, compared with the discharge count in column 1. The incremental impact on ALOS, average charges, and hospital mortality was almost negligible in comparison.

Finally, the incremental effects of switching from the AHA hospital identifier to the SID hospital identifier (columns 7 and 8) were miniscule for all four outcomes.

In summary, based on the changes implemented in the redesign, we expect overall trends in discharge counts to decline by about 4.3 percent, overall trends in ALOS to decline by about 1.5 percent, overall trends in total charges to decline by about 0.5 percent, and overall trends in hospital mortality to decline by about 2.0 percent.

[Table 2](#) summarizes the effects of these modifications on the margin of error for sample statistics. The entries in [Table 2](#) show the margin of error for the new sample design in relation to the margin of error for the present NIS design. For example, an entry of 0.50 means that the margin of error for a statistic generated from a sample under the new design is half that of a statistic generated from a sample under the present sample design (for a sample of about 8 million discharges). In other words, an entry of 0.50 means that confidence intervals under the new design would be about half the length of confidence intervals under the old design. These results (based on 2011 data) were very similar to last year's results (based on 2010 data).

For discharge counts, the entries of 1.0 indicate that there is no improvement to the margin of error for estimates of total discharges at the national level. This is by design. At the national level, the sample weights always sum to the total number of discharges in the universe. However, the estimates of total discharges for subsets of the population showed substantial improvements, as is shown in the results chapter of this report.

For ALOS, average charges, and hospital mortality, the improvements were substantial at the national level. The margins of error under the new design are expected to be about 53 percent of the old design for ALOS estimates, about 55 percent of the old design for average charge estimates, and about 51 percent of the old design for estimates of hospital mortality. As can be seen by comparing entries across the columns of [Table 2](#), the improvements continue through the incremental changes to the universe definition.

Moreover, as shown in the results chapter of this report, these improvements persist for discharges classified by age, sex, and DRGs. For example, across all 752⁸ DRGs, the margins of error for the new design compared with the old design average 46 percent lower for total discharges, 36 percent lower for ALOS, 41 percent lower for average charges, and 28 percent lower for in-hospital mortality rates. Further, for 90 percent of DRGs the new margins of error are at least 41 percent lower for total discharges, 29 percent lower for ALOS, 34 percent lower for average charges, and 22 percent lower for in-hospital mortality rates.

Conclusions

In sum, the NIS redesign planned to take effect for the 2012 NIS (to be released in 2014) is expected to provide more stable and precise estimates than previous versions of the NIS. Because long-term acute care hospitals will be excluded and because the accuracy of discharge weights will be improved, NIS users should expect a one-time decrease to historical trends for discharge counts of about 4 percent. They should also expect smaller one-time disruptions to historical trends for rates and means estimated from the NIS, beginning with data year 2012. To address this, we recommend that AHRQ provide NIS users with “trend” discharge weights for historical NIS files to minimize the effects of the redesign on estimated trends that cross the 2012 data year.

⁸ For calendar year 2011, the data combined DRG version 28 (effective 10/1/2010 with 747 DRGs) and version 29 (effective 10/1/2011 with 751 DRGs). One DRG (number 15) in version 28 was replaced by two DRGs (numbers 16 and 17) in version 29, resulting in 752 different DRGs.

Table 1: Impact of Incremental Modifications to the Universe on Universe Statistics.

	<i>Old Universe Definition</i> (1998–2011)		Impact of Incremental Modifications to the Universe					
	Include LTAC Hospitals		Exclude LTAC Hospitals					
	Use AHA Discharge Counts		Use AHA Discharge Counts		Use SID Discharge Counts*			
	Use AHA Hospital ID		Use AHA Hospital ID		Use AHA Hospital ID		<i>New Universe Definition</i> Use SID Hospital ID	
	Total Discharges	Percentage of Original Discharges	Total Discharges	Percentage of Original Discharges	Total Discharges	Percentage of Original Discharges	Total Discharges	Percentage of Original Discharges
Column Number	1	2	3	4	5	6	7	8
Discharge Count	38,590,733	100.0	38,338,545	99.3	36,935,306	95.7	36,939,183	95.7
ALOS	4.59	100.0	4.53	98.5	4.52	98.5	4.53	98.5
Average Charges	\$34,962	100.0	\$34,711	99.3	\$34,779	99.5	\$34,790	99.5
Hospital Mortality	0.01905	100.0	0.01867	98.0	0.01866	97.9	0.01866	98.0

Data sources: HCUP State Inpatient Databases (SID) and American Hospital Association (AHA) Survey Data for 2011

* When discharge counts or hospital identifiers are not available from the SID, estimates from the AHA will be used. This is expected to affect fewer than 10 percent of hospitals.

Abbreviations: ALOS, average length of stay; ID, identification number; LTAC, long-term acute care.

Table 2: Impact of Incremental Modifications to the Universe on the Margin of Error for Sample Statistics

	<i>Old Universe Definition (1998–2011)</i>	Impact of Incremental Modifications to New NIS Design			
	Include LTAC Hospitals	Exclude LTAC Hospitals			
	Use AHA Discharge Counts	Use AHA Discharge Counts	Use SID Discharge Counts*		
	Use AHA Hospital ID	Use AHA Hospital ID	Use AHA Hospital ID	<i>New Universe Definition</i> Use SID Hospital ID	
Column Number	1	2	3	4	
Discharge Count	1.00	1.00	1.00	1.00	
ALOS	0.53	0.52	0.52	0.53	
Average Charges	0.55	0.58	0.57	0.55	
Hospital Mortality	0.57	0.55	0.55	0.51	

Based on 500 Simulated Samples, HCUP 2011 Data.

* When discharge counts or hospital identifiers are not available from the SID, estimates from the AHA will be used. This is expected to affect fewer than 10 percent of hospitals.

Abbreviations: AHA, American Hospital Association; ALOS, average length of stay; LTAC, long-term acute care; SID, State Inpatient Databases