
**eAppendix 1.** Study Selection  
**eAppendix 2.** Description of the National Inpatient Sample  
**eAppendix 3.** Worksheet for Evaluating the Methodology of NIS Studies  
**eAppendix 4.** Detailed Criteria for Adjudicating Study Methodology  
**eFigure 1.** Distribution of Selected Studies by Impact Factor  
**eFigure 2.** Distribution of Selected Studies by Journal Type: Medical vs Surgical  
**eFigure 3.** Distribution of Selected Studies by Medical Subspecialty  
**eReferences**

This supplementary material has been provided by the authors to give readers additional information about their work.
eAppendix 1. Study Selection

We identified all studies published using the National Inpatient Sample (NIS) data between January 1, 2015 and December 31, 2016. The Agency for Healthcare Research and Quality’s (AHRQ) website for its Healthcare Cost and Utilization Project (HCUP) has an online repository of all studies published using the NIS by calendar-year of publication. This is publicly available at: https://www.hcup-us.ahrq.gov/reports/pubsearch/pubsearch.jsp. We identified all 1056 studies published during 2015 and 2016 that were listed on this repository. Further, to ensure completeness, in conjunction with an experienced librarian, we identified additional studies from OVID Medline and OVID Medline-in-Process. Using search terms “National Inpatient Sample” and “Nationwide Inpatient Sample” within the title and abstract sections, and manual sorting, we identified an additional 88 NIS studies for a total of 1144 studies. Finally, after manual removal of duplicates and conference proceedings, there were a total of 1082 unique studies included in the sampling frame. The search and study selection process was conducted on March 19, 2017. We selected all 25 studies published in journals with a Journal Citation Reports® impact factor (2015) of 10 or higher, and a simple random sample of 100 studies published in journals with an impact factor of less than 10 (Figure 1 in main paper). We pre-specified an impact factor of 10 as an arbitrary indicator of high-impact journals.
eAppendix 2. Description of the National Inpatient Sample

The National Inpatient Sample (NIS) is a random 20% sample of inpatient discharges from U.S. community hospitals regardless of the payer, constructed by AHRQ annually since 1988 to specifically allow estimates of national healthcare resource utilization. Under its Healthcare Cost and Utilization Project (HCUP), AHRQ has forged a partnership between multiple statewide data organizations to contribute all-payer healthcare utilization data annually, and has leveraged a federal-state-industry partnership to construct the NIS.

From 1988-2011, the NIS (then called Nationwide Inpatient Sample) was constructed as an annual sample of nearly 1000 U.S. community hospitals after stratification by hospital ownership, bed size, teaching status, urban/rural location, and the 9 U.S. census divisions (summarized in Table 1 of the main paper). These hospitals were sampled only from participating U.S. states, whose numbers increased from 8 states in 1988 to 46 in 2011, and were selected in a manner to represent 20% of all U.S. hospitals defined by the American Hospital Association survey across participating and non-participating states. All discharges captured in the comprehensive, respective State Inpatient Databases from the sampled hospitals, regardless of disposition and outcome, were included in the NIS. The State Inpatient Databases, themselves, capture data in a uniform format allowing effective pooling of data across states.

Discharge weights were calculated by AHRQ and were based on the number of sampled hospitals in each stratum relative to the number of U.S. community hospitals within similar strata defined nationally across both participating and non-participating states. During this period, the NIS underwent its first major redesign, which included how a hospital unit in the NIS was defined, specifically, rehabilitation centers were excluded from its sampling pool. Revised discharge weights were made available to ensure consistency in trend analyses spanning the 1998 data.

The second major redesign of the NIS occurred with the 2012 data. With increasing state participation in HCUP in recent years, the NIS directly covered approximately 97% of the U.S. population. Therefore, beginning in 2012, to improve the representativeness of the sample, and the precision and stability of weighted national estimates, the NIS was redesigned from a 20% sample of hospitals to a 20% sample of patients captured from all hospitals in participating states (now called National Inpatient Sample). In the redesigned sample, discharges from all hospitals are sorted by hospital ownership, bed size, teaching status, urban/rural location, and 4 U.S. census regions, and every 5th discharge is selected. Further, while the American Hospital Association Survey was used to define hospital-units in the sample prior to 2012, the redesigned NIS defined hospital-systems based on the data from State Inpatient Databases (SIDs) of the participating states. Discharges from psychiatric hospitals, rehabilitation centers, long-term acute care centers, and federal institutions are not included. Discharge weights to obtain national estimates too were recalculated at a discharge level using the number of discharges within each stratum as opposed to the number of hospitals in earlier years. This redesigned approach improved the margin of error for national estimates by 42-48% over the previous sampling methodology. The database redesign has major implications for trend analyses. To allow assessment of trends across these distinct designs, the revised sampling approach was applied retrospectively to 1993-2011 data to obtain revised discharge weights, available as ‘trend weights’ from HCUP. The validity of this approach was rigorously evaluated and is currently recommended as the standard analysis approach. Further, it is recommended that trend analyses using redesigned data do not include data from the years 1988-1992.

The NIS data are organized such that each observation in the sample represents a unique hospitalization with information on more than 100 clinical characteristics: patient demographics (e.g., age, sex, race, median income for ZIP code), hospital characteristics (e.g., ownership, size, teaching status, census region and division), primary and up to 24 secondary diagnoses as well as 15 procedures as administrative codes, Diagnosis Related-Group codes for disease severity, discharge status and disposition, total charges, and length of stay.

NIS is best used for examining the utilization of hospital health services, practice variation, cost assessment, and to assess the impact of health policy interventions. Moreover, since it includes data on nearly 8 million hospitalizations each year, it affords the opportunity to examine healthcare utilization for rare disease conditions that are managed primarily in an inpatient setting.

However, there are some important considerations related to its design and analysis that may be prone to misinterpretation: (a) NIS does not identify individual patients and hence longitudinal information is unavailable; repeated inpatient encounters for the same patient may be captured more than once; (b) it does not capture post-discharge outcomes, and estimates that rely on this information are not possible; (c)
it does not capture information on outpatient encounters or observation-only stays, and conditions and procedures predominantly occurring in these settings may be underrepresented in the sample; (d) the sampling methodology does not allow accurate estimates of subset volumes (e.g. state-level estimates cannot be obtained by applying discharge weights to all sampled hospitals/patients within a state) and similarly, after changes in methodology in 2012, due to variable sampling rates across hospitals, applying discharge weights to the sample from a hospital does not generate hospital-level estimates; (e) analyses of national estimates need to account for both sample design, specifically clustering and stratification, as well as changes in sampling strategies over time; (f) NIS uses administrative diagnosis and procedure codes listed on the billing record for the corresponding hospitalizations and these codes are not individually validated. Hence, the diagnosis and procedure codes that do not affect billing substantially may be prone to variation in reliable coding practices.

Additional resources:

1. Data Documentation: The NIS data are supported with robust documentation on the HCUP website. The data for the years 2011 and 2012 are representative of the pre-2012 and post-2012 design respectively. There are available at:

   NIS 2011 data documentation

   NIS 2012 data documentation
   https://www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2012.jsp

2. Analytical support:
   
   a) Tutorials: The following webpage details the survey analysis process step-by-step:

   Other tutorials provide details on data handling for NIS analyses:
   https://www.hcup-us.ahrq.gov/tech_assist/tutorials.jsp

   b) HCUP Online Methods Series:

   The following webpage links to methodological considerations that apply broadly to HCUP data analysis and data interpretation in selected scenarios:

   https://www.hcup-us.ahrq.gov/reports/methods/methods.jsp

   c) Frequently Asked Questions:

   https://www.hcup-us.ahrq.gov/tech_assist/faq.jsp#nis

3. HCUPnet: To allow investigators to test their analyses and results against those generated through the appropriate methodology, the ‘HCUPnet’ portal presents a point-and-click interface to assess the overall capture rate of specific diagnosis and procedure claims codes nationally. It is available at:
https://hcupnet.ahrq.gov/
eAppendix 3. Worksheet for Evaluating the Methodology of NIS Studies

Section A: Data Interpretation
1. Did the research study infer that captured encounters represent distinct patients?
   ☐ Yes [1] ☐ No [0]

Section B: Research Design
2. Did the study obtain estimates for one or more U.S. states?
   ☐ Yes [1] ☐ No [0]
3. Did the study obtain estimates for hospital- or facility-level volumes using data from NIS 2012-14?
   ☐ Yes [1] ☐ No [0]
4. Did the study assign diagnosis or procedure volume estimates to individual physicians?
   ☐ Yes [1] ☐ No [0]
5. Were the primary or secondary International Classification of Diseases (ICD) diagnosis codes used to infer in-hospital events without a prior validation of such an approach?
   ☐ Yes [1] ☐ No [0]

Section C: Data Analysis
6. Did the study not explicitly account for the survey design and its components - clustering, stratification, and weighting?
   ☐ Yes [1] ☐ No [0]
7. Did the study not adequately address changes in data structure over time (for trend analyses)?
   ☐ Yes [1] ☐ No [0]
eAppendix 4. Detailed Criteria for Adjudicating Study Methodology

General instructions
All questions are structured so that an incorrect research practice scores a point (1), and an accurate practice gets a zero (0).

Question-specific instructions

1. Did the research study infer that captured encounters represent distinct patients?
In addition to reviewing the entire manuscript, focus on the following sections to assess this research practice:
   a. Abstract
   b. Methods (patient selection)
   c. Results (Opening paragraph)
If the above do not refer to observations as ‘discharges’ or ‘encounters,’ but as unique ‘patients,’ review ‘Limitations’ section of the Discussion for an acknowledgment that the data represent encounters and not individual patients.

If all sections above, including the discussion section, do not mention this aspect, score a 1.
If they report NIS encounters as discharges or encounters, then score a 0.

2. Did the study obtain estimates for one or more U.S. states?
If there is any mention of state-level estimates, or estimates for groups of states using the NIS data, score a 1, else a 0.

3. Did the study obtain estimates for hospital- or facility-level volumes using data from NIS 2012-14?
A study will be included in the denominator for this research practice if it performed hospital- or facility-level volume assessments.
If study assessed hospital volumes in the NIS data for 2012 or after, score a 1. Hospital volume assessments can only be performed for NIS 1988-2011.
If not, score a 0.

4. Did the study assign diagnosis or procedure volume estimates to individual physicians?
Physician-level volume estimates cannot be accurately performed with the NIS due to differences in reporting practices across states, inability to link a physician with a given aspect of healthcare (e.g. a procedure), and the use of the provider codes to identify groups of physicians in some practices and individual providers in other states. If performed, score a 1. Else score a 0.

5. Were the primary or secondary International Classification of Diseases (ICD) diagnosis codes used to infer in-hospital without a prior validation of such an approach?
Were any of the outcomes defined by presence of ICD-9 codes, such that you cannot differentiate if these were present on admission (comorbidities) or true complications? If yes, score 1.

As an example, “anemia” as a secondary diagnosis code can be wrongly assumed to represent a bleeding complication from a procedure, but it may have been present on admission.

However, if the study assesses procedural complications, the use of ICD-9CM diagnosis codes 996.xx to 999.xx specifically identifies post-procedural complications. Therefore, score a 0 for these studies.

If a study used the Agency of Healthcare Research and Quality’s validated patient safety indicators (PSIs) to identify complications, score a 0.
6. Did the study not explicitly account for the survey design and its components - clustering, stratification, and weighting?

If the study mentions one of the following in the methods or other sections of the manuscript, score a 0 (i.e. study accounts for this):

   a) “SURVEY or SVY commands were used to analyze the data”
   b) “Analyses account for stratification and clustering of data”
   c) “Analyses account for the complex survey design of the data”

Please note that studies that say only ‘weighting’ was done to obtain national estimates and do not mention that the data’s complex survey structure was accounted for in the analyses will be scored a 1. All studies that do not mention the above at all will also score a 1.

7. Did the study not adequately address changes in data structure over time (for trend analyses)?

Sampling design changed in 1998 and 2012. A study will be included in the denominator for this metric if it used data that included these transition points.

For evaluation of the research practice among eligible studies, did the candidate study mention these changes, and account for them in their analyses by explicitly stating in the methods section? They may account for these changes by (1) stating that these changes were accounted for in their analyses, or (2) specifically mentioning the use of modified discharge weights. If a study uses either of the two approaches, score a 0, otherwise, score a 1.
eFigure 1. Distribution of Selected Studies by Impact Factor

Legend: Distribution of selected studies across journal impact factors (for studies in journals with impact factor <10).
Legend: Distribution of selected studies by nature of the source journal, medical vs. surgical (for studies in journals with impact factor <10).
eFigure 3. Distribution of Selected Studies by Medical Subspecialty

Legend: Distribution of selected studies by the field of study of the source journal (for studies in journals with impact factor <10).

- Medicine/Medical subspecialty
- Surgery/Surgical subspecialty
- Obstetrics and Gynecology
- Pediatrics
- Mental and behavioral health

Selected studies
NIS Universe

Percentage of studies