



HHS Public Access

Author manuscript

JAMA. Author manuscript; available in PMC 2018 November 28.

Published in final edited form as:

JAMA. 2017 November 28; 318(20): 2011–2018. doi:10.1001/jama.2017.17653.

Adherence to Methodological Standards in Research using the National Inpatient Sample

Rohan Khera, MD¹, Suveen Angraal, MBBS², Tyler Couch, BS¹, John W. Welsh, BS², Brahmajee K. Nallamothu, MD, MPH³, Saket Girotra, MD SM⁴, Paul S. Chan, MD, MS⁵, and Harlan M. Krumholz, MD SM^{2,6}

¹Division of Cardiology, University of Texas Southwestern Medical Center, Dallas, Texas

²Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, Connecticut

³Division of Cardiology, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan

⁴Division of Cardiovascular Medicine, Department of Internal Medicine, University of Iowa Carver College of Medicine, Iowa City, Iowa

⁵Saint Luke's Mid America Heart and Vascular Institute and the University of Missouri-Kansas City, Kansas City, Missouri

⁶Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine and the Department of Health Policy and Management, Yale School of Public Health, New Haven, Connecticut

Abstract

Importance—Publicly available datasets hold much potential, but their unique design may require specific analytic approaches.

Objective—To determine adherence to appropriate research practices for a frequently used large public database, the National Inpatient Sample (NIS) of the Agency for Healthcare Research and Quality (AHRQ).

Design, Setting and Participants—In this observational study, of the 1082 studies published using the NIS from January 2015 – December 2016, a representative sample of 120 studies was systematically evaluated for adherence to practices required by AHRQ for design and conduct of research using the NIS.

Exposure—None

Corresponding Author. Dr. Harlan M. Krumholz; 1 Church St., Suite 200, New Haven, CT 06510, 203-764-5885; harlan.krumholz@yale.edu.

Access to Data and Data Analysis: Drs. Khera and Krumholz had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Disclosures: The other authors report no potential conflicts of interest.

Main Outcomes—All studies were evaluated on 7 required research practices based on AHRQ’s recommendations, compiled under 3 domains: (A) data interpretation (interpreting data as hospitalization records rather than unique patients); (B) research design (avoiding use in performing state-, hospital-, and physician-level assessments where inappropriate; not using non-specific administrative secondary diagnosis codes to study in-hospital events), and (C) data analysis (accounting for complex survey design of the NIS and changes in data structure over time).

Results—Of 120 published studies, 85% (n=102) did not adhere to 1 required practices and 62% (n=74) did not adhere to 2 required practices. An estimated 925 (95% CI 852–998) and 696 (95% CI 596–796) NIS publications had violations of 1 and 2 required practices, respectively. A total of 79 sampled studies, representing 68.3% (95% CI 59.3–77.3) of the 1082 NIS studies, did not account for the effects of sampling error, clustering, and stratification; 62 (54.4%, 95% CI 44.7–64.0) extrapolated non-specific secondary diagnoses to infer in-hospital events; 45 (40.4%, 95% CI 30.9–50.0) miscategorized hospitalizations as individual patients; 10 (7.1%, 95% CI 2.1–12.1) performed state-level analyses; and 3 (2.9%, 95% CI 0.0–6.2) reported physician-level volume estimates. Of 27 studies (weighted: 218 studies, 95% CI 134–303) spanning periods of major changes in the data structure of the NIS, 21 (79.7%, 95% CI 62.5–97.0) did not account for the changes. Among the 24 studies published in journals with an impact factor 10, 16 (67%) and 9 (38%) did not adhere to 1 and 2 practices, respectively.

Conclusions and Relevance—In this study of 120 recent publications that used data from the NIS, the majority did not adhere to required practices. Further research is needed to identify strategies to improve the quality of research using the NIS and assess whether there are similar problems with use of other publicly available data sets.

BACKGROUND

Publicly available datasets hold much potential and support the assessment of patterns of care and outcomes. Further, they lead to democratization of research, thereby allowing novel approaches to studying disease conditions, processes of care, and patient outcomes.¹ However, their design properties may require specific analytic approaches. The National Inpatient Sample (NIS) is a large administrative database produced by the Agency for Healthcare Research and Quality (AHRQ) that has been increasingly used as a data source for research.² Developed under the AHRQ Healthcare Cost and Utilization Project (HCUP), the NIS includes administrative and demographic data from a 20% sample of inpatient hospitalizations in the United States, and has been compiled annually since 1988 through a partnership between multiple statewide data organizations to contribute all-payer healthcare utilization data annually.^{3,4}

The NIS, however, has design features that require specific methodological considerations. Therefore, AHRQ supports the data with robust documentation, including a detailed description of sampling strategies and data elements for each year,⁵ a step-by-step description of the required analytic approach in multiple online tutorials,⁶ and a section on known pitfalls.⁷ Further, it allows investigators to examine the accuracy of the analytical approach using the web-based tool ‘HCUPnet,’ which provides weighted national estimates for every diagnosis and procedure claim code using a simple interface.⁸ The inferences and

interpretation drawn from studies that use data from the NIS but do not adhere to these resources may contain inaccuracies.

Given the recent proliferation of research using data from the NIS, this study systematically assessed the use of appropriate research practices in contemporary investigations using the NIS across a spectrum of biomedical journals.

METHODS

Study Selection

We performed a systematic evaluation of a randomly selected subset of peer-reviewed articles published from the NIS from January 1, 2015 through December 31, 2016 using a checklist of major methodological considerations relevant to the database. Using data from a public repository of publications from the NIS on the HCUP website,⁹ supplemented with data from bibliographic repositories, we identified 1082 unique studies (Appendix A in the Supplement). From these, we selected all 25 studies that were published in journals with a Journal Citation Reports[®] impact factor (2015) of ≥ 10 and a simple random sample of 100 additional studies that were published in journals with an impact factor of <10 (Figure 1). The sampling of studies was performed using the SURVEYSELECT procedure in SAS 9.4 (Cary, NC), wherein all studies ($N=1057$) in journals with an impact factor <10 were assigned a random number, and 100 studies were selected with each study having an equal probability of being selected in the sample (sampling-probability = $100 \div 1057$). The representativeness of the sample was assessed against the NIS universe for (a) distribution of studies across the spectrum of journal impact factors, (b) the nature of the source journal (medical or surgical), and (c) the clinical field of the journals (medicine/medical subspecialties, surgery/surgical subspecialties, pediatrics, obstetrics and gynecology, or mental/behavioral health) in which the articles were published.

The inverse of the sampling-probability, or 10.57, represents the sampling weight for the studies published in journals with an impact factor <10 . Since all studies with an impact factor ≥ 10 were selected, their corresponding sampling weight was 1.

Evaluation Criteria

All selected studies were evaluated for 7 research practices in the major domains of data interpretation, research design, and data analysis. These research practices were compiled based on the publicly accessible recommendations by AHRQ for the use of the NIS.^{3,5-7,10-14} The design of the NIS and required research practices for use of the data are described in Appendices B and C in the Supplement. Adherence to these research practices is essential to draw appropriate conclusions using data from the NIS, and are therefore required of all studies using these data. The 7 research practices (Table 1) are described briefly below.

Data Interpretation—The NIS is a record of inpatient hospitalization events.^{4,12} Therefore, we evaluated if studies using the NIS correctly portrayed observations as hospitalization events or discharges rather than as unique patients (*Practice 1*).

Research Design—*Practice 2* requires the avoidance of using the NIS to assess state-level patterns of care or outcomes.¹¹ To permit assessment of national estimates, the NIS is constructed using a complex survey design in which sampling of hospitalizations is based on pre-defined hospital strata.^{10,12,14} This sampling design does not include states, and sampling from states may not be representative of hospitalizations in that state.¹¹ Similarly, since 2012 the data structure for the NIS changed from a sample of 100% discharges from 20% of hospitals in the United States to a national 20% sample of patients, precluding hospital volume-based analyses beyond data from 1988–2011.^{10,14} Therefore, we evaluated if studies limited hospital-level analyses to data from the NIS for 1988–2011 (*Practice 3*). In addition, given the inconsistent meaning of the available provider field code, which refers to either individual physicians or physician groups, physician-level volumes cannot be reliably assessed.^{13,15} Therefore, we evaluated if the NIS was used to obtain physician-level estimates (*Practice 4*).

Since the record of hospitalization in NIS includes 1 principal and up to 24 secondary diagnosis codes without a present-on-admission indicator, there is limited ability to distinguish complications from comorbid conditions.^{16,17} Thus, it is recommended that validated algorithms that use a combination of diagnosis-related groups and secondary diagnosis codes to specifically identify comorbid conditions (e.g., Elixhauser’s comorbidities) and complications (patient safety indices developed by AHRQ or secondary codes specific to post-procedure complications) be used.^{18–20} Therefore, we evaluated if studies used non-specific secondary diagnosis codes to infer in-hospital events (*Practice 5*).

Data Analysis—The appropriate interpretation of data from the NIS, which are compiled using a complex survey design, requires the use of survey-specific analysis tools that simultaneously account for clustering and stratification as well as the potential for sampling bias, and allow weighting of estimates to generate national estimates with an accompanying measure of variance of the estimate.⁶ Therefore, we assessed if analyses were performed using appropriate survey methodology (*Practice 6*). In addition to the sampling redesign in data after 2011, major data changes in 1998 necessitated the use of modified discharge weights in studies spanning these transition years.^{14,21} Thus, we assessed if studies spanning these 2 transition points followed special considerations to ensure accurate assessment of trends, specifically through the use of modified discharge weights to obtain accurate estimates (*Practice 7*).

Evaluation of Selected Studies

The 7 practices were assessed by an objective set of criteria for grading each study (Appendix D in the Supplement). Five of the practices are applicable in all settings, thus they applied to all the studies. The remaining 2 (Practices 3 and 7) are only applicable to studies that performed hospital-volume assessments and therefore would require analyses that are limited to data before 2012, and the performance of trend analyses spanning transitions in the NIS with the required modifications to analyses, respectively. Before study evaluation, all investigators involved in data abstraction (S.A., T.C, J.W., and R.K.) reviewed a standard summary of the methodological design of the NIS compiled by all investigators and reviewed the official data documentation reflecting the 2 different sampling designs

(before 2011, and 2012 and later). Each study was evaluated independently by 2 of 3 investigators (S.A., T.C., and J.W.) and results were collated and confirmed by a fourth abstractor (R.K.). There was good inter-rater reliability (kappa statistic 0.88) and disagreements were resolved with mutual agreement and/or discussion with the senior author (H.M.K.). All study outcomes are reported as the percentage of eligible studies that did not adhere to a research practice.

To estimate the overall frequency of violations in the universe of studies published using the NIS in 2015 and 2016, we used survey methodology that accounted for our stratified sampling of studies based on journal impact factor. For these analyses, we used a journal impact factor of ≥ 10 or <10 as the stratification variable and the corresponding sampling weights for these strata (sampling weights 1 and 10.57, respectively) to obtain weighted estimates for the universe of NIS studies published during this period.

Further, to examine the association between the publications and other investigations and guidelines, we evaluated their citation record using Google Scholar[®] citations on April 4, 2017. We repeated all analyses after stratification of publications as having an impact factor ≥ 10 and <10 . We used chi-square and Fisher's exact test to compare differences in categorical outcomes, and the non-parametric Wilcoxon rank-sum test and non-parametric regression to compare continuous outcomes.

Data Simulation

To demonstrate the practical implications of these errors, we present an example based on our own analyses. We used the NIS data from the years 2010 through 2013 to simulate errors in the assessment of *hospitalization-level* trends in the use of coronary artery bypass grafting (CABG) in the United States, emphasizing the need for using a survey-specific methodology and accounting for major changes in data structure over time (*practices 6 and 7*). In this example, hospitalizations with CABG procedures were identified using the clinical classification software procedure code 44. We examined temporal trends in CABG procedures during 2010–2013, using a set of modified discharge weights for the years 2010–2011 (*correct weights*) that accounted for changes in the NIS data structure for subsequent years. We then simulated these trends using discharge weights that did not account for changes in data structure over time (*incorrect weights*). Differences in time-trends with these two approaches were assessed using the analysis of covariance test.

All analyses were performed using SAS 9.4 (Cary, NC) using 2-sided statistical tests, and a level of significance set at an alpha of 0.05. The study's use of the NIS data was exempted from the purview of Yale University's Institutional Review Board since the data were de-identified.

RESULTS

Of the 125 publications in our initial cohort (all 25 studies published in journals with an impact factor ≥ 10 and a random sample of 100 studies published in journals with an impact factor <10), 5 studies (1 (4%) in a journal with an impact factor ≥ 10 and 4 (4%) in journals with an impact factor <10) used multiple datasets with limited information on NIS-specific

methodology, precluding methodological evaluation, leaving 120 studies for detailed evaluation of research practices (Figure 1). The selected studies were representative of the universe of NIS studies with respect to journal impact factor, nature, and clinical field of the source journal (eFigures 1–3 in Appendix E).

Of these, 78 (65%) qualified for evaluation on 5 research practices, 40 (33%) for 6 practices, and 2 (2%) for all 7 practices.

Of the 120 studies, only 18 satisfied all required practices, representing 10.5% (95% CI 4.7–16.4) of the 1082 studies published using the NIS during the study period. A total of 28 studies (21.2%, 95% CI 13.2–29.1) violated 1 required research practice and 74 (64.3%, 95% CI 55.0–73.6) violated 2 or more practices (36 (31.6%; 95% CI 22.6–40.7) violated 2 practices, 30 (24.9%; 95% CI 16.5–33.3) violated 3 practices, and 8 (7.8%; 95% CI 2.5 to 13.1) violated 4 or more practices) (Table 2). Therefore, an estimated 925 (95% CI 852–998) studies had 1 and 696 (95% CI 596–796) studies had 2 violations of required research practices, among the 1082 unique studies published using the NIS during 2015–2016 (Table 2).

The percentage of studies that did not adhere to individual required practices varied considerably (Table 3). Denominators varied by each of the evaluated research practices. Of the 120 studies, 79 did not account for the complex survey design of NIS in their analyses, corresponding to 68.3% (95% CI 59.3–77.3) of the studies in the universe of 1082 NIS studies, 62 (54.4%, 95% CI 44.7–64.0) used non-specific secondary diagnosis codes to infer complications, 45 (40.4%, 95% CI 30.9–50.0) reported results to suggest that NIS included individual patients rather than hospitalizations (without addressing this in the interpretation of their results), 10 (7.1%, 95% CI 2.1–12.1) improperly performed state-level analyses, and 3 (2.9%, 95% CI 0.0–6.2) improperly performed physician volume estimates. Seventeen studies performed an assessment of diagnosis and/or procedure volumes at the hospital, corresponding to an estimated 141 (95% CI 71–212) overall. Of these, 2 studies in the sample (8.2%, 95% CI 0.0–22.5) included data from 2012 when such estimates were unreliable. In addition, although 27 studies (weighted: 218 studies, 95% CI 134–303) periods of major data redesign in the NIS, the analyses in 21 (79.7%, 95% CI 62.5–97.0) of these did not account for the changes.

Studies published in journals with an impact factor ≥ 10 frequently did not adhere to required research practices (Table 2). Of the 24 publications in high-impact-factor journals, 16 (67%) had at least 1 violation and 9 (38%) had 2 violations. These rates were higher among the 96 studies sampled from publications in journals with an impact factor <10 , in which nearly 90% (86 of 96 studies) had at least 1 or more violations of required research practices (absolute difference, 23%, 95% CI 0%–45%; $P = .01$); two-thirds of all studies (65 studies) had 2 or more practices that were not appropriate for data from the NIS (absolute difference, 30%, 95% CI 7%–52%; $P = .009$). Moreover, compared with studies published in journals with impact factor ≥ 10 , those published in journals with an impact factor <10 had a higher number of violations of required research practices per study (median [IQR] of violations per study, 1[0 to 2] for impact factor ≥ 10 vs 2 [1 to 3], $P = .006$) (Table 2). The nature of the violations followed a similar pattern in studies with an impact factor of <10 or ≥ 10 (Table 3).

Studies using data from the NIS were cited a median of 4 times (IQR: 0 to 9) during a median follow-up period of 16 months since their publication (Table 4). A higher number of violations was associated with fewer citations, with a median of 6.5 citations (IQR 2 to 12) among studies with zero or 1 violations compared with a median of 2 citations (IQR: 0 to 7) for studies with 2 violations of required practices ($P = .01$). Further, among studies published in journals with an impact factor <10 , while the median number of citations was higher in studies with zero to 1 violations (median [IQR]: 4 [0 to 7]) compared with studies with 2 or more citations (median [IQR]: 2 [0 to 6]), these differences were not statistically significant ($P=.49$). For studies published in journals with an impact factor ≥ 10 , there was no significant difference in the median number of citations among studies with zero to 1 violations, compared with those with 2 or more violations (Table 4).

In the simulation of NIS data for CABG trends in the years 2010–2013, the use of incorrect weights for the years 2010–2011 would erroneously suggest that there was a steep decline in CABG volumes over this period (slope of linear regression line \pm standard error, -6342 ± 1034 per year). This contrasts with a more gradual actual decline when the correct weights are used (-2366 ± 156 per year, P for difference in slopes $.019$) (Figure 2).

DISCUSSION

In this overview of a random sample of 120 published studies drawn from 1082 unique studies published using data from the NIS during the years 2015–2016, 85% of studies did not adhere to 1 or more required research practices. Most studies did not account for the complex design of the sample in their analyses and therefore did not address the effects of sampling error, clustering, and stratification of data on the interpretation of their results. Similarly, 80% of the studies did not account for major changes in the data structure of the NIS over time and were thus likely to ascribe effects of data changes to temporal changes in the disease condition of interest. Investigations using data from the NIS also frequently misinterpreted the NIS as a patient-level dataset rather than a record of hospitalization, thereby inflating prevalence estimates. Furthermore, 52% of the studies extrapolated information from the available data to infer in-hospital events using non-specific secondary diagnosis codes. Several studies performed state-, hospital- and physician-level analyses in conditions for which such analyses would not be considered appropriate. The quality issues identified were pervasive in the literature based on the NIS, even among articles published in high-impact journals. In addition, despite limited follow-up, publications based on the NIS have been frequently cited, regardless of the number of violations of required research practices.

Within the NIS, the limited agreement between robust official recommendations and actual practice raises questions about the inferences that have been made from many published investigations. Further, it raises questions about the reasons for the non-adherence of the investigations with the research practices required by AHRQ. First, the data can be obtained by anyone with access to a computer and there is no requirement for statistical training or analytic support for those wishing to use the database to conduct investigations. While the NIS has robust documentation and tutorials, such resources may not be known to researchers. Second, even experienced investigators may incorrectly design studies or

misinterpret data from the NIS. In particular, the sampling strategy ensures representativeness, but requires an understanding of more advanced survey-analysis procedures that can appropriately account for the stratification and clustering of data from the NIS. The NIS has a data structure that is similar to other common administrative datasets, like Medicare, wherein each observation represents a discrete healthcare encounter and includes a set of administrative diagnosis and procedure codes that correspond to that encounter. However, the NIS data include several additional variables that identify the sampling strata and clusters for each observation, which are necessary for its appropriate use. Further, features such as the inability to track patients longitudinally, or obtain estimates for states or physicians, require that investigators invest the time to understand the nuances of data analysis using the NIS as opposed to transposing methodology from analyses of more conventional administrative datasets such as Medicare. Therefore, a careful review of its required practices is essential to ensure its appropriate use. In addition, it is critical that investigators ensure that the NIS represents the most appropriate database for their research question, and not predicate their decision on its easy accessibility compared with other data sources.

While the research practices assessed in this study are specific to the NIS, the findings do not impugn the NIS or the open-source science platform. Rather, these findings highlight the possibility that lack of adherence to required research practices could undermine the potential of this national resource and that the conscientious dissemination of information, such as that provided by AHRQ, may not be sufficient to address the problem. The use of checklists and standardized reporting of adherence to standards within publications could be one means of promoting high-quality studies.²² However, such checklists would need to incorporate database-specific standards.

This study has several limitations. First, the study only includes an evaluation of studies from a recent 2-year period, and the quality of investigations using the NIS in preceding years may be different. However, given the forward-feeding nature of science, and limited familiarity with the NIS that was observed in the studies examined, superior quality in earlier years would not be expected. Second, the present study performed a limited evaluation of study quality focused on 7 NIS-specific practices but did not evaluate other aspects of quality. Therefore, the study does not suggest that investigations that followed all the NIS practices that we examined are of the highest quality, given the potential for additional limitations and incorrect research practices. These include inflating the generalizability of the study's population and/or its outcomes to those outside of an inpatient clinical setting. Third, the present study did not independently examine the direct implications of the identified research practices in the context of their specific field. However, it would be prudent to confirm the results of studies using data from the NIS that did not adhere to research practices, particularly those that are of major importance to a research field. Fourth, this study was not designed to compare quality in different types of studies or in other publicly available databases, and an independent assessment of studies published using other data sources is needed. Fifth, while the present study followed objective criteria and performed multiple independent evaluations of the studies, there is a potential for misclassifying studies if the authors did not report the methods clearly. Sixth, the present study uses study citations as a marker of the association between a publication

using the NIS database and subsequent investigations in the field; however, it does not specifically address the nature of these study citations.

Conclusions

In this study of 120 recent publications that used data from the NIS, the majority did not adhere to required practices. Further research is needed to identify strategies to improve the quality of research using the NIS and assess whether there are similar problems with use of other publicly available data sets.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Anne Elixhauser, PhD from the Agency for Healthcare Research and Quality for advising us on the design and application of the study evaluation checklist used in this study. Dr. Elixhauser did not receive compensation for this work. The contents of the study do not represent the official views of the Agency for Healthcare Research and Quality and are not endorsed by any Federal agency.

Funding: Dr. Khera is supported by the National Heart, Lung, and Blood Institute (5T32HL125247-02) and the National Center for Advancing Translational Sciences (UL1TR001105) of the National Institutes of Health. Dr. Girotra (K08 HL122527) and Dr. Chan (R01HL123980) are supported by funding from the National Heart, Lung, and Blood Institute. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Dr. Krumholz is a recipient of research agreements from Medtronic and Johnson & Johnson (Janssen), through Yale, to develop methods of clinical trial data sharing; is the recipient of a grant from Medtronic and the Food and Drug Administration, through Yale, to develop methods for post-market surveillance of medical devices; works under contract with the Centers for Medicare & Medicaid Services to develop and maintain performance measures that are publicly reported; chairs a cardiac scientific advisory board for UnitedHealth; is a participant/participant representative of the IBM Watson Health Life Sciences Board; is a member of the Advisory Board for Element Science and the Physician Advisory Board for Aetna; and is the founder of Hugo, a personal health information platform.

References

1. Shah RU, Bairey Merz CN. Publicly Available Data: Crowd Sourcing to Identify and Reduce Disparities. *J Am Coll Cardiol*. 2015; 66(18):1973–1975. [PubMed: 26515999]
2. Khera R, Krumholz HM. With great power comes great responsibility: "Big data" research from the National Inpatient Sample. *Circ Cardiovasc Qual Outcomes*. 2017; 10(7):e003846. [PubMed: 28705865]
3. NIS Database Documentation Archive. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Jun. 2016 at www.hcup-us.ahrq.gov/db/nation/nis/nisarchive.jsp [Accessed September 25, 2017]
4. HCUP Databases. Healthcare Cost and Utilization Project - Overview of the National (Nationwide) Inpatient Sample (NIS). Agency for Healthcare Research and Quality; Rockville, MD: Nov. 2016 at www.hcup-us.ahrq.gov/nisoverview.jsp [Accessed September 25, 2017]
5. NIS Description of Data Elements. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Nov. 2016 at www.hcup-us.ahrq.gov/db/nation/nis/nisdde.jsp [Accessed September 25, 2017]
6. Producing National HCUP Estimates - Accessible Version. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Oct. 2015 at www.hcup-us.ahrq.gov/tech_assist/nationalestimates/508_course/508course.jsp [Accessed September 25, 2017]

7. HCUP Frequently Asked Questions. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Dec. 2016 at www.hcup-us.ahrq.gov/tech_assist/faq.jsp [Accessed September 25, 2017]
8. HCUPnet. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: 2017. at <https://hcupnet.ahrq.gov/#setup> [Accessed September 25, 2017]
9. HCUP Publications Search. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Feb. 2010 at www.hcup-us.ahrq.gov/reports/pubsearch/pubsearch.jsp [Accessed September 25, 2017]
10. 2011 Introduction to the NIS. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Jul. 2016 at www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2011.jsp [Accessed September 25, 2017]
11. Why the NIS should not be used to make state-level estimates. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Jan. 2016 at www.hcup-us.ahrq.gov/db/nation/nis/nis_statelevelestimates.jsp [Accessed September 25, 2017]
12. 2014 Introduction to the NIS. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Dec. 2016 at www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2014.jsp [Accessed September 25, 2017]
13. HCUP NIS Description of Data Elements. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: 2014. at www.hcup-us.ahrq.gov/db/vars/mdnum2_r/nisnote.jsp [Accessed September 25, 2017]
14. [Accessed September 25, 2017] Nationwide Inpatient Sample Redesign: Final Report. Apr. 2014 at <https://www.hcup-us.ahrq.gov/db/nation/nis/reports/NISRedesignFinalReport040914.pdf>
15. Khera R, Cram P, Girotra S. Letter by Khera et al. regarding, "Impact of annual operator and institutional volume on percutaneous coronary intervention outcomes: A 5-year United States experience (2005–2009)". *Circulation*. 2015; 132(5):e35. [PubMed: 26240269]
16. HCUP NIS Description of Data Elements. Healthcare Cost and Utilization Project (HCUP). DRG_NoPOA - DRG in use on discharge date, calculated without POA. Agency for Healthcare Research and Quality; Rockville, MD: Sep. 2008 at www.hcup-us.ahrq.gov/db/vars/drg_nopoa/nisnote.jsp [Accessed September 25, 2017]
17. Healthcare Cost and Utilization Project (HCUP) Methods Series. [Accessed September 25, 2017] The Case for the POA Indicator. May. 2011 at https://www.hcup-us.ahrq.gov/reports/methods/2011_05.pdf
18. HCUP Elixhauser Comorbidity Software. Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality; Rockville, MD: Jun. 2017 at www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp [Accessed September 25, 2017]
19. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998; 36(1):8–27. [PubMed: 9431328]
20. Healthcare Cost and Utilization Project (HCUP) Methods Series. [Accessed September 25, 2017] Applying AHRQ Quality Indicators to Healthcare Cost and Utilization Project (HCUP) Data for the Eleventh (2013) NHQR and NHDR Report. Mar. 2012 at https://www.hcup-us.ahrq.gov/reports/methods/2012_03.pdf
21. Changes in the NIS Sampling and Weighting Strategy for 1998. Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality; Rockville, MD: Oct. 2015 January 2002. at https://www.hcup-us.ahrq.gov/db/nation/nis/reports/Changes_in_NIS_Design_1998.pdf [Accessed September 25, 2017]
22. Motheral B, Brooks J, Clark MA, Crown WH, Davey P, Hutchins D, Martin BC, Stang P. A checklist for retrospective database studies--report of the ISPOR Task Force on Retrospective Databases. *Value Health*. 2003; 6(2):90–97. [PubMed: 12641858]

KEY POINTS**Question**

Do secondary analyses of publicly-accessible datasets adhere to required research practices?

Findings

In a representative sample of 120 studies from the National Inpatient Sample published during 2015–2016, despite accompanying documentation of the required methodology, 85% of studies violated 1 or more required research practices pertaining to data structure, analysis, or interpretation.

Meaning

Lack of adherence to methodological standards was prevalent in published research using the National Inpatient Sample database.

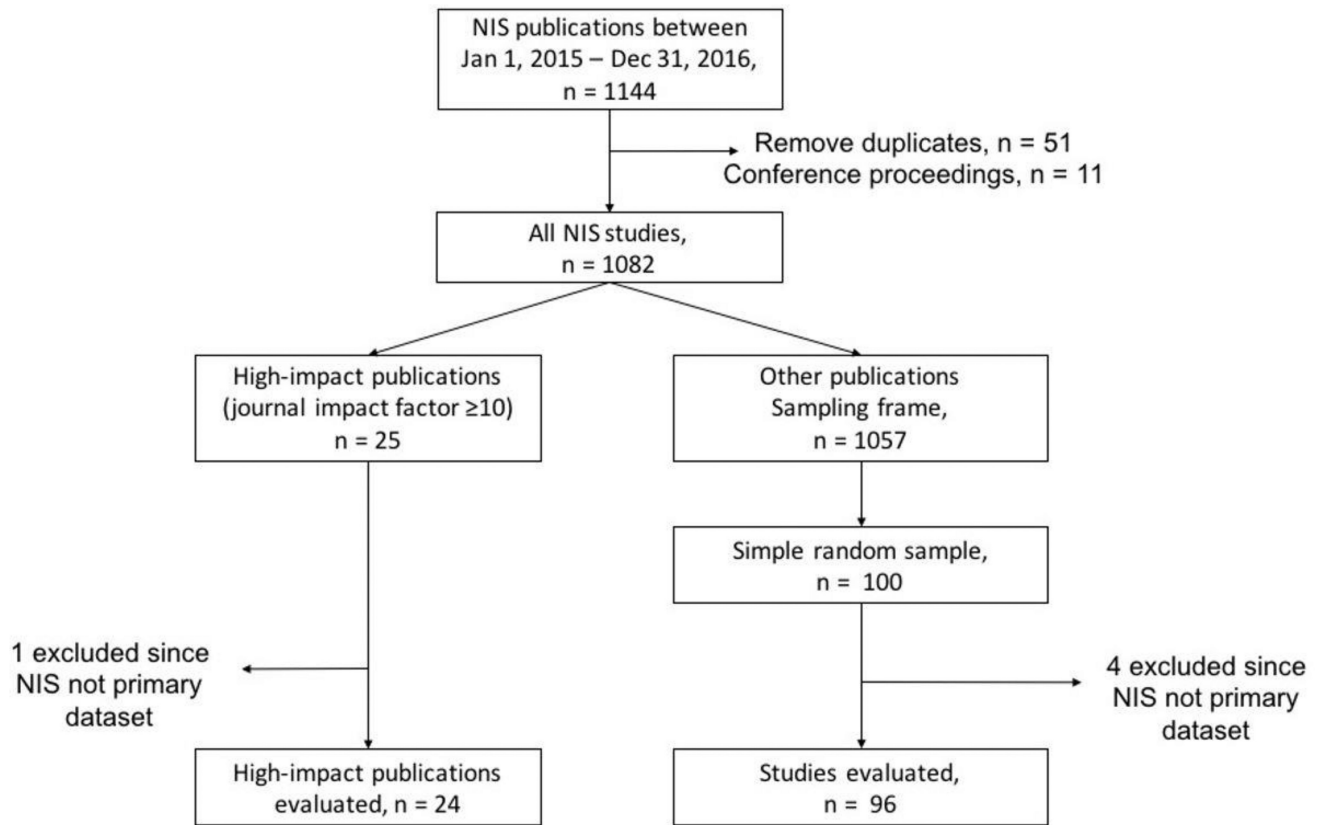


Figure 1. Study Selection Flowsheet

A sample of 120 studies was evaluated from the 1082 publications using the NIS during 2015–2016.

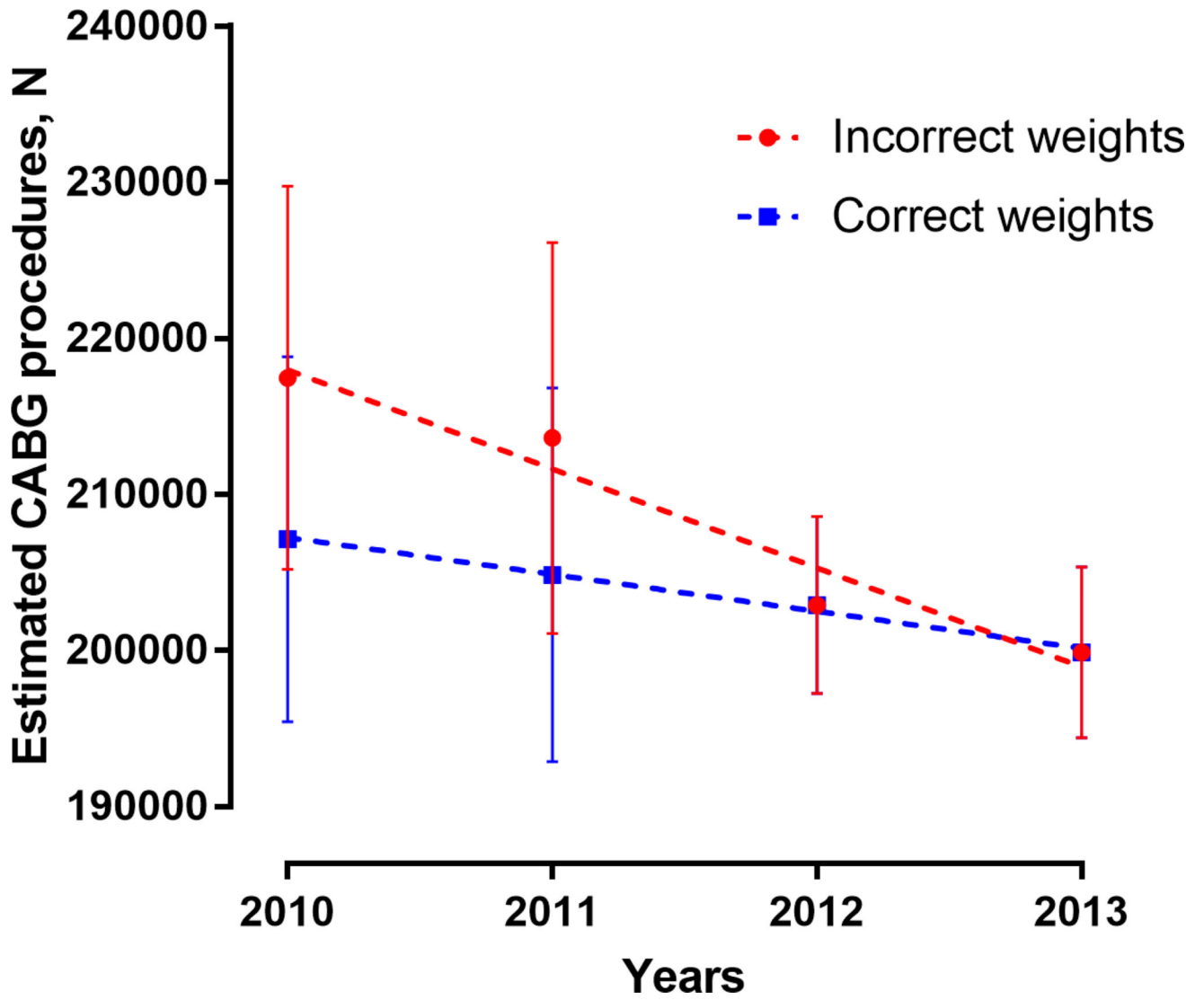


Figure 2. Simulation of Data to Demonstrate Incorrect Assessment of Hospitalization-level Trends

National-level trends in the number of coronary artery bypass grafting (CABG) hospitalizations were evaluated using the National Inpatient Sample data from the years 2010 through 2013. Trends were assessed using discharge weights for 2010–2011 that accounted for changes in data structure over time (*correct weights*), compared to those without this adjustment (*incorrect weights*).

Table 1

Required research practices for research that used the National Inpatient Sample.

Domain	Research Practice No.	Required Research Practices For Conducting Research using the National Inpatient Sample
Data Interpretation	1	Identifying observations as hospitalization events rather than unique patients ^{4,12}
Research design	2	Not performing state-level analyses ¹¹
	3	Limiting hospital-level analyses to data from years 1988–2011 ^{10,14}
	4	Not performing physician-level analyses ^{13,15}
	5	Not using non-specific secondary diagnosis codes to infer in-hospital events ^{16–20}
Data analysis	6	Using survey-specific analysis methods that account for clustering, stratification and weighting ⁶
	7	Accounting for data changes in trend analyses spanning major transition periods in the dataset (1997–1998 and 2011–2012) ^{14,21}

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Total number of violations of required research practices per study, for publications in 2015–2016 using the National Inpatient Sample.

Table 2

Number of violations of required practices	Overall (N = 120)	Journal impact factor <10 (N = 96)	Journal impact factor 10 (N = 24)	Estimates for the universe of NIS studies, N=1082*	
	Number (%) of studies			Projected number of studies (95% CI)	Projected percentage of studies (95% CI)
0	18 (15.0%)	10 (10.4%)	8 (33.3%)	114 (50 to 177)	10.5% (4.7 to 16.4)
1	28 (23.3%)	21 (21.9%)	7 (29.2%)	229 (143 to 315)	21.2% (13.2 to 29.1)
2	36 (30.0%)	32 (33.3%)	4 (16.7%)	342 (244 to 440)	31.6% (22.6 to 40.7)
3	30 (25.0%)	25 (26.0%)	5 (20.8%)	269 (178 to 360)	24.9% (16.5 to 33.3)
4	8 (6.7%)	8 (8.3%)	0 (0.0%)	85 (28 to 142)	7.8% (2.5 to 13.1)

P value = 0.025 (Chi-square test for comparison of studies in journals with impact factor <10 vs 10)

* Results reflect weighted estimates that applied to all studies using NIS data during the study period. An additional estimated 43 studies (95% CI 2 to 85) or 4.0% (95% CI 0.2% to 7.8%) correspond to the 5 studies excluded for using the NIS as a secondary source of data.

NIS, National Inpatient Sample

Rate of violations of required research practices for publications in 2015–2016 that used the National Inpatient Sample.

Table 3

Research Practice No.	Required Research Practices	Overall (N = 120)		Estimates for the universe of NIS studies (N = 1082)		Impact factor <10 (N = 96)		Impact factor 10 (N = 24)		P-value* (Fisher's exact test)
		Studies violating practice, n (%)	Eligible, N	Projected number of studies (95% CI)	Projected percentage of studies (95% CI) [†]	Studies violating practice, n (%)	Eligible, N	Studies violating practice, n (%)	Eligible, N	
1	Identifying observations as hospitalization events rather than unique patients	45 (37.5%)	120	437 (334 to 541)	40.4% (30.9 to 50.0)	41 (42.7%)	96	4 (16.7%)	24	0.02
2	Not performing state-level analyses	10 (8.3%)	120	77 (23 to 131)	7.1% (2.1 to 12.1)	7 (7.3%)	96	3 (12.5%)	24	0.42
3	Limiting hospital-level analyses to data from years 1988–2011	2 (1.8%)	17	12 (0 to 33)	8.2% (0.0 to 22.5) [‡]	1 (7.7%)	13	1 (25.0%)	4	0.43
4	Not performing physician-level analyses	3 (2.5%)	120	32 (0 to 68)	2.9% (0.0 to 6.2)	3 (3.1%)	96	0 (0.0%)	24	1.0
5	Not using non-specific secondary diagnosis codes to infer in-hospital events	62 (51.7%)	120	588 (484 to 693)	54.4% (44.7 to 64.0)	55 (57.3%)	96	7 (29.7%)	24	0.02
6	Using survey-specific analysis methods that account for clustering, stratification and weighting	79 (65.8%)	120	739 (642 to 837)	68.3% (59.3 to 77.3)	69 (71.9%)	96	10 (41.7%)	24	0.008
7	Accounting for data changes in trend analyses spanning major transition periods in the dataset (1997–1998 and 2011–2012)	21 (77.8%)	27	174 (97 to 251)	79.7% (62.5 to 97.0) [§]	16 (80.0%)	20	5 (71.4%)	7	0.63

* For comparison of studies in journals with impact factor <10 vs 10

[†] Unless otherwise specified, results reflect estimated percentage of all 1082 studies using NIS data during the study period with a given violation. Therefore, an estimated 43 studies (95% CI 2 to 85) that correspond to the 5 studies excluded for using the NIS as a secondary source of data, are represented in the denominator for these percentages.

[‡] Percentage of the estimated 141 studies performing hospital-level analyses. These studies correspond to the 17 sampled studies that performed these analyses.

[§] Percentage of the estimated 218 studies performing analyses spanning major data transitions. These studies correspond to the 17 sampled studies that performed these analyses.

NIS, National Inpatient Sample

Table 4

Citation count for publications in 2015–2016 that used the National Inpatient Sample, by number of violations of required research practices.*

Group	Number of publications	Citations [†] , median (interquartile range)	P-value for differences in citations by number of violations
All studies	120	4 (0 to 9)	N/A
All studies, by number of violations, N = 120			<.001 [‡]
0 practice violations	18	9.5 (7 to 24)	
1 practice violation	28	4.5 (0.5 to 7.5)	
2 practice violations	36	4.5 (1 to 10)	
3 practice violations	30	2 (0 to 6)	
4 practice violations	8	1 (0 to 1)	
All studies, by number of violations (2 groups), N = 120			0.01 [§]
0 to 1 practice violations	46	6.5 (2 to 12)	
2 practice violations	74	2 (0 to 7)	
Impact factor <10, by number of violations, N = 96			0.49 [§]
0 to 1 practice violations	31	4 (0 to 7)	
2 practice violations	65	2 (0 to 6)	
Impact factor >10, by number of violations, N = 24			0.08 [§]
0 to 1 practice violations	15	18 (8 to 35)	
2 practice violations	9	6 (1 to 19)	

* Median follow up from publication to the assessment of citations were: All studies - 16 months (IQR: 9 to 21 months, impact factor <10 – 14 months (IQR: 8 to 21 months), impact factor ≥ 10 – 17 months (IQR: 12 to 22 months).

[†] Per Google Scholar (assessed on April 4, 2017)

[‡] Non-parametric regression for trend in number of citations with increasing number of violations, negative trend

[§] Wilcoxon sum-of-ranks test for citation differences between studies with 0 to 1 vs 2 violations

N/A Not applicable