JAMA | Original Investigation

Adherence to Methodological Standards in Research Using the National Inpatient Sample

Rohan Khera, MD; Suveen Angraal, MBBS; Tyler Couch, BS; John W. Welsh, BS; Brahmajee K. Nallamothu, MD, MPH; Saket Girotra, MD, SM; Paul S. Chan, MD, MSc; Harlan M. Krumholz, MD, SM

IMPORTANCE Publicly available data sets 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 through December 2016, a representative sample of 120 studies was systematically evaluated for adherence to practices required by AHRQ for the design and conduct of research using the NIS.

EXPOSURES None.

MAIN OUTCOMES AND MEASURES All studies were evaluated on 7 required research practices based on AHRQ's recommendations and compiled under 3 domains: (1) data interpretation (interpreting data as hospitalization records rather than unique patients); (2) research design (avoiding use in performing state-, hospital-, and physician-level assessments where inappropriate; not using nonspecific administrative secondary diagnosis codes to study in-hospital events); and (3) 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 or more required practices and 62% (n = 74) did not adhere to 2 or more required practices. An estimated 925 (95% CI, 852-998) NIS publications did not adhere to 1 or more required practices and 696 (95% CI, 596-796) NIS publications did not adhere to 2 or more required practices. A total of 79 sampled studies (68.3% [95% CI, 59.3%-77.3%]) among the 1082 NIS studies screened for eligibility did not account for the effects of sampling error, clustering, and stratification; 62 (54.4% [95% CI, 44.7%-64.0%]) extrapolated nonspecific 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 the changes. Among the 24 studies published in journals with an impact factor of 10 or greater, 16 (67%) did not adhere to 1 or more practices, and 9 (38%) did not adhere to 2 or more practices.

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.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Harlan M. Krumholz, MD, SM, 1 Church St, Ste 200, New Haven, CT 06510 (harlan.krumholz@yale.edu).

Publicly available data sets 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, the design properties of publicly available data sets may require specific analytic approaches. The National Inpatient Sample (NIS), which is a large administrative database produced by the Agency for Healthcare Research and Quality (AHRQ), 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 annually contribute data for all-payer health care utilization.[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 about 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 NIS data without adhering to these resources may contain inaccuracies.

Given the recent proliferation of research using NIS data, 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 using 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[9] and supplemented with data from bibliographic repositories, we identified 1082 unique studies (eAppendix 1 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 or greater and a simple random sample of 100 additional studies that were published in journals with an impact factor of less than 10 (**Figure 1**). The sampling of studies was performed using the SURVEYSELECT procedure in SAS 9.4; all studies (n = 1057) in journals with an impact factor of less than 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 ÷ 1057). The representativeness of the sample was assessed against the NIS universe for (a) distribution of studies across the

Key Points

Question Do secondary analyses of publicly accessible data sets 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 did not adhere to 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.

spectrum of journal impact factors, (b) the nature of the source journal (medical or surgical), and (c) the clinical field of the journals (medicine and medical subspecialties, surgery and surgical subspecialties, pediatrics, obstetrics and gynecology, or mental and 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 of less than 10. Because all studies with an impact factor of 10 or greater were selected, the corresponding sampling weight for each 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 eAppendices 2 and 3 in the Supplement. Adherence to these research practices is essential for drawing appropriate conclusions using data from the NIS and is 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, studies using the NIS were evaluated to determine whether observations were correctly portrayed 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 predefined 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% of discharges from 20% of hospitals in the United States to a national 20% sample of patients, precluding hospital volume-based analyses beyond data from

Figure 1. Flow of Study Selection for Studies Evaluated Using the National Inpatient Sample
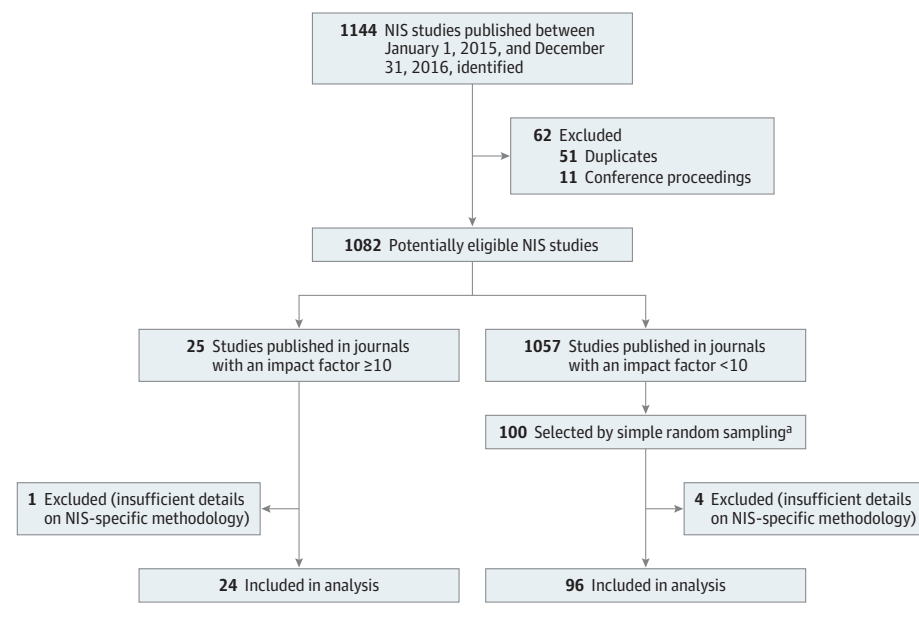
A sample of 120 studies was evaluated from the 1082 publications using the National Inpatient Sample (NIS) during 2015-2016.

[a] The sampling of studies was performed using the SURVEYSELECT procedure in SAS 9.4; all studies (n = 1057) in journals with an impact factor of less than 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 ÷ 1057).

1988-2011.[10,14] Therefore, studies were evaluated to determine if they 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, studies were evaluated to determine if the NIS was used to obtain physician-level estimates (practice 4).

Because the record of hospitalization in the NIS includes 1 principal and as many as 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 (eg, the Elixhauser comorbidity index) and complications (patient safety indices developed by AHRQ or secondary codes specific to postprocedure complications) be used.[18-20] Therefore, studies were evaluated to determine if nonspecific secondary diagnosis codes to infer in-hospital events were used (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—allowing weighting of estimates to generate national estimates with an accompanying measure of variance of the estimate.[6] Therefore, an evaluation was performed to determine if analyses used 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, studies spanning these 2 transition points were evaluated to determine if they 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 (eAppendix 4 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) were applicable to fewer studies. Practice 3 required that studies performing hospital-volume assessments be limited to data before 2012, and practice 7 required that studies performing trend analyses spanning transitions in the NIS make required modifications to their analyses. Before study evaluation, all investigators involved in data abstraction (S.A., T.C., J.W.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; 2012 and later). Each study was evaluated independently by 2 of 3 investigators (S.A., T.C and J.W.W.), and results were collated and confirmed by a fourth abstracter (R.K.). There was good interrater reliability (κ statistic, 0.88) and disagreements were resolved with mutual agreement, discussion with the senior author (H.M.K.), or both. 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 nonadherence in the universe of studies published using the NIS in 2015 and 2016, a survey methodology that accounted for stratified sampling of studies based on journal impact factor was used. For these analyses, a journal impact factor of 10 or greater or of less than

10 was used as the stratification variable and the corresponding sampling weights for these strata (1 for ≥10; 10.57 for <10) 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, the citation record was evaluated using Google Scholar citations on April 4, 2017. All analyses were repeated, after stratification of publications, and categorized as having an impact factor of 10 or greater or of less than 10. The χ^2 and Fisher exact tests were used to compare differences in categorical outcomes, and the nonparametric Wilcoxon rank-sum test and nonparametric regression analyses were used 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 (AHRQ-recommended 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 2 approaches were assessed using analysis of covariance.

All analyses were performed using SAS 9.4, 2-sided statistical tests, and a level of significance set at an α of .05. The study's use of NIS data was exempted from the purview of Yale University's institutional review board because the data were deidentified.

Results

Of the 125 publications in our initial cohort (all 25 studies published in journals with an impact factor of ≥10 and a random sample of 100 studies published in journals with an impact factor of <10), 5 studies were excluded because they used multiple data sets with limited information on NIS-specific methodology, precluding methodological evaluation (1 [4%] in a journal with an impact factor of ≥10 and 4 [4%] in journals with an impact factor of <10), 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, medical or surgical nature, and clinical field of the source journal (eFigures 1-3 in the Supplement). 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%]) did not adhere to 1 required research practice, and 74 (64.3% [95% CI, 55.0%-73.6%]) did not adhere to 2 or more practices (of which 36 [31.6%] 95% CI, 22.6%-40.7% did not adhere to 2 practices; 30 [24.9%] 95% CI, 16.5%-33.3% did not adhere to 3 practices; and 8 [7.8%] 95% CI, 2.5%-13.1% did not adhere to 4 or more practices; Table 2). Therefore, an estimated 925 (95% CI, 852-998) studies did not adhere to 1 or more required research practices, and 696 (95% CI, 596-796) studies did not adhere to 2 or more required research practices among the 1082 unique studies published using the NIS during 2015-2016 (Table 2).

Table 1. Required Research Practices for Studies Using the NIS

Research Practice No. by Domain	Required Research Practices for Conducting Studies Using the NIS
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	Avoiding use of nonspecific 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 data set (1997-1998 and 2011-2012)[14,21]

Abbreviation: NIS, National Inpatient Sample.

Table 2. Total Number of Instances of Nonadherence to Required Research Practices per Study for Publications in 2015-2016 Using the National Inpatient Sample (NIS)

No. of Instances of Nonadherence to Required Practices	Nonadherence, No. (%) of Studies			Estimates of Nonadherence for the Universe of NIS Studies (N = 1082)[a]	
	Overall (N = 120)	Journal Impact Factor <10 (n = 96)[b]	Journal Impact Factor ≥10 (n = 24)[b]	No. (95% CI)	% (95% CI)
0	18 (15.0)	10 (10.4)	8 (33.3)	114 (50-177)	10.5 (4.7-16.4)
1	28 (23.3)	21 (21.9)	7 (29.2)	229 (143-315)	21.2 (13.2-29.1)
2	36 (30.0)	32 (33.3)	4 (16.7)	342 (244-440)	31.6 (22.6-40.7)
3	30 (25.0)	25 (26.0)	5 (20.8)	269 (178-360)	24.9 (16.5-33.3)
≥4	8 (6.7)	8 (8.3)	0	85 (28-142)	7.8 (2.5-13.1)

[a] Reflect weighted estimates that applied to all studies using NIS data in the study period. An additional estimated 43 studies (95% CI, 2-85; [4.0% {95% CI, 0.2%-7.8%}]) correspond to the 5 studies excluded for having insufficient details on NIS-specific methodology.

[b] P value = .03 (χ^2 test for comparison of studies in journals with an impact factor of <10 vs ≥10).

Table 3. Rate of Instances of Nonadherence to Required Research Practices for Publications in 2015-2016 Using the National Inpatient Sample (NIS)

Research Practice No. and Description	Nonadherence, No. of Studies/Total No. (%)				Estimates of Nonadherence for the Universe of NIS Studies (N = 1082)	
	Overall	Journal Impact Factor <10	Journal Impact Factor ≥10	P Value[a]	No. (95% CI)	% (95% CI)[b]
1: Identifying observations as hospitalization events rather than unique patients	45/120 (37.5)	41/96 (42.7)	4/24 (16.7)	.02	437 (334-541)	40.4 (30.9-50.0)
2: Not performing state-level analyses	10/120 (8.3)	7/96 (7.3)	3/24 (12.5)	.42	77 (23-131)	7.1 (2.1-12.1)
3: Limiting hospital-level analyses to data from years 1988-2011	2/17 (11.8)	1/13 (7.7)	1/4 (25.0)	.43	12 (0-33)	8.2 (0.0-22.5)[c]
4: Not performing physician-level analyses	3/120 (2.5)	3/96 (3.1)	0/24	>.99	32 (0-68)	2.9 (0.0-6.2)
5: Avoiding use of nonspecific secondary diagnosis codes to infer in-hospital events	62/120 (51.7)	55/96 (57.3)	7/24 (29.7)	.02	588 (484-693)	54.4 (44.7-64.0)
6: Using survey-specific analysis methods that account for clustering, stratification and weighting	79/120 (65.8)	69/96 (71.9)	10/24 (41.7)	.008	739 (642-837)	68.3 (59.3-77.3)
7: Accounting for data changes in trend analyses spanning major transition periods in the data set (1997-1998 and 2011-2012)	21/27 (77.8)	16/20 (80.0)	5/7 (71.4)	.63	174 (97-251)	79.7 (62.5-97.0)[d]

[a] P values are for comparison of studies in journals with an impact factor of less than 10 vs 10 or greater using the Fisher exact test.

[b] Unless otherwise specified, results reflect estimated percentage of all 1082 studies using NIS data during the study period with a given act of nonadherence. Therefore, an estimated 43 studies (95% CI, 2-85) that correspond to the 5 studies excluded for having insufficient details on NIS-specific methodology, are represented in the denominator for these percentages.

[c] Value indicates the percentage of the estimated 141 studies performing hospital-level analyses. These studies correspond to the 17 sampled studies that performed these analyses.

[d] Value indicates the percentage of the estimated 218 studies performing analyses spanning major data transitions. These studies correspond to the 17 sampled studies that performed these analyses.

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 nonspecific 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]) had 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 of 10 or greater frequently did not adhere to required research practices (Table 2). Of the 24 publications in journals with high impact factors, 16 (67%) did not adhere to at least 1, and 9 (38%) did not adhere to 2 or more required research practices. These rates were higher among the 96 studies sampled from publications in journals with an impact factor of less than 10, in which nearly 90% (86 of 96 studies) had at least 1 or more instances of nonadherence to 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 an impact factor of 10 or greater, those published in journals with an impact factor of less than 10 had more instances of nonadherence to required research practices per study (median, 2 [interquartile range {IQR}, 1-3] vs 1 [IQR, 0-2] for journals with an impact factor of ≥10 [P = .006]) (Table 2). The nature of the nonadherences followed a similar pattern in studies with an impact factor of less than 10 vs 10 or greater (Table 3).

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

Table 4. Citation Count for Publications in 2015-2016 That Used the National Inpatient Sample by Instances of Nonadherence to Required Research Practices[a]

No. of Instances of Nonadherence	No. of Publications	Citations, Median (IQR)[b]	P Value for Difference[c]
Overall	120	4 (0-9)	
0	18	9.5 (7-24)	
1	28	4.5 (0.5-7.5)	
2	36	4.5 (1-10)	<.001[d]
3	30	2 (0-6)	
≥4	8	1 (0-1)	
Grouped by range	120		
0-1	46	6.5 (2-12)	.01[e]
≥2	74	2 (0-7)	
Journal impact factor <10	96		
0-1	31	4 (0-7)	.49[e]
≥2	65	2 (0-6)	
Journal impact factor ≥10	24		
0-1	15	18 (8-35)	.08[e]
≥2	9	6 (1-19)	

Abbreviation: IQR, interquartile range.

[a] Median (IQR) follow-up from publication to the assessment of citations were 16 months (IQR, 9-21) for all studies, 14 months (IQR, 8-21) for impact factors of less than 10, and 17 months (IQR, 12-22) for impact factors of 10 or greater.

[b] Accessed using Google Scholar on April 4, 2017.

[c] P value for difference in No. of citations between groups based on No. of instances of nonadherence.

[d] Nonparametric regression for trend in number of citations with increasing numbers of nonadherence to practice, negative trend.

[e] Calculated using the Wilcoxon rank-sum test for citation differences between studies with 0 to 1 instances of nonadherence to practice vs 2 or greater.

Figure 2. Simulation of Data to Demonstrate Incorrect Assessment of Hospitalization-Level Trends
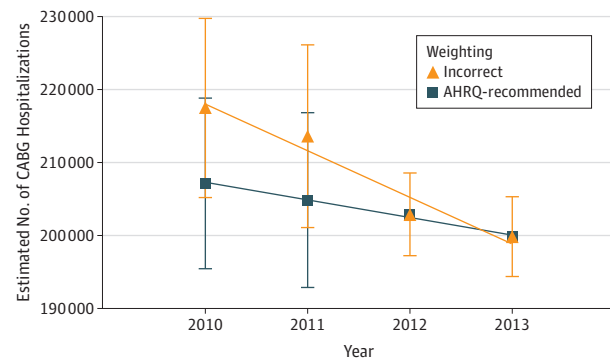
National-level trends in the number of coronary artery bypass grafting (CABG) hospitalizations were evaluated using National Inpatient Sample data from 2010 through 2013. Trends (slope of linear regression) were assessed using discharge weights for 2010-2011 that accounted for changes in data structure over time (Agency for Healthcare Research and Quality [AHRQ]–recommended weighting), compared with trends shown without this adjustment (incorrect weighting). Vertical bars indicate standard error.

studies with zero to 1 instance of nonadherence compared with those with 2 or more (Table 4).

In the simulation of NIS data for CABG trends in the years 2010-2013, the use of incorrect weighting 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 [SE], −6342 [1034] per year). This contrasts with a more gradual actual decline with use of correct weighting (−2366 [156] per year; [P value for difference in slopes, .02]; **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 data set 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 nonspecific 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 journals with high impact factors. In addition, despite limited follow-up, publications based on the NIS have been frequently cited, regardless of the number of required research practices not followed.

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 investigations' nonadherence to 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 individuals using the database to conduct investigations. Although the NIS has robust documentation and tutorials, the 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 to 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 data sets, such as Medicare, in which each observation represents a discrete health care encounter and includes a set of administrative diagnosis and procedure codes that correspond with 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 researchers invest the time to understand the nuances of data analysis using the NIS as opposed to transposing methodology from analyses of more conventional administrative data sets such as Medicare. Therefore, a careful review of the required practices is essential to ensure appropriate use of NIS data. 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.

Although 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.[2]

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 were examined are of the highest quality, given the potential for additional limitations and incorrect research practices (eg, inflating the generalizability of the study's population or its outcomes to those outside of an inpatient clinical setting or both). 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.

REFERENCES

1. Shah RU, Merz CNB. Publicly available data: crowd sourcing to identify and reduce disparities. *J Am Coll Cardiol*. 2015;66(18):1973-1975.

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.

3. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. NIS database documentation archive. Rockville, MD; June 2016. www.hcup-us.ahrq.gov/db/nation/nis/nisarchive .jsp. Accessed September 25, 2017.

4. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. Overview of the National (nationwide) Inpatient Sample (NIS). Rockville, MD; November 2016. www.hcup-us.ahrq .gov/nisoverview.jsp. Accessed September 25, 2017.

5. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. NIS description of data elements. Rockville, MD; November 2016. www.hcup-us.ahrq.gov/db/nation /nis/nisdde.jsp. Accessed September 25, 2017.

6. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. Producing national HCUP estimates. Rockville, MD; October 2015. www.hcup-us.ahrq.gov/tech_assist /nationalestimates/508_course/508course.jsp. Accessed September 25, 2017.

7. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. HCUP frequently asked questions. Rockville, MD; December 2016. www.hcup-us.ahrq.gov/tech _assist/faq.jsp. Accessed September 25, 2017.

8. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. HCUPnet. Healthcare Cost and Utilization Project. Free

health care statistics. Rockville, MD; 2017. https://hcupnet.ahrq.gov/#setup. Accessed September 25, 2017.

9. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. HCUP publications search. Rockville, MD; February 2010. www.hcup-us.ahrq.gov/reports/pubsearch /pubsearch.jsp. Accessed September 25, 2017.

10. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. Introduction to the HCUP Nationwide Inpatient Sample (NIS) 2011. Rockville, MD; July 2016. www.hcup-us.ahrq .gov/nation/nis/NIS_Introduction_2011.jsp. Accessed September 25, 2017.

11. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. Why the NIS should not be used to make state-level estimates. Rockville, MD; January 2016. www.hcup-us.ahrq .gov/db/nation/nis/nis_statelevelestimates.jsp. Accessed September 25, 2017.

12. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. Introduction to the HCUP National Inpatient Sample (NIS) 2014. Rockville, MD; December 2016. www.hcup-us.ahrq .gov/db/nation/nis/NIS_Introduction_2014.jsp. Accessed September 25, 2017.

13. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. NIS Description of Data Elements. Rockville, MD; 2014. www.hcup-us.ahrq.gov/db/vars/mdnum2 _r/nisnote.jsp. Accessed September 25, 2017.

14. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. HCUP Nationwide Inpatient Sample Redesign Report. April 2014. https://www.hcup -us.ahrq.gov/db/nation/nis/reports /NISRedesignFinalReport040914.pdf. Accessed September 25, 2017.

15. Khera R, Cram P, Girotra S. Letter by Khera et al regarding article, "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.

16. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality.NIS Description of Data Elements. DRG_NoPOA—DRG in use on discharge date, calculated without POA. Rockville, MD; September 2008. www.hcup-us .ahrq.gov/db/vars/drg_nopoa/nisnote.jsp. Accessed September 25, 2017.

17. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality.HCUP Methods Series. The case for the POA indicator: update 2011; report 2011-05; May 2011. https://www.hcup-us .ahrq.gov/reports/methods/2011_05.pdf. Accessed September 25, 2017.

18. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. Elixhauser comorbidity software, version 3.7. Rockville, MD; June 2017. 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.

20. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. HCUP Methods Series: Methods applying AHRQ quality indicators to healthcare cost and utilization project (HCUP) data for the eleventh (2013) NHQR and NHDR report #2012-03. Rockville, MD; March 2012. https://www.hcup-us.ahrq.gov/reports/methods /2012_03.pdf. Accessed September 25, 2017.

21. Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. Changes in the NIS sampling and weighting strategy for 1998. Rockville, MD; January 18, 2002. 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, et al. A checklist for retrospective database studies—report of the ISPOR Task Force on Retrospective Databases. *Value Health*. 2003;6(2): 90-97.