Clinical Orthopaedics
and Related Research®
A Publication of The Association of Bone and Joint Surgeons®

**Systematic Review**

# Most Orthopaedic Studies Using the National Inpatient Sample Fail to Adhere to Recommended Research Practices: A Systematic Review

**Teng L. Teng MBS, Mariano E. Menendez MD, Kanu Okike MD, Charles Cassidy MD, Matthew Salzler MD**

## Abstract

*Background* The National Inpatient Sample (NIS) database is accessible, inexpensive, and increasingly used in orthopaedic research, but it has complex design features that require nuanced methodological considerations for appropriate use and interpretation. A recent study showed poor adherence to recommended research practices for the NIS across a broad spectrum of medical and surgical fields, but the degree and patterns of nonadherence among orthopaedic publications remain unclear.

*Questions/purposes* In this study, we sought: (1) to quantify nonadherence to recommended research practices provided by the Agency for Healthcare Research and Quality (AHRQ) for using the NIS data in orthopaedic studies from 2016-2017; and, (2) to identify the most common nonadherence practices.

*Methods* We evaluated all 136 manuscripts published across the 74 orthopaedic journals listed on Scimago Journal & Country Rank that used the NIS from January 2016 through December 2017. Of those studies, 2% (3 of 136) were excluded because NIS was not used for analysis. The studies were evaluated for adherence to seven recommended research practices by the AHRQ: (1) identifying observations as hospitalization events rather than unique patients; (2) not performing state-level analyses; (3) limiting hospital-level analyses to data from year 1988-2011; (4) not performing physician-level analyses; (5) avoiding the use of nonspecific secondary diagnosis codes to infer in-hospital events; (6) using survey-specific analysis methods that account for clustering, stratification, and weighting; and (7) accounting for data changes in trend analyses spanning major transition periods in the data set (1997-1998 and 2011-2012).

*Results* Overall, 93% (124 of 133) of the studies did not adhere to one or more practices. For each of the research practices assessed, 80% (106 of 133) of the studies did not account for the clustering and stratification in survey design; 56% (75 of 133) implied records were unique patients rather than hospitalization events; 41% (54 of 133) inappropriately used secondary diagnosis codes to infer in-hospital events.

*Conclusions* Nearly all manuscripts published in orthopaedic journals using the NIS database in 2016 and 2017 failed to adhere to recommended research practices. Future research quantifying variations in study results on the basis

T. L. Teng, Tufts University School of Medicine, Boston, MA, USA

M. E. Menendez, C. Cassidy, M. Salzler, Department of Orthopaedics, Tufts Medical Center, Boston, MA, USA

K. Okike, Department of Orthopaedics, Moanalua Medical Center, Honolulu, HI, USA

M. Salzler ✉, Department of Orthopaedics, Tufts Medical Center, Boston, MA 02111, Email: msalzler@tuftsmedicalcenter.org

Wolters Kluwer

of adherence to recommended research practices would be of value.

*Clinical Relevance*  With the ubiquitous presence of large-database studies in orthopaedic journals, our work points to the importance of rigorous methodological appraisal when evaluating results, and encourages journals to require the use of the methodology checklists upon submission of such studies. More research is needed to determine whether deviations from recommended research practices actually lead to biased conclusions that affect patient care and policy-related decisions.

## Introduction

Large, publicly available databases have transformed the ability to conduct epidemiological research over the past few decades. These datasets provide a readily available source of nationwide data and are increasingly used for epidemiological, effectiveness, and safety outcomes studies [9, 14, 15]. With the rapid expansion of technologically advanced interventions that have grown faster than clinical outcome studies to support their use, national databases are appropriate resources with their extremely large pool of cases that investigators can use to demonstrate the effectiveness of new interventions [13]. Despite holding much potential, the unique design properties of these national databases require specific practices for their appropriate use [2, 3]. Within orthopaedic surgery, there has been a surge of publications using claims data from the National Inpatient Sample (NIS), the largest all-payer inpatient care database in the United States [1, 4, 12]. Compiled annually since 1988 by the Agency for Healthcare Research and Quality (AHRQ), the NIS includes administrative and demographic data from a nationally representative 20% sample of in-patient hospitalizations in the United States [6, 7].

Because the NIS contains a wealth of healthcare data and is relatively easy to access, many analyses may be performed without an understanding of, or adherence to, the complex structure and statistical design of the NIS. As the compiling body for NIS, the AHRQ has offered recommended research practices when using the NIS. Khera et al. [11] recently analyzed 120 studies published in 2015-2016 using the NIS from a broad spectrum of medical and surgical fields and found that most did not adhere to seven recommended practices provided by the AHRQ. However, it is unclear from their study how many of those studies were actually related to orthopaedic research. A more comprehensive analysis, with an exclusive focus on studies published in orthopaedic journals, will provide better insight into the degree and patterns of nonadherence in our field. This is especially important given the dramatic growth of orthopaedic research using NIS data in recent years [10].

In this study, we sought to (1) quantify nonadherence to recommended research practices provided by the Agency for Healthcare Research and Quality (AHRQ) for using the NIS data in orthopaedic studies from 2016-2017; and, (2) identify the most common nonadherence practices.

## Materials and Methods

### Study Selection

We performed a systematic evaluation of all manuscripts published in orthopaedic journals using the NIS database from January 1, 2016 through December 31, 2017 using criteria similar to Khera et al. [10]. We decided a priori to use the list of orthopaedic journals provided on Scimago Journal Rank (SJR), a commonly used bibliometric database in orthopaedic research [17]. Using SJR [16], we compiled the list of orthopaedic journals (Fig. 1). PubMed was used to search for articles published within those journals from 2016-2017 that contained search terms "Nationwide Inpatient Sample" or "National Inpatient Sample" in the title or the abstract section. After a review of all manuscripts, any paper that did not use NIS for primary data analysis was excluded.

A list of 74 journals in the field of orthopaedics was generated from SCImago under Orthopaedics and Sports Medicine in the United States as of June 2018 (see Appendix 1; Supplemental Digital Content 1, http://links.lww.com/CORR/A381). Of the identified 136 NIS (see Appendix 2; Supplemental Digital Content 2, http://links.lww.com/CORR/A382) studies over the time frame, 32% (43 of 136) studies were published in 2016 and 68% (93 of 136 were published in 2017. From the 136 NIS studies, 2% (3 of 136) were excluded because they did not use NIS data for primary analysis. As such, a total of 133 articles were evaluated for compliance to AHRQ recommendations when conducting NIS-based studies.

### Evaluation Criteria

The selected articles were evaluated on the basis of the seven AHRQ recommended research practices described by Khera et al. [11] (Table 1). These practices are as follows: The NIS is a record of inpatient hospital events, thus should not be portrayed as unique patients (Practice 1). As a national database, the NIS is constructed with complex survey design that ensures hospitalization records are representative at a national level. Thus, this database may not be representative for hospitalizations at the state level (Practice 2). Practice 3 addressed the major redesign in data gathering in 2011-2012 that changed from sampling 100% of discharges from 20% of hospitals in the United States to sampling 20% of discharges from all participating hospitals. Such change rendered
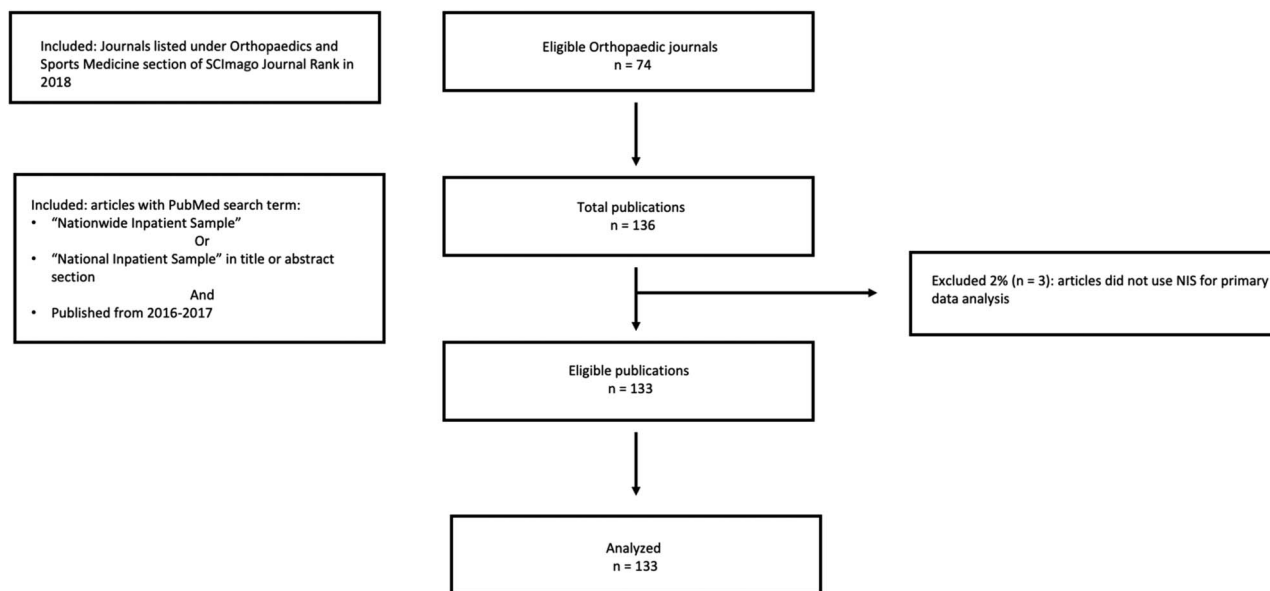
**Fig. 1** The STROBE checklist, shown here, details the study inclusion and exclusion criteria.

hospital-level analysis on data gathered afterwards inappropriate. Provider field code could be referring to either individual physician or physician groups, making the NIS not suitable in assessing physician-level volume (Practice 4). Practice 5 addressed the use of nonspecific secondary diagnosis codes which runs the risk of mixing comorbidity with true complication. Researchers should use validated algorithms such as Elixhauser or patient safety indices from AHRQ when performing such analysis (see Appendix 3; Supplemental Digital Content 3, http://links.lww. com/CORR/A383). Researchers should be using survey-specific analysis tools to interpret data from NIS (Practice 6). Methods such as clustering and stratifications must be factored in to obtain appropriate weights for the estimates. In 1998, AHRQ underwent another major redesign in sample gathering, including changing the definition of a hospital unit. Practice 7 checked if studies that performed trend analyses covering the major transition periods (1997-1998 and 2011-2012) accounted for such data change.

**Evaluation of Selected Studies**

Each article was evaluated for adherence to these practices. All study investigators (MS, MEM, KO, CC, TLT) reviewed a summary of the methodological design of the NIS. Each study was evaluated by one author (TLT), curated by another (MEM), and results were reviewed by three authors (MS, KO, CC). After training from the evaluation criteria, there was excellent agreement between TLT and MEM. Difference in evaluation occurred infrequently and consensus was reached every time.

**Distribution of Selected Studies**

A total of 133 papers were published in orthopaedic journals from 2016 to 2017 that used the NIS database. Of the seven practices evaluated, Practice 3 was applicable to one study and Practice 7 was applicable to 58 studies. The rest of the practices were evaluated against all of the studies.

**Table 1.** Required research practices for studies using the National Inpatient Sample[a]

| Research practice criteria | Required research practices for conducting studies using the NIS |
|---|---|
| 1 | Identifying observations as hospitalization events rather than unique patients |
| 2 | Not performing state-level analyses |
| 3 | Limiting hospital-level analyses to data from years 1988-2011 |
| 4 | Not performing physician-level analyses |
| 5 | Avoiding use of nonspecific secondary diagnosis codes to infer in-hospital events |
| 6 | Using survey-specific analysis methods that account for clustering, stratification, and weighting |
| 7 | Accounting for data changes in trend analyses spanning major transition periods in the data set (1997-1998 and 2011-2012) |

[a]Adapted from Khera et al. [11].

## Results

### Frequency of Nonadherence to AHRQ Recommended Research Practices in Orthopaedic Studies

There was a high prevalence of nonadherence to AHRQ recommended research practices in orthopaedic studies that used NIS published during 2016 to 2017. Of the entire cohort, 7% (9 of 133) satisfied all practices, 93% (124 of 133) did not follow one or more research practice, and 87% (116 of 133) did not follow two or more research practices (Table 2). Of the publications that did not follow two or more practices: 32% (42 of 133) did not follow two practices; 29% (38 of 133) did not follow three practices; 20% (27 of 133) did not follow four or more practices.

### Types of Nonadherence

The most common practice of nonadherence in the 80% (106 of 133) of the studies we evaluated was not accounting for the complex survey design of the NIS (Table 3). The second most common nonadherence in 56% (75 of 133) of the studies was the incorrect assumption that the NIS records represented individual patients instead of discharges, followed by 41% (54 of 133) of the studies inappropriately inferring in-hospital events from secondary diagnosis codes, and 35% (47 of 133) inappropriately performing state-level analyses (Table 3). There was one paper that performed estimates of procedure volume at hospital-level, which also inappropriately included NIS data after 2012. Of the 58 studies that included trend analysis spanning the major transition periods, 74% (43 of 58) did not account for the change in sampling redesign.

### Discussion

Owing to its accessibility, inexpensiveness, and relative ease of use, the adoption of NIS in orthopaedic research has increased dramatically over the past decade [10]. Careful

**Table 2.** Prevalence of nonadherence stratified by number of nonadherences found in each article.

| Number of instances of nonadherence to research practices | Nonadherence, % (number) of studies Overall (n = 133) |
|---|---|
| 0 | 7% (9 of 133) |
| 1 | 13% (17 of 133) |
| 2 | 32% (42 of 133) |
| 3 | 29% (38 of 133) |
| ≥ 4 | 20% (27 of 133) |

**Table 3.** Prevalence of nonadherence stratified by each of the seven research practices

| Research practice number (as listed in Table 1) | Nonadherence, % (number of studies/total number) Overall |
|---|---|
| 1 | 56% (75 of 133) |
| 2 | 35% (47 of 133) |
| 3 | 1 of 1 |
| 4 | 0 of 133 |
| 5 | 41% (54 of 133) |
| 6 | 80% (106 of 133) |
| 7 | 74% (43 of 58) |

attention to the methodology in database-mining studies is important to avoid biased conclusions. Recently, Khera et al. [11] reviewed 120 NIS studies across different medical and surgical fields and found that most did not adhere to recommended research practices. However, it is likely that only a very small fraction of those studies were orthopaedic-related. A more comprehensive analysis focused solely on orthopaedic studies is important to better understand the degree and patterns of nonadherence in our field. Therefore, this study sought (1) to determine nonadherence to AHRQ recommended research practices for using the NIS data in orthopaedic studies published between 2016-2017, and (2) to explore the most common nonadherence practices. We found that nearly all manuscripts published in orthopaedic journals using the NIS failed to adhere to recommended research practices. Our findings call for rigorous methodological considerations when evaluating orthopaedic claims database research, and the implementation of strategies to enhance methodological quality of such studies.

### Limitations

Our study was subject to several limitations that generate questions for future research. First, we were unable to determine whether nonadherence to the AHRQ recommendations actually causes harm or leads to biased decisions that negatively impact patient care. Although it is possible that some of these instances of nonadherence are just a numbers issue, we find it hard to believe that systematic errors in analysis design do not affect study results to some extent. Future research quantifying variations in study results on the basis of adherence to recommended research practices would be of value. Nevertheless, our study serves as warning that careful methodological consideration is warranted when using large claims databases such as the NIS. Second, there could be justifiable reasons for when a recommended research practice is not followed.

For instance, relying on using all available diagnosis codes for a given admission may not necessarily be an inappropriate practice if supported by validated coding algorithms or by using diagnosis codes in combination with procedure codes. However, we believe that this is unlikely to drastically change our findings. Third, there could be studies that followed the guidelines appropriately but failed to mention that in the methodology. Given the number of the papers and the fact that the description of the methodology is the most important aspect of a paper, we chose to evaluate the data based on what was presented in the journals. We expect this would have been rare, with minimal effect on our findings, as article methodologies generally involve detailed descriptions so that they can be replicated by other interested authors. Finally, the complete recommendations from AHRQ cover more information than the seven practices evaluated in this study, but we and others [11] believe these seven practices encompass most, if not all, of the methodological inaccuracies that one could make using the NIS data.

## Impact of Inappropriate Use of Database

The stark discrepancy between official NIS research practice recommendations and actual practice raises questions about the inferences that have been made from many published articles. The effects of each type of nonadherence can bias results differently depending on the study design, and multiple nonadherences likely compound the bias and their effects. In certain instances, nonadherence may not affect the final result. For example, when performing a cross-sectional analysis on the volume of certain procedures, distinguishing between hospital events and unique patients may not affect the result. However, failing to account for the clustering, sampling, and weighting of hospitals and hospital events would have potentially important effects on the validity of the results. Khera's analysis [11] found that 85% of studies failed to adhere to one or more recommended research practices. The proportion was similarly high (93%) in our analysis. Although the reasons for practice nonadherence to methodological standards were not within the scope of our study, we believe that these analyses may be adversely impacted by a lack of awareness of the AHRQ recommended research practices requirements [8]. All users of AHRQ data must complete a data use agreement training course before receiving the data, but some of the methodological nuances of using the data may be overlooked. To highlight the effect of appropriate NIS data analysis, a simulation of coronary artery bypass grafting volume from 2010 to 2013 accounting for the major changes in NIS data structure during that transition period (Practices 6 and 7) would result in a gradual decline (-2366 per year)

compared with a much steeper decline (-6342 per year) when the data structure was not accounted for [11]. The correct usage of the NIS database would facilitate drafting more effective health policies with updated national trends.

## Suggestions for Improvement

With the notion that NIS-derived findings are often used in patient care and health policy initiatives, our study calls for rigorous appraisal of methodology approaches of claims database studies and the development of efforts to minimize systematic errors in data analysis and interpretation. First, it is important for investigators to familiarize themselves with the correct usage of NIS data by reading the introductory material provided by Healthcare Cost and Utilization Project (HCUP), which is accessible online (https://www.hcup-us.ahrq.gov/nisoverview.jsp) [7]. Given that information on the HCUP website may be difficult to navigate, improvements in the user interface may result in improved adherence to research practices for investigators who are new to the NIS. Furthermore, the AHRQ could facilitate this process by providing an online forum for researchers to ask questions. Based on the most frequently asked questions, AHRQ could publish periodic user tips targeting the most common methodological concerns and issues when using the NIS data. From the journal's perspective, our study findings could encourage adoption of a brief methodology checklist based on the seven described criteria upon submission of NIS-based studies. Such checklist may not just improve adherence rates, but also make the job of the reviewers and editors easier. A more comprehensive checklist can be found on the AHRQ website (https://www.hcup-us.ahrq.gov/db/nation/nis/nischecklist.jsp) [5].

We found that most publications in orthopaedic journals using the NIS database fail to adhere to recommended research practices. Future research quantifying variations in study results on the basis of adherence to recommended research practices would be of value. Nonetheless, given the ubiquitous presence of database-mining research in orthopaedics, our work calls for rigorous methodological appraisal when evaluating results, and encourages journals to require the use of methodology checklists upon submission of such studies.

## References

1. Bedard NA, Pugely AJ, McHugh M, Lux N, Otero JE, Bozic KJ, Gao Y, Callaghan JJ. Analysis of outcomes after TKA: Do all databases produce similar findings? *Clin Orthop Relat Res*. 2018; 476:52-63.
2. Bekkers S, Bot AG, Makarawung D, Neuhaus V, Ring D. The National Hospital Discharge Survey and Nationwide Inpatient Sample: The databases used affect results in THA research. *Clin Orthop Relat Res*. 2014;472:3441-3449.

3. Bohl DD, Basques BA, Golinvaux NS, Baumgaertner MR, Grauer JN. Nationwide Inpatient Sample and National Surgical Quality Improvement Program give different results in hip fracture studies. *Clin Orthop Relat Res*. 2014;472:1672-1680.

4. Bohl DD, Russo GS, Basques BA, Golinvaux NS, Fu MC, Long WD, 3rd, Grauer JN. Variations in data collection methods between national databases affect study results: A comparison of the nationwide inpatient sample and national surgical quality improvement program databases for lumbar spine fusion procedures. *J Bone Joint Surg Am*. 2014;96:e193.

5. Healthcare Cost and Utilization Project AHRQ. Checklist for Working with the NIS. Available at: https://www.hcup-us.ahrq. gov/db/nation/nis/nischecklist.jsp. Accessed May 15, 2020.

6. Healthcare Cost and Utilization Project AHRQ. NIS database documentation archive. Available at: https://www.hcup-us.ahrq. gov/db/nation/nis/nisarchive.jsp. Accessed June 20, 2018.

7. Healthcare Cost and Utilization Project AHRQ. Overview of the National (nationwide) Inpatient Sample (NIS). 2016. Available at: https://www.hcup-us.ahrq.gov/nisoverview.jsp. Accessed June 20, 2018.

8. Healthcare Cost and Utilization Project AfHRaQ. Introduction to the HCUP national inpatient sample (NIS) 2017. 2019. Available at: https://www.hcup-us.ahrq.gov/db/nation/nis/NIS_Introduction_2017.jsp. Accessed January 5, 2019.

9. Ju DG, Rajaee SS, Mirocha J, Lin CA, Moon CN. Nationwide analysis of femoral neck fractures in elderly patients: A receding tide. *J Bone Joint Surg Am*. 2017;99:1932-1940.

10. Karlson NW, Nezwek TA, Menendez ME, Tybor D, Salzler MJ. Increased utilization of American administrative databases and large-scale clinical registries in orthopaedic research, 1996 to 2016. *J Am Acad Orthop Surg Glob Res Rev*. 2018;2:e076.

11. Khera R, Angraal S, Couch T, Welsh JW, Nallamothu BK, Girotra S, Chan PS, Krumholz HM. Adherence to methodological standards in research using the National Inpatient Sample. *JAMA*. 2017;318:2011-2018.

12. Khera R, Krumholz HM. With great power comes great responsibility: "Big data" research from the National Inpatient Sample. *Circ Cardiovasc Qual Outcomes*. 2017;10: e003846.

13. Levin PE. Apples, oranges, and national databases: commentary on an article by Daniel D. Bohl, MPH, et al.: "Variations in data collection methods between national databases affect study results: a comparison of the nationwide inpatient sample and national surgical quality improvement program databases for lumbar spine fusion procedures". *J Bone Joint Surg Am*. 2014;96: e198.

14. Molloy IB, Martin BI, Moschetti WE, Jevsevar DS. Effects of the length of stay on the cost of total knee and total hip arthroplasty from 2002 to 2013. *J Bone Joint Surg Am*. 2017;99: 402-407.

15. Nwachukwu BU, McLawhorn AS, Simon MS, Hamid KS, Demetracopoulos CA, Deland JT, Ellis SJ. Management of end-stage ankle arthritis: Cost-utility analysis using direct and indirect costs. *J Bone Joint Surg Am*.. 2015;97:1159-1172.

16. Scimago Journal & Country Rank. Orthopedics and sports medicine. 2017. Available at: https://www.scimagojr.com/journalrank.php?category=2732&country=US&page=1&total_size=80. Accessed June 20, 2018.

17. Sundaram K, Warren J, Anis HK, Klika AK, Piuzzi NS. Publication integrity in orthopaedic journals: the self-citation in orthopaedic research (SCOR) threshold. *Eur J Orthop Surg Traumatol*. 2019:1-7.