Arthroplasty Tips & Tricks: The Use of Big Clinical Databases for Orthopedic Research

Matthew Sloan, MD, MS
Neil Sheth, MD

Introduction

Research has changed. This is nowhere more evident than the ubiquitous appearance of articles published using data derived from big clinical databases across the most popular research journals in every field of medicine. For example, the National Surgical Quality Improvement Project (NSQIP) database, which is produced by the American College of Surgeons, reports that its database has been used to publish over 1,100 articles in peer-reviewed journals as of June 30, 2016.[1]

There are numerous big clinical databases available, both administrative and clinically derived, which use data collected from clinical settings across the United States. The administrative databases derive data from payments and claim information which include: Medicare claims database, Hospital Cost and Utilization Project (HCUP), and University Healthsystem Consortium. The clinical databases derive data from review of direct patient information which include: Automated Central Tumor Registry, NSQIP, National Trauma Data Bank, National Cancer Database, and the Surveillance, Epidemiology, and End Results (SEER) Program. There are also innumerable state-based and institution-based administrative and clinical databases and registries in use.[2]

What is driving the use of these big clinical databases? The obvious answer is time and money. While randomized controlled trials represent the gold standard of clinical research, they require enormous financial investment and time to be performed properly. Consider the requirements to perform a randomized trial that hopes to evaluate the superiority of low molecular weight heparin versus aspirin in prevention of venous thromboembolic events following total joint replacement. How many patients must be enrolled in such a trial? In order to evaluate this complication with sufficient power to detect a difference between these interventions, if one should exist, thousands of patients must be enrolled. If this study were conducted at a single institution, enrollment on that scale would take several years. A multi-center collaborative effort would require additional funds for oversight among participant sites and quality of data collection. Who is going to collect and analyze all this data? Research assistants and analysts need to be compensated to organize and interpret the collected data.

The performance of this rigorous investigation is essential to the scientific and medical community. However, before such investments can be made, preliminary research must establish a basis for embarking on a project of this scale. Herein lies the fundamental role for the big clinical database: readily available data on millions of patients that can be queried in minutes with minimal cost.

How to Perform Research Using Big Clinical Databases

The first step to beginning any research project is asking a question. The question should be informed by a rigorous literature review to determine what is known on the topic and what questions remain. Once the question is identified, one can determine if using a big clinical database is an option. All databases have their own benefits and limitations. It is imperative to understand the limitations of a database prior to acquiring it for research purposes. To illustrate the importance of understanding database limitations, we will walk through the process of using a big clinical database to answer the following question:

> How has the volume of primary and revision Total Joint Arthroplasty (TJA) changed in the United States since the year 2000?

This question is ideally suited for a big clinical database. It is not a question that can be answered by prospective study. Furthermore, the answer to this question may generate future questions that may be evaluated initially with data already present in the database. This work may then be used to inform a prospective study to evaluate these subsequent questions.

Identify the Database

To identify the appropriate database to answer this question, we will require one that has nationally collected data. This rules out institutional and state databases. We are interested in TJA performed among all patients, bearing in mind that a subsequent question may assess demographic (age, sex, race, socioeconomic) or clinical (obesity, diabetes, comorbidity index) variations among patients over time. So, Medicare claims database would be a poor choice, as it does not provide comprehensive data for patients under 65-years-old, a rapidly growing segment of TJA recipients.

One tip is to evaluate what database has been used to ask similar questions in the past. We reviewed a highly-cited article by Kurtz and colleagues that predicts expected growth in primary and revision TJA between 2005 to 2030.[3] To build their prediction model, the HCUP National Inpatient Sample (NIS) was used. This database collects patient data from a nationally representative sample of 1,000 hospitals and includes data on over 8 million patient discharges with demographic and clinical variables.[4,5] Patient data is not limited by age or region, which makes this database an ideal candidate for following national trends in clinical procedures.

Methods

Access to national databases varies in cost and restrictions of use. For our preferred database, the HCUP NIS, one must complete an online course and a data use agreement. After completion, each year of the database is available for purchase. The files are large, 5-10 GB per year, in a compressed file format.

After downloading the database files, the majority of the work involved in accessing the data is unzipping and merging the files. Data files of this scale, with roughly 8 million patient observations per year, cannot be opened using standard spreadsheet software. Common statistical packages with this capability include SAS, SPSS, Stata, and R. Depending upon familiarity, these program are more or less user friendly to those with a statistical programming background. However, R is the least user friendly and will require significant familiarity with the coding procedures even to upload the files into the program for analysis. Stata, on the other hand, uses numerous dropdown menus to enabling a novice to upload files and perform analyses with minimal prior experience.

We used Stata for our analysis. The steps for accessing data are outlined in Figure 1. Briefly, after download, each file set must be unzipped into a file type readable by the statistics program (an ASCII file in this case). Once all the files have been converted to ASCII format the files can be read into Stata and converted to a data (.DTA) file. The distributor of HCUP NIS provides coding to add labels to the variables. The various raw data files can then be merged so that the dataset with the patient procedures can be combined with the patient comorbidities and hospital characteristics files, for example. Once the data has been organized, it is ready for analysis.

To start, it is recommended to run simple means on patient ages and other demographic variables for comparison to reference values. This ensures that all the data was uploaded and merged correctly. If these values match, you can proceed to perform the analysis of interest. In our case, we wanted to know about trends in TJA and revision TJA since 2000. Each patient admission lists up to 15 procedure codes obtained from the patient discharge record. Using the 2000-2014 database years, we coded for all patients who underwent total

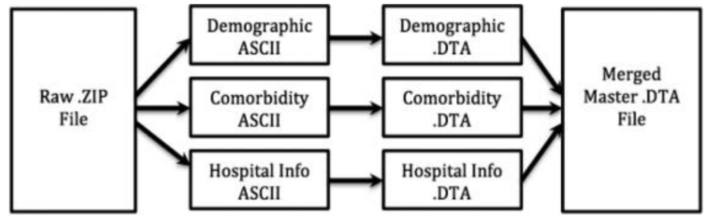

Figure 1. Process for unzipping and merging big clinical data file for use in statistical analysis programs.

hip arthroplasty (THA), total knee arthroplasty (TKA), revision THA, or revision TKA procedures in each year. The HCUP NIS database provides a representative weighted sample of 20% of hospital discharges from hospitals in the United States. Using each hospital's weighted average data, a national estimate for each procedure can be determined and cross-referenced against distributor data for sample procedures.

Once the trend data is obtained, outside data is combined with the raw volume data. We obtained data from the American Association of Orthopaedic Surgeons (AAOS) Orthopaedic Surgeon Census[6] to determine the change in annual cases per surgeon on a regional and national scale.

Results

Analysis of overall TJA demonstrates an increase in THA from 278,596 in 1997 to 523,280 in 2014. TKA procedures increased from 316,257 in 1997 to 752,941 in 2014. (Figure 2) Over the time period from 2000-2014 revision THA rose from 34,493 to 50,425. During the same period, revision TKA increased from 24,763 to 63,205, surpassing the number of revision THA performed annually. (Figure 3) From 2004 to 2016 the total number of orthopedic surgeon in the United States rose from 17,486 to 29,585. (Figure 4)

The volume of TJA per surgeon and revision TJA per surgeon annually remained constant over the time period 2004-2014. In 2004 overall TJA per surgeon was 47.2, with an insignificant change to 47.4 in 2014. Revision TJA per surgeon was 4.15 in 2004 and 4.22 in 2014. (Figure 5) Similar trends were seen regionally, with rises in TJA procedures, revision TJA

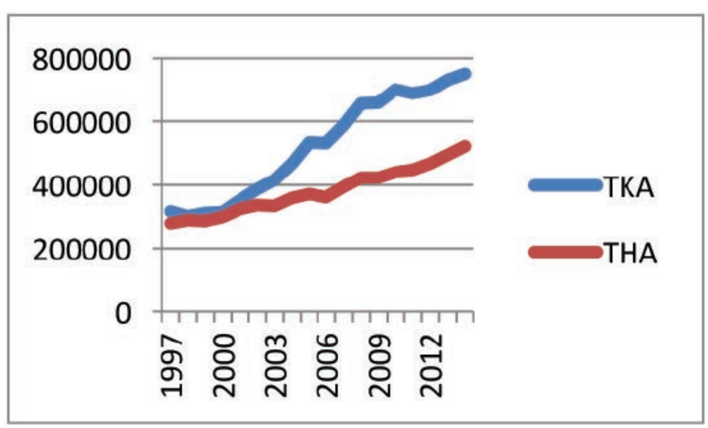

Figure 2. Total hip arthroplasty and total knee arthroplasty overall volume in the United States, 1997-2014.

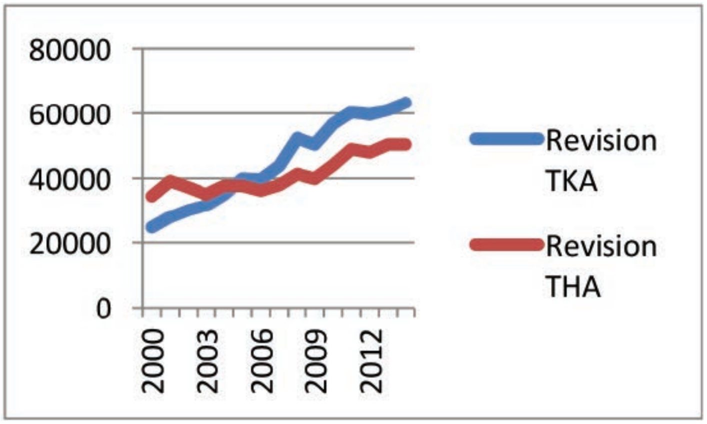

Figure 3. Revision total hip arthroplasty and revision total knee arthroplasty volume in the United States, 2000-2014.

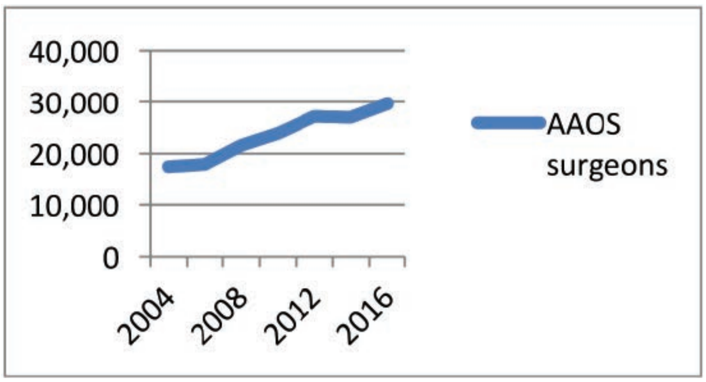

Figure 4. AAOS orthopedic surgeons in the United States according to AAOS Orthopaedic Surgeon Census, 2004-2016.

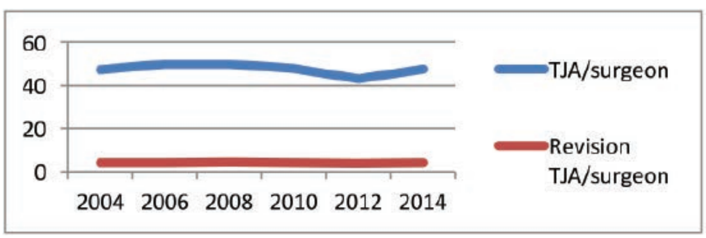

Figure 5. Overall total joint arthroplasty procedure and revision total joint arthroplasty procedure per orthopedic surgeon in the United States, 2004-2014.

procedures, and orthopedic surgeon volume in the Northeast, Midwest, South, and West. TJA and revision TJA procedures per surgeon, however, remained stable over the time period.

Table 1 demonstrates projected primary and revision TJA procedures for 2010 and 2020 according to the Kurtz paper. Comparison is made with the actual 2010 and 2014 data.

Discussion

The use of big clinical databases is allowing analysis of trends in orthopedics on a national level. In 2007, Kurtz and colleagues used this type of database to project growth of primary and revision total joint arthroplasty.[3] Using data from the HCUP NIS survey from 1990-2003, their team projected exponential growth of primary and revision TKA and continued linear growth of primary and revision THA. In the development of the Kurtz predictions, the model incorporated United States population estimates to develop estimates by age subgroup. The assumption inherent to this model was that projected increases in the aging population are driving increased demand for TJA procedures.

This is a logical assumption: older Americans are the primary recipients of TJA, thus an aging population drives demand for TJA procedures. However, the outcome modeled, that is total TJA performed in the United States is not a marker of demand, it is a demonstration of surgical performance, and thus, directly related to supply of surgeons available to perform these procedures. Our study incorporates the number of surgeries performed per orthopedic surgeon over the time period from 2004-2014. This analysis demonstrates a flat growth rate of surgeries performed per surgeon.

We assert that the dramatic rise that has been observed in TJA over the past two decades may be related to the rise in orthopedic surgeons in the United States. Furthermore, demand may be well above the saturation point at which surgeries can be performed. The ramifications of such unmet demand could be continued growth in TJA in parallel with increasing orthopedic surgeon density. It may be that the number of TJA procedures is far below patient demand, limited by the number of available surgeons. This would negate the Kurtz projection that primary and revision TKA will proceed to undergo exponential growth.

The year 2010 and 2020 Kurtz projections are demonstrated in Table 1. Their study uses a more narrow definition of TJA,

Table 1. Comparison of Predicted 2010 and 2020 primary and revision TJA procedures from Kurtz et. al with Actual primary and revision data from 2010 and 2014 using same coding scheme. * *Denotes significant difference from estimate*

Procedure	Predicted 2010	Actual 2010	Predicted 2020	Actual 2014
Primary THA	253,000 (232,000-276,000)	291,994*	384,000 (339,000-435,000)	371,605
Primary TKA	663,000 (618,000-711,000)	632,862	1,520,000 (1,362,000-1,700,000)	680,886*
Revision THA	47,800 (40,300-56,100)	44,032	67,600 (54,000-83,900)	50,425*
Revision TKA	55,300 (46,500-65,100)	56,586	55,300 (46,500-65,100)	63,205

incorporating only a single ICD-9 code for each of these procedures. Using the same coding scheme, actual data for primary and revision TJA procedures is provided for 2010 and 2014, the most recent available data year in the HCUP NIS database. It is clear that even at this earliest projected time point, 2010, the Kurtz projection has underestimated the number of primary THA procedures. For 2020, it appears the Kurtz projection has grossly underestimated the number of primary THA procedures that will be performed, as this estimate has already been met in 2014. The primary TKA projection, however, has not yet reached half of its 2020 estimate in 2014. Similarly for revision TJA, the 2020 estimate for revision TKA procedures has already been met and revision THA procedures have nearly been met by 2014. It may be the case that primary TKA procedures have not yet begun their exponential growth. However, it is evident that the Kurtz projections have been much too conservative for primary and revision THA, and revision TKA. We believe that the outlook is overly optimistic with regard to primary TKA procedures.

Following analysis of the Kurtz et al. methodology, we propose modeling future projections by taking into account expected surgeon growth as the strongest predictive variable. In addition, we plan to categorize the analysis into trends by demographic and comorbidity variables such as age, gender, and obesity to better understand what other factors are playing a significant role in driving the rise in TJA procedures.

References

1. American College of Surgeons. ACS NSQIP. "NSQIP in the Literature". 2017 March 10.

2. Murphy M, Alavi K, Maykel J. Working with existing databases. Clin Colon Rectal Surg 2013 March 01;26(1):5-11.

3. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am 2007 April 01;89(4):780-5.

4. Agency for Healthcare Research and Quality, Rockville, MD. HCUP Nationwide Inpatient Sample (NIS). Healthcare Cost and Utilization Project (HCUP). 2000-2011.

5. Agency for Healthcare Research and Quality, Rockville, MD. HCUP National Inpatient Sample (NIS). Healthcare Cost and Utilization Project (HCUP). 2012-2014.

6. American Association of Orthopaedic Surgeons. AAOS Orthopaedic Surgeon Census 2004-2016. 2016.