Big Data, Big Research

Implementing Population Health-Based Research Models and Integrating Care to Reduce Cost and Improve Outcomes

Afshin A. Anoushiravani, MD[a], Jason Patton, MD[a],
Zain Sayeed, MSc, MHA[a], Mouhanad M. El-Othmani, MD[a],
Khaled J. Saleh, MD, MSc, FRCS (C), MHCM, CPE[b],*

KEYWORDS

- Big data • NIS • NSQIP • Medicare data • Research advances

KEY POINTS

- Big data, in the health care setting, may be defined as a collection of information extracted from traditional and digital sources used to drive future discoveries and analyses.
- Awareness of advantages and limitations of large data registries, such as Medicare claims data, National Surgical Quality Improvement Program, National Inpatient Sample, Kid's Inpatient Database, and private alternatives, are necessary for orthopedic surgeons to conduct meaningful outcomes research.
- Use of the International Society of Arthroplasty Registries' recommendations on creation of registries will enable orthopedic surgeons to target quality improvement initiatives and better track patient outcomes.

INTRODUCTION

In the early twentieth century, Dr Codman became the first advocator for the collection and analysis of patient outcomes.[1] With the aid of medical and technological advancements throughout the century, the collection, analysis, and interpretation of collected data gave rise to an era of Big Data.[2] Simply put, big data is a collection of information extracted from traditional and digital sources used to drive medical advancements.[3] Over the past decade, several factors have converged allowing for the rise of big data and its implementation in clinical research.[4] Advanced electronic devices with data mining capabilities as well as the reduced cost of data storage and analysis has provided clinicians with the capabilities to answer questions beyond the scope of randomized controlled trials and meta-analyses.[4,5]

Big Data: Big Business

In the last 5 years, big data research applications are quickly turning into big business. International Business Machine, a prominent leader in information technology (IT), recently reported that 90% of the world's data was collected within the last 2 years, further demonstrating the rapid

---

rate at which information is being collected.[6] Within the next 5 years this IT sector is projected to grow by 400% to $50 billion annually.[7] As a result, many corporations and entrepreneurs are focusing on IT applications in health care. Early leaders in this emerging sector have been large hospital organizations who have used patient data to implement population-based health care initiatives. Through this approach, clinical and economic advantages in the form of improved quality of care, increased efficiency, and reduced resource consumption have been realized.[8–10] Although big data can be beneficial to patients, health care providers must use this information in an appropriate manner, as improper use may lead to harmful outcomes and clinical practices.

Large national databases have emerged as a viable means of reporting surgical outcomes and resource consumption patterns within orthopedic surgery. In the United States, several large publically governed patient registries exist. The strengths and limitations of each database depend on the purpose and design of the database. Before incorporating data from a specific database, an investigator must have a good understanding of the research question being asked as well as strengths and limitations associated with available databases.[11] The purpose of this article is to review various advantages and disadvantages of large data sets available for orthopedic use, examine their ideal use, and report how they are being implemented nationwide. Improvements that can be made to more efficiently collect relevant data and introducing model orthopedic practices that have embraced the big data big research model are also presented. Throughout this article, the reader should also be mindful that clinicians cannot ignore big data. Furthermore, clinicians must ensure that it is used in an ethical manner to improve patient outcomes.[12]

CURRENT BIG DATA REGISTRIES
International Databases
The first national orthopedic registry was created in Sweden in 1975 to collect information on total knee arthroplasty (TKA).[13] Since then, all Scandinavian and several English-speaking countries have developed independent total joint registries.[14] The Swedish joint registry assigns each patient a single national health identifier ensuring that a given primary prosthesis implanted at one institution can be connected to subsequent revisions at a different institution.[14] In 2007, the Nordic Arthroplasty Register Association (NARA) was created to enable collaboration among the TKA and total hip arthroplasty (THA) registries of Sweden, Denmark, and Norway.[15] Although the collaboration has allowed for a robust analysis of total joint arthroplasty (TJA) outcomes, NARA demonstrates how intraoperative techniques, such as cement fixation during THA, depend on regional norms rather than evidence-based practices.[14] Additionally, infrequently used implants have been removed from the Swedish market because of insufficient outcomes data.[14] Ultimately, the relatively small sample size and national tendencies in regard to surgical technique and prosthesis selection may limit the comparative capabilities of Scandinavian TJA registries.[14] Although not without limitation, these evidence- and population-based international registries have served as prototypes for several American data sets.

Medicare Claims Data
The United States has assigned specific governmental agencies with the task of collecting orthopedic care administrative data, whereas European registries were developed under the impetus of professional societies.[14] In 1965, the US Congress established Medicare as Title XVII of the Social Security Act; on July 1, 1966 the program was initiated.[10,16] The Centers for Medicare and Medicaid Services (CMS) is the national insurance program offering health care to 4 groups of US citizens: those 65 years or older, the disabled, those with end-stage renal disease, and those with amyotrophic lateral sclerosis.[16] The Medicare claims data set has recorded administrative claims data, reimbursement, and payment information on more than 45 million beneficiaries; therefore, it is the most robust nationwide database available to clinicians.[10,16] However, it is primarily limited to elderly and disabled patients and is not representative of the US population.

Additionally, the database has the ability to track beneficiaries through both ambulatory and inpatient settings. This feature enables researchers to assess short- and long-term outcomes and trends in resource utilization. Although the Medicare data set is considered the most comprehensive and robust database available in the United States, it is heavily regulated and is associated with significant up-front costs ranging from $3000 to $20,000 per data file.[10] Furthermore, the data set relies heavily on *International Classification of Disease* (ICD) codes and is ineffective at assessing nonbillable events. It should also be noted that the Medicare database does not include information on