Orthopaedic Healthcare Worldwide

The Big To Do About “Big Data”

Peter L. Schilling MD, MSc, Kevin J. Bozic MD, MBA

William Osler, a founding father of modern medicine, emphasized the importance of a good patient history by saying, “Listen to your patient, he’s telling you the diagnosis.” This adage is as true today as it was in Dr. Osler’s day, but with an interesting distinction: Our ability to listen to

patients has never been greater than it is today.

The data and information we are familiar with are undergoing a revolutionary transformation that will fundamentally change the way we interact with and understand our world. One phrase is increasingly used to describe this transformation: “Big data.” The exact definition of big data is still a moving target, but in general, it refers to the explosion of digitized data created by people, machines, sensors, tools, and other mechanisms. This data acts like an audit trail or data footprint, telling a story about the events and interactions of humans, machines, markets, natural systems, and other entities. The big to-do about big data is not so much its size, but that it contains hidden as signal within noise [9] some valuable predictive information.

The business world already has begun making large-scale investments to understand and utilize this wealth of information for competitive advantage. The motivation behind this is remarkably simple: They want to understand customer behavior, and extract value from it.

Therein lies the question for orthopaedics, and all of medicine: If big data can inform businesses’ understanding of their customers, can it also be used to better understand and treat our patients? The answer is yes, and it is currently happening. If big data still feels abstract, then look no further than the Electronic Medical Record (EMR) to get better acquainted with it. The medical record, that cozy, comfortable bastion of daily medical practice, has all the hallmarks of big data in its aggregated digital form. They generate and store massive quantities of complex clinical information. Like most big data, EMRs are text-heavy and unstructured. Unstructured data lacks the predefined organization needed to readily analyze or process the information held within it. Even in the Digital Age, unstructured data is challenging to deal with, but this is changing.

Commercial analytics software has emerged to meet the challenges of big,

A Note from the Editor-in-Chief: We are pleased to present to readers of Clinical Orthopaedics and Related Research® the latest Orthopaedic Healthcare Worldwide column. This section explores the political, social, and economic issues associated with delivering musculoskeletal care in the many environments in which our specialty is practiced, both in the United States and around the world. We welcome reader feedback on all of our columns and articles; please send your comments to eic@clinOrthop.org.

Each author certifies that they, or any members of their immediate families, have no funding or commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article. All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research* editors and board members are on file with the publication and can be viewed on request. The opinions expressed are those of the writers and do not reflect the opinion or policy of CORR® or the Association of Bone and Joint Surgeons®.

P. L. Schilling MD, MSc
Sports Orthopedics and Rehabilitation
Medicine Associates (SOAR), 500
Arguello Street, Suite 100, Redwood
City, CA, USA
e-mail: pschillin@gmail.com

K. J. Bozic MD, MBA (✉)
Department of Orthopaedic Surgery,
University of California San Francisco,
500 Parnassus Avenue, MU320W, San
Francisco, CA 94143-0728, USA
e-mail: bozick@orthosurg.ucsf.edu

Orthopaedic Healthcare Worldwide

unstructured data, and now, the widespread adoption of EMRs is paving the way for analytic tools specific to healthcare. Healthcare analytics is medicine's version of big-data analytics, and it has already been anointed by industry experts as the next big thing for reducing costs, improving outcomes, and coordinating care [6]. By integrating medicine, statistics, and computer science, healthcare analytics will one day enable us to unify sources of clinical data and information across formats and platforms, ultimately creating a longitudinal record for patients across inpatient, outpatient, specialty, and ambulatory settings. This seamless integration of data, combined with the analytics to see and communicate insightful patterns within it, will be an invaluable tool for improving quality, reducing cost, and advancing research. Industry experts estimate that analytics could save the US healthcare system as much as USD 300 billion per year [1]. Most exciting of all is what clinical data analytics is expected to do for the day-to-day care of patients in the hospital, clinic, and throughout the community.

Granted, clinical data analytics software can be expensive, and small hospitals on tight budgets will not be able to invest in the technology initially. We suspect that the adoption of clinical data analytics is likely to behave like other technological innovations in

healthcare that have large upfront costs (like MRI machines and EMR systems). We anticipate there will be early adopters, most likely large hospitals with deep pockets and an appreciation for how the technology can be immediately applied, and those that adopt the technology later, such as smaller hospitals on tight budgets who need to see the benefits before making a large investment.

Facile use of clinical data analytics will one day become as familiar as radiograph imaging and the stethoscope. Analytics applied to clinical information will give physicians greater insight into patients' health profiles by bringing to light patterns and signals within noise that can be used to assess, monitor, and predict treatment outcomes based on patient profile. And when this insight is coupled with our ever-increasing connectivity, we will have new opportunities for patient engagement and care coordination.

But the “Age of Big Data” has opportunities that go far beyond the confines of EMRs. We are only just starting to understand and apply a broad array of nontraditional data sources to healthcare. History puts these opportunities into context. Recall Dr. Osler's advice about listening to patients. Today, we have the unprecedented ability to listen to the collective voice of millions of people around the globe in almost real-time. Roughly

15 years ago, Google Inc. figured out that people speak through the information they seek online. Google listened to these voices on its way to becoming the second most valuable company in the world [10] by converting the information they gather about its global users into highly targeted online advertising.

Suspend judgment, and think about how this approach can be used for understanding trends in healthcare. A researcher named Gunther Eysenbach showed that these queries are an intrinsically valuable source of information about health trends, and he used it to lay the foundation for a web query-based approach to influenza surveillance that Google later carried forward with modest success [2]. The ongoing program called Google Flu Trends operates based on the clever observation that searches for flu-related information become more popular in areas where flu is prevalent [3, 4]. Such information may one day be used to compliment the Centers for Disease Control and Prevention's traditional influenza surveillance methods by acting as an early warning system.

Google Flu has been far from perfect; nonetheless, it represents a new way to think about population health management [7, 8]. This is big data in action, a new interpretation of Osler's adage, and it works by listening to the combined voice of millions of people

Orthopaedic Healthcare Worldwide

seeking health-related information on the Internet every day on topics as vast and varied as health and medicine itself. For example, web searches for the word “injury” increase in popularity in the United States every September at the start of the football season [5]. This pattern represents an opportunity for professional medical societies. Our patients are asking these injury-related questions every September, why not answer them through targeted, online public health announcements?

It is good to be skeptical. Big data-related efforts have had as many failures as they have successes. Google Flu has had well-publicized problems with its algorithm’s ability to make accurate estimates of influenza activity [7, 8]. It is important to remember that a data-rich environment does not obviate the need for established rules of probability and statistics. It is irresponsible to completely forgo hypothesis testing in favor of mining large data sets for patterns and using these patterns to assign causality and predict the future. Moreover, bias must always be considered, because regardless of the size of the data, it will never go away completely. Finally, we must remember to protect patients’ privacy as well as their individuality. Physicians treat patients, not

their profiles, and it is impermissible to allow the Age of Big Data to compromise patients’ health information no matter what the perceived gain.

Pitfalls aside, we now have an exciting array of new data sources and complimentary analysis tools that are likely to revolutionize every facet of medicine, from the ways we interact with and treat patients to the ways we conduct research. This article is not only a call to action, it is a call to dream. Big data is not a threat, but a tremendous opportunity, and this is only the beginning. Providers should seek out opportunities to learn more about the power and potential of big data, and how it can be used for the benefit of our patients. Keep an open mind, think creatively, but proceed with cautious optimism: Predicting the future has always been a dicey business.

References

1. Eysenbach G. Infodemiology: The epidemiology of (mis)information. *Am J Med.* 2002;113:763–765.
2. Eysenbach G. Infodemiology: Tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc.* 2006:244–248.
3. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature.* 2009;457:1012–1014.
4. Google.org. Flu Trends: How does this work? Available at: <http://www.google.org/flutrends/about/how.html>. Accessed June 30, 2014.
5. Google Trends. June 2014 charts. Available at: <http://www.google.com/trends>. Accessed on June 30, 2014.
6. Kayyali B, Knott D, Van Kuiken S. The big-data revolution in US health care: Accelerating value and innovation. Available at: http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care. Accessed July 17, 2014.
7. Lohr S. Google flu trends: The limits of big data. Available at: <http://bits.blogs.nytimes.com/2014/03/28/google-flu-trends-the-limits-of-big-data/>. Accessed June 15, 2014.
8. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: Traps in big data analysis. *Science.* 2014;343:1203–1205.
9. Silver N. *The Signal and The Noise: Why So Many Predictions fail — But Some Don't.* New York, NY: Penguin Press. 2012.
10. Solomon J. Google worth more than Exxon. Apple next? Available at: <http://money.cnn.com/2014/02/07/investing/google-exxon-market-value/>. Accessed August 7, 2014.