Big Data: The Paradigm Shift Needed to Revolutionize Musculoskeletal Clinical Research

David L. Helfet, MD, MBChB, Beate P. Hanson, MD, MPH, and Diarmuid De Faoite, MBS, BBS

One year ago, we wrote an editorial in *The American Journal of Orthopedics* on missing data.[1] This year, data is once again the focus of our editorial but from a different perspective. Rather than focus on the problems of incomplete data, we want to talk about the possibilities of collecting all data through advanced technology, a phenomenon better known as "Big Data."

New Technology

The factors driving Big Data's ascendency are the digitalization of useful data, increased means to gather digitalized data, and cheaper analytic power.[2] Computer behemoth IBM claims that 90% of the data in the world today has been created in the last 2 years alone.[3] Big Data is not just an industry buzzword; it is already an industry in itself. Revenue from Big Data reached $18 billion in 2013 and is predicted to rise to $50 billion in the next 5 years.[4] While it is easy to see how Internet companies like Amazon can both collect and use all of the data they receive from customers (to suggest their next purchase, for example), it might be less easy to see how Big Data concepts can be applied to clinical research.

Health Care

Electronic data records are propelling the development of pools of information in health care. Almost half of all hospitals in the United States are participating in health information exchanges (HIEs).[5] When these sources of data pools are integrated, the information collected can be used in a powerful way. For example, the health maintenance organization Kaiser Permanente uses a new computer system that drives data exchange between medical facilities. Patient benefits include improved outcomes of cardiovascular disease, and an estimated $1 billion has been saved due to reduced office visits and laboratory tests.[5]

Contemporary Studies

Let's quickly consider how we currently conduct clinical studies. Because we do not usually collect data from the entire population, contemporary clinical studies offer only a snapshot of a subsection of patients. The results from this sample are then usually extrapolated to the general population. This was fine when there were insurmountable technological and logistical issues. So instead of trying to collect data from everyone in the population of interest, we select a sample of patients and expend our energy on controlling for the suboptimal methods we currently employ, techniques which are the best ones available to us.

> "Since we cannot ignore Big Data, we must do what we can to ensure that its potential is harnessed to reduce costs and improve patient outcomes."

What are the consequences of all this? Those of us in clinical research are usually very concerned about dealing with confounding factors: selection bias, adjusting for missing data, controlling for errors, and so on. We can also see how imprecise our current methods are by how often a scientific manuscript ends with a call for larger-scale research. Indeed, a scientific research paper that does not list the study's limitations is often regarded with suspicion, a telling indictment of the problems we expect to encounter in clinical research.

So what has historically been the best current solution to overcome these challenges? A meta-analysis of random-

Dr. Helfet is Associate Editor of Trauma of this journal; Professor, Department of Orthopedic Surgery, Cornell University Medical College; and Director of the Orthopaedic Trauma Service, at the Hospital for Special Surgery and New York–Presbyterian Hospital, New York, New York. Dr. Hanson is Director and Mr. De Faoite is Education Manager, AO (Arbeitsgemeinschaft für Osteosynthesefragen/Association for the Study of Internal Fixation) Clinical Investigation and Documentation (AOCID), Davos, Switzerland.

Authors' Disclosure Statement: The authors report no actual or potential conflict of interest in relation to this article.

Am J Orthop. 2014;43(9):399-400.

ized controlled trials sits atop the evidence pyramid as being the best level of evidence. However, even the use of meta-analyses can be problematic. One group of researchers found that in 2005 and 2008, respectively, 18% and 30% of orthopedic meta-analyses had major to extensive flaws in their methodology.[6] Indeed, implicit in the use of a meta-analysis is a criticism that our current studies with their limited sample sizes do not tell the whole story.

Paradigm Shift

We are in the middle of a paradigm shift in the way we can collect and analyze data. Our focus until now has been on identifying a causal relationship in our studies. New technology which allows for large-scale data collection and analysis means that we can now collect **ALL** patient data, in other words **N = all**. When you can collect all data, the *why* (causality) something is happening becomes less important than the *what* (correlation) is happening.[7] Studies will therefore begin to focus on effectiveness in the real world as opposed to measurements taken under the ideal (or nearly ideal) conditions of efficacy.

All of this is going to have implications, the greatest of which is the change in mindset that we are going to have to go through. How we conduct our studies and what their focus is will both change and expand. For example, the Mini-Sentinel project uses preexisting electronic health care data from multiple sources to monitor the safety of medical products that are regulated by the US Food and Drug Administration (FDA). This FDA-sponsored initiative, which only began in 2008, had already collected data on 178 million individuals by July 2014.[8]

Since we cannot ignore Big Data, we must do what we can to ensure that its potential is harnessed to reduce costs and improve patient outcomes. Given the potential of using electronic clinical data, it is also necessary to strike a note of caution. We have to keep uppermost in mind that new technologies like Big Data can unsettle a lot of people. A central tenet of clinical research is that patient data belong to the patient. Robust and transparent processes need to be developed to ensure that patients do not feel compromised in any way by their data being used in such new and widespread methods. The need to rethink and implement safeguards is already being addressed. For example, the university-associated Regenstrief Institute does not pass along even deidentified data to their Big Data industry partner.[9]

However, we need to also be cognizant of the fact that society is changing in the way people use and regard their own information. Patient-reported data is already being shared among patients online, for both common and rare diseases. The data are also richer and can go beyond the usual outcomes that are recorded to give a bigger picture, eg, why patients are not adhering to treatment regimens.[10]

In summary, it is our earnest belief that if the health care industry can embrace the concept of Big Data and utilize it properly, our patients and medical practices will be all the better for it.

References

1. Helfet DL, Hanson BP, De Faoite D. Publish or perish; but what, when, and how? *Am J Orthop*. 2013;42(9):399-400.
2. Nash DB. Harnessing the power of big data in healthcare. *Am Health Drug Benefits*. 2014;7(2):69-70.
3. What is big data? IBM website. http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html. Accessed July 22, 2014.
4. Upbin B. Visualizing the big data industrial complex. *Forbes* website. http://www.forbes.com/sites/bruceupbin/2013/08/30/visualizing-the-big-data-industrial-complex-infographic/. Published August 30, 2013. Accessed July 22, 2014.
5. Kayyali B, Knott D, Van Kuiken S. The big-data revolution in US health care: accelerating value and innovation. McKinsey & Company website. http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care. Published January 2013. Accessed July 22, 2014.
6. Dijkman BG, Abouali JA, Kooistra BW, et al. Twenty years of meta-analyses in orthopaedic surgery: has quality kept up with quantity? *J Bone Joint Surg Am*. 2010;92(1):48-57.
7. Cukier K, Mayer-Schonberger V. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, NY: Eamon Dolan/Houghton Mifflin Harcourt; 2013.
8. Mini-Sentinel distributed data "at a glance." Mini-Sentinel website. http://www.mini-sentinel.org/about_us/MSDD_At-a-Glance.aspx. Accessed July 22, 2014.
9. Jain SH, Rosenblatt M, Duke J. Is big data the new frontier for academic-industry collaboration? *JAMA*. 2014;311(21):2171-2172.
10. Okun S, McGraw D, Stang P, et al. Making the case for continuous learning from routinely collected data. Institute of Medicine website. http://www.iom.edu/Global/Perspectives/2013/~/media/Files/Perspectives-Files/2013/Discussion-Papers/VSRT-MakingtheCase.pdf. Published April 15, 2013. Accessed July 22, 2014.