

Not the Last Word: Big Data Will Make You Confront Big Ethical Questions—Here's Why

Joseph Bernstein MD

This column has its regular readers, and I can count on both of them to examine everything I write. Some of you, though, are new visitors and I'll venture a guess that it was the title of the column that drew you in. I base my surmise on a study of 100 million Facebook and Twitter posts, which found that placing the words "will make you" in the title stimulated

users to open the link more than any phrase [9]. The study also found that teasers like "here's why" were also powerful lures.

Discoveries like that show some of the power of big data [13]. Big data can drive so-called machine learning: Namely, using "modern computer and mathematic algorithms to recognize complex combinations of predictors with the capacity of handling huge amounts of data" [14]. Accordingly, machine learning can uncover prediction rules with seeming serendipity and uncanny accuracy.

In orthopaedics, we are accustomed to learning prediction rules from datasets. Consider the development and validation of the Mangled Extremity Severity Score (MESS) established by Johansen and colleagues [6]. The MESS is a rating system used to help surgeons identify "trauma victims whose irretrievably injured lower extremities warrant primary amputation" (as opposed to an attempt at repair). The authors developed the MESS by first retrospectively evaluating charts of patients presenting with lower extremity long bone fractures. They then identified the variables that lead to amputation: The extent of limb ischemia; age, the degree of shock; and the energy imparted during injury.

Last, they proposed and prospectively validated a weighting system that awarded points in these categories to discriminate between patients who had salvaged limbs and those who had amputated limbs.

The development of the MESS certainly was data-driven, but its creation is more-accurately described as a "small-data" approach. For one thing, the sample size was modest ($n = 51$). But more to the point, the researchers based their investigation on a priori guesses of which variables were likely to matter. Big data methods, by contrast, not only utilize a vastly larger population, they do not begin with assumptions about which parameters are important (and in turn consider only those); rather, they are empowered to consider every known variable, including those that might seem irrelevant.

The big-data approach—learning prediction rules from large datasets, without assumptions in advance about which factors matter—is coming to orthopaedics, especially now that electronic medical record (EMR) systems are ubiquitous. Methods driven by big data might be used, for example, to refine the indications of total knee replacement. Generally, we know that total knee replacement is appropriate for patients with end-stage arthritis, especially if other, less-drastic measures did not result in adequate pain relief. Yet we don't really know which patients are most appropriate for surgery—at least one in five patients do not achieve a

A note from the Editor-in-Chief: We are pleased to present to readers of Clinical Orthopaedics and Related Research® the next Not the Last Word. The goal of this section is to explore timely and controversial issues that affect how orthopaedic surgery is taught, learned, and practiced. We welcome reader feedback on all of our columns and articles; please send your comments to eic@clinorthop.org.

The author certifies that neither he, nor any members of his immediate family, have any commercial associations (such as consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article.

The opinions expressed are those of the writer, and do not reflect the opinion or policy of Clinical Orthopaedics and Related Research® or The Association of Bone and Joint Surgeons®.

Joseph Bernstein MD (✉), University of Pennsylvania, 424 Stemmler Hall, Philadelphia, PA 19104, USA, Email: orthodoc@uphs.upenn.edu

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.

J. Bernstein, Department of Orthopaedic Surgery, University of Pennsylvania, Philadelphia, PA, USA

Not the Last Word

clinically important improvement after knee replacement [1], and even experienced surgeons are terrible at identifying who those patients might be [3].

Perhaps a machine-learning program, let loose on a trove of various sources of medical big data (the administrative claim record, billing information, physician notes, and radiology reports, among others), can deduce a rule that anticipates preoperatively those who will be dissatisfied [14]. Big data may help us discover, hypothetically, that patients older than 70 years of age with slightly high blood pressure and slightly low hemoglobin are 40% less likely to be happy with their result. If such patients choose surgery less frequently (as they should), overall satisfaction rates should rise.

Of course, not all of the deduced rules will be as clean as the one just suggested, and that's where the ethical issues arise. Let's say our machine-learning program examines patient satisfaction in context of the full record and discovers 100% dissatisfaction among, say, HIV-positive men who have been late to at least two medical appointments in the past year. (Recall that the program examines everything in the record, not only those attributes we suggest.) What do we do with that information? Indeed, you may deem this finding offensive and oppressive. But your feelings do not make it any less true.

It is not controversial to assert that if a well-powered study discovered that slightly high blood pressure coupled with slightly low hemoglobin lead to a 100% failure rate, it's reasonable to not offer surgery to such a patient. (What better defines the term "contraindicated" than those attributes that point to universal failure?) Yet on the other hand, it seems wrong to deny treatment to a patient for seemingly irrelevant considerations, especially when such action is redolent of discrimination. And I don't

think claiming "just tell the patients about the risks and let them decide" will do. Ultimately, the surgeon's presentation dictates the patient's response [2]. Hence the ethical conundrum.

I am tempted to say, to paraphrase Fermat [12], that I have discovered a truly remarkable resolution to this problem, but this this essay is just too small to contain it. But if I did, I (like Fermat, most likely) would be bluffing. I'm not sure what the right answer is.

I'd guess that the ethical resolution begins with a reminder that big data can only reveal correlations, and that correlations can be confused by confounding. Maybe all the HIV-positive men are taking a medication that causes both bone loss and forgetfulness—explaining both the disappointments and missed appointments. Such patients might be ideal surgery candidates if they were given a different medication. Also, there are biases in our databases, in that some information is collected, and other information is not—and without these missing data points, the confounding features may remain undiscovered.

For now, we are left with a dilemma. We know that big data can teach us valuable things; we know that big data might mislead; but we lack the ability to discern always which rule applies.

Drake G. LeBrun MD, MPH Orthopaedic Residency Program Hospital for Special Surgery

It is hard to deny that big data holds promise in orthopaedics. The capacity to handle vast data sets with numerous outcomes and exposures can (ideally) lead to predictive models that are precise, accurate, and useful.

But big data, despite its promise, is not without its risks. In particular, big data poses major problems related to poor data analysis and erroneous interpretation.

The advent of big data brings to light an important issue in orthopaedic

clinical research regardless of the size of the dataset—the need for rigorous methodology and appropriate data analysis. In fact, it is likely that in many cases, traditional statistical methods using a conventional "small data" approach can actually be more accurate and more appropriate than big data techniques [5].

Big data analysis, and its controlling for confounding factors, missing data, nonindependence, simultaneous statistical tests, misclassification bias, and lumping, among other statistical complexities, can be challenging and more complex than small data analysis, particularly for those without a strong background in biostatistics [8]. As the complexity of data analysis grows, so too does the statistical foundation required to understand big data. Indeed, it is likely that the availability of large datasets and the capacity of our statistical software to analyze them are growing at a faster rate than the biostatistical and epidemiological expertise of orthopaedic surgeons. Thus, there is a growing disparity between the power of our data and our ability to wield that power effectively. Such a divide leads to a harrowing possibility: An increased risk of drawing spurious conclusions from incorrectly analyzed data and using those conclusions to affect clinical care.

How do we bridge that divide? The answer is deceptively simple: Train orthopaedic surgeons in clinical epidemiology. Understanding the basic tenets of epidemiology can help orthopaedic surgeons conduct studies more rigorously and critique studies more effectively. Indeed, the concepts underlying clinical epidemiology are not going away with the advent of big data. Rather, they are even more important now than ever.

What does all this have to do with ethics? When we conduct clinical research, we generate or test hypotheses

Not the Last Word

and make claims based on data. Those claims, in turn, can and often do influence clinical decision-making. If those claims are based on imprecise or inaccurate data, then spurious conclusions may ultimately impact clinical care. It is therefore incumbent on the researcher, the reviewer, and the reader to be aware of problems arising from inappropriate data analysis [4]. Furthermore, it is our prerogative to critically evaluate studies involving big data to ensure that the author's interpretation, and our own, are epidemiologically and statistically sound. In doing so, we avoid any risk of applying spurious results to our patients.

Charles Carroll IV MD
Orthopaedic Surgeon, NorthShore
Orthopedic Institute
Feinberg School of Medicine,
Northwestern University

Dr. Bernstein's column defines a conundrum of big data. It may not be the only one, but one that warrants attention and thought. As we spend countless hours on EMRs, we all should wonder, "What will happen with all of these data?" At least I wonder about it as I click through the EMR. We will face clinical decisions influenced or perhaps made by artificial intelligence or derived from large data sets. In some cases, the data and the decisions will be concordant and related to our clinical training and professional expertise. That is the easy part.

When the data present a perceived contraindication to accepted treatment we may be in a difficult spot. Where do we turn? How does it affect the doctor-patient interaction? Do we use the data to deny or alter accepted treatment?

I would offer that we turn to a patient-centered approach that includes consideration of the four pillars of ethics: Autonomy, beneficence, non-maleficence, and social justice. We cannot send our patients home with a

set of facts and related data and ask them to think about it and get back to us. We should take the data and interpret it taking into consideration the four pillars. We can then offer a patient-centered course of action, as well as discuss a reasonable course of treatment.

A surgeon's presentation does matter. If one interprets and considers conflicts, a process that promotes shared decision-making and autonomy can follow. A reasonable decision can be then made for the care of the patient.

Seth Haberman BA
Chief Executive Officer
Sense Education (<https://www.sense.education/>)

In the past, medical discoveries arrived by observation, association, hypothesis, and then when possible and ethical, experimentation. While this process is tried and true, the scientific method fails with extensive datasets, as our ability to make observations and associations breaks down. This trend is not confined to medicine. The particle physics field, which once relied on humans interpreting pictures of particles colliding, now uses large-scale computing power to analyze the interaction of hundreds of thousands of particles to reveal complex interactions, which are ultimately tied to some fundamental theory—and that's the key. In physics, big data is always a means not an end.

Perhaps it's different in Medicine. At a recent Google AI meeting Jeff Dean PhD, the head of Artificial Intelligence at Google Inc, discussed a deep-learning model trained on 248,000 patients capable of discerning in retinal scans cardiovascular risk factors, gender, smoking, and systolic blood pressure to within 11.23 mmHg [7]. Regarding gender detection from retinal scans, the system is highly capable, whereas the best humans are no better than chance at this endeavor [10]. It also

has practical applications—from making it more difficult to create a false identity, to improving the speed of searches in large iris databases, to potentially counting the number of people of different ethnicities who enter a country without recording his or her identity [10]. And it all works despite us not even knowing the relevant features that drive the model.

Dr. Bernstein is correct to point out that deep-learning, the particular subset of machine-learning that imitates the workings of the human brain, is mostly correlations with regressions and data reduction. What's critical is that deep learning does not uncover causality, and in some realms, that would indicate we should wait to understand the underlying science.

But if machine-learning systems can better predict all sorts of maladies than our current methods, should we be asking new ethical questions? If we were to impose the full scientific method on today's treatments, how many patients would be denied care?

Ethics deal with the moral decisions that govern our lives or professions. In medicine, those are already well understood. In that regard, the Talmud notes that a correct action, done for the wrong reason, eventually becomes a corrective action done for the right reason [11]. Perhaps that applies in medicine as well, but it's hardly assured.

References

- Berliner JL, Brodke DJ, Chan V, SooHoo NF, Bozic KJ. Can preoperative patient-reported outcome measures be used to predict meaningful improvement in function after TKA? *Clin Orthop Relat Res.* 2017;475:149-157.
- Bernstein J, Kupperman E, Kandel LA, Ahn J. Shared decision making, fast and slow: Implications for informed consent, resource utilization, and patient satisfaction in orthopaedic surgery. *J Am Acad Orthop Surg.* 2016;24:495-502.

Not the Last Word

3. Ghomrawi HMK, Mancuso CA, Dunning A, Gonzalez Della Valle A, Alexiades M, Cornell C, Sculco T, Bostrom M, Mayman D, Marx RG, Westrich G, O'Dell M, Mushlin AI. Do surgeon expectations predict clinically important improvements in WOMAC scores after THA and TKA? *Clin Orthop Relat Res.* 2017;475:2150-2158.
4. Hailer NP. Orthopedic registry research - limitations and future perspectives. *Acta Orthop.* 2015;86:1-2.
5. Househ MS, Aldosari B, Alanazi A, Kushniruk AW, Borycki EM. Big data, big problems: A healthcare perspective. *Stud Health Technol Inform.* 2017;238:36-39.
6. Johansen K, Daines M, Howey T, Helfet D, Hansen ST Jr. Objective criteria accurately predict amputation following lower extremity trauma. *J Trauma.* 1990;30:568-572.
7. Peng L. Assessing cardiovascular risk factors with computer vision. Available at: <https://ai.googleblog.com/2018/02/assessing-cardiovascular-risk-factors.html>. Accessed December 5, 2018.
8. Perry DC, Parsons N, Costa ML. 'Big data' reporting guidelines: How to answer big questions, yet avoid big problems. *Bone Joint J.* 2014;96-B:1575-1577.
9. Rayson S. We analyzed 100 million headlines. Here's what we learned (new research). Available at: <http://buzzsumo.com/blog/most-shared-headlines-study/>. Accessed October 25, 2018.
10. Sandhana L. Iris scanner could tell your race and gender. Available at: <https://www.newscientist.com/article/mg21228346-000-iris-scanner-could-tell-your-race-and-gender/>. Accessed December 6, 2018.
11. Sefaria. The William Davidson Talmud. Available at: <https://www.sefaria.org/Sanhedrin.105b>. Accessed December 6, 2018.
12. Singh S. *Fermat's Last Theorem*. New York, NY: Anchor Books; 1997.
13. Stephens-Davidowitz S. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*. New York, NY: Dey Street Books; 2018.
14. Thio QCBS, Karhade AV, Ogink PT, Raskin KA, De Amorim Bernstein K, Lozano Calderon SA, Schwab JH. Can machine-learning techniques be used for 5-year survival prediction of patients with chondrosarcoma? *Clin Orthop Relat Res.* 2018;476:2040-2048.