Surgical Research Review

Integrating "big data" into surgical practice

Brittany Mathias, MD,[a] **Gigi Lipori, MBA,**[b] **Lyle L. Moldawer, PhD,**[a] *and* **Philip A. Efron, MD,**[a]
Gainesville, FL

'Big data' is the next frontier of medicine. We now have the ability to generate and analyze large quantities of healthcare data. Although interpreting and integrating this information into clinical practice poses many challenges, the potential benefits of personalized medicine are seemingly without limit. (Surgery 2016;159:371-4.)

From the Department of Surgery,[a] *University of Florida College of Medicine; and the University of Florida Health and Shands Hospital,*[b] *Gainesville, FL*

THE RAPID PACE of medical discovery is evident in the daily news, but few advances have changed fundamentally the practice of medicine in the way that radiographic imaging or antibiotic pharmacotherapy revolutionized medical diagnosis and management. Today, we are on the cusp of the next revolution in medicine. In 2001, the human genome was sequenced, birthing the age of "personalized medicine."

This work was also supported by grants R01 GM-040586-24 and R01 GM-081923-06 awarded by the NIGMS. In addition, PAE was also supported by P30 AG-028740 from the National Institute on Aging and R01 GM-113945-01 (NIGMS). Finally, PAE and LLM were supported by P50 GM-111152-01 (NIGMS). The Glue Grant database was supported by the Inflammation and the Host Response to Injury Large Scale Collaborative Research Program (Glue Grant U54 GM-062119), awarded to Dr Ronald G. Tompkins, Massachusetts General Hospital, by the National Institute of General Medical Sciences (NIGMS). The work represents a secondary use of this public database, and the conclusions and discussion are the authors, and do not necessarily represent the views of either Massachusetts General Hospital, or the NIGMS. BM was supported by a training grant in burn and trauma research from the National Institute of General Medical Sciences (NIGMS) (T32 GM-008721).

Conflict of Interest and Financial Disclosure Statement: No conflict of or competing interests have been declared.

Accepted for publication August 12, 2015.

Reprint requests: Philip A. Efron, MD, FACS, FCCM, Department of Surgery, University of Florida College of Medicine, Room 6116, Shands Hospital, 1600 SW Archer Road, Gainesville, FL 32610-0286. E-mail: Philip.Efron@surgery.ufl.edu.

0039-6060/$ - see front matter

http://dx.doi.org/10.1016/j.surg.2015.08.043

Large-scale, genome-wide sequencing, transcriptomics, proteomics, microbiomics, and metabolomics coupled with wide-spread implementation of the electronic health records (EHR) has resulted in immense quantities of individual health data. Implementing the parameters of quality control and merging and extracting meaningful information from an exponentially growing dataset poses an immense challenge. The implementation of "big data" is the next revolution in the field of health care requiring a reevaluation of the entire system of cost-effective best practices. Education of physicians will need to adapt, starting in medical school and extending to the effective use of EHRs in clinical practice. The design of clinical diagnostic laboratories requires modification. This process will even include how patients are educated on their personalized medicine.

The term *big data* refers to information collected or collated on a large scale and evaluated to obtain new insights or forms of usefulness.[1] The creation of large databases and the capability to mine them for data is not innovative. Large clinical databases such as The Human Genome Project (available from: www.genome.gov), Encyclopedia Of DNA Elements (ENCODE, a project to identify all functional elements in the human genome sequence; available from: www.encodeproject.org), 1000Genomes (available from: www.1000genomes.org), the International HapMap Project (available from: http://hapmap.ncbi.nlm.nih.gov/), the National Surgical Quality Improvement Program (available from: http://site.acsnsqip.org), as well as smaller projects such as

the "Inflammation and Host Response to Injury" Glue Grant (available from: www.gluegrant.org), have been completed or are ongoing. What is novel is the imminent combination of an individual's EHR with that same individual's gene sequence, polymorphisms, and expression patterns from specific tissues.

The immediate frontier is the technologic capacity to integrate big data into health care practice in a cost-efficient manner. The National Institutes of Health (NIH) considers big data a research priority and created funding opportunities via the "NIH Big Data to Knowledge (BD2K)" (available from: http://bd2k.nih.gov/#sthash.hxjXbQ3y.dpbs). The goal of this project is to overcome the major challenges in using big data, which are quoted as (1) locating data and software tools, (2) getting access to the data and software tools, (3) standardizing data and metadata, (4) extending policies and practices for data and software sharing, (5) organizing, managing, and processing biomedical big data, (6) developing new methods for analyzing and integrating biomedical data, and (7) training researchers who can use biomedical big data effectively. Achieving these goals may lead ultimately to the routine use of big data in clinical medicine and across other scientific fields.

In addition to the NIH BD2K efforts, several other large-scale projects are working to integrate big data into the clinical arena. The Electronic Medical Records and Genomics (eMERGE) Network was initiated in September 2007 with the primary goal of combining DNA biorepositories with EHR systems for large-scale, high-throughput genetic research. A key goal of eMERGE Phase II is to explore the best avenues to incorporate genetic variants into the EHR for use in clinical care, such as for improving genetic risk assessment. In addition, it aims to improve the prevention, diagnosis, treatment, and/or accessibility of genomic medicine (available from: https://emerge.mc.vanderbilt.edu/). Phase III of eMERGE began in August of 2015 and is intended to broaden participation to an increasing number of academic and nonacademic health care centers. Furthermore, the Clinical Sequencing Exploratory Research consortium is a program sponsored by the National Human Genome Research Institute, which explores applications of whole-genome and whole-exome sequencing.[2] Finally, President Obama has announced a new initiative to accelerate our capacity to use precision medicine and has earmarked NIH and funding via the Centers for Medicare and Medicaid Services for this purpose.[3]

Big data will be coming to your local hospital! The only question now is, "How long until it can be used easily and cost effectively as a tool from the armamentarium of the clinician?" Big data is not to be feared but rather embraced, although the challenges are immense. Big data provides certain opportunities in clinical medicine, including (1) the generation of and dissemination of new knowledge, (2) the translation of personalized medicine into health care, and (3) increased patient involvement by creating accessible and understandable data.[2] One field that is primed to incorporate big data into standard health practice is personalized medicine through medical genetics. Genomic medicine, which incorporates genomics-based diagnostics into practice, is already becoming the standard of care.[4] Pharmacogenetics uses genetics to individualize drug therapy.[5] The entire human genomic and exomic sequence, or whole-genome sequencing, is now available and can be used for rapid and relatively inexpensive individual analysis.[1,6] The cost of exon sequencing is predicted to become <$500 in the next decade; indeed every patient may likely have their genome sequenced.

Genomic medicine will enable clinicians to use "tailor-made" medicine and minimize disparate outcomes that result from "off-the-rack" therapy.[7] Risk:benefit analyses from drug administration to interventional therapies will become routine as more information becomes available. The surgeon's practice will certainly not remain insulated from this revolution. The dosing of drugs, such as warfarin, norepinephrine, and vasopressin, are already known to be affected by specific genetic variants.[4,6,8] Recently, our comprehension of pharmacogenomics for appropriate patient treatment has expanded exponentially and now includes, but is not limited to, antiseizure, antimicrobial, antiviral, antifungal, and antiplatelet medications.[4]

To optimally use big data in a clinical setting, refined EHR systems will be necessary.[7] Future EHR systems will need to include "genomic medicine clinical decision support" to enable point-of-care personalized medicine.[7] The sheer magnitude of information available to the clinician will require clinical decision support to aid in the interpretation of genetic variations and their application to a particular disease or medicine for each individual patient.[6] Clinical decision support will enable point-of-care analysis of a patient's genetic profile, such as whole-genome sequencing, that can be incorporated readily and cost effectively into daily clinical practice. This is a feasible alternative to a genetic professional who spends

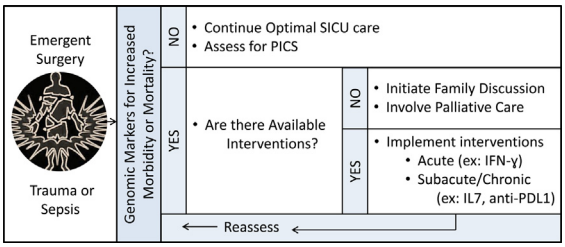

Fig 1. Algorithm for utilization of 'big data' in the clinical setting.

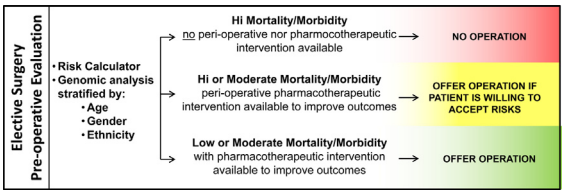

Fig 2. Utilization of risk calculators stratify patients allowing for a more informed pre-operative evaluation for elective surgery.

typically at least 7 hours in preparation for admitting a new patient.[6] A basic comprehension of this will be so important to the surgeon that we predict it will become a routine part of a surgery resident's training and questions regarding this will be added to future certification examinations.

Similarly, the transcriptome of tumors is already being used to predict pharmacologic and therapeutic responsiveness for some leukemias and solid tumors. Expression patterns from individual tissues and cell populations are being used for diagnostic and prognostic purposes.[9] As translational research progresses, not only will genome-wide sequencing be commonplace, but so will expression analysis, proteomics, and metabolomics.

It is clear that surgical oncology is already benefiting from the use of big data toward enhancing precision medicine, but personalized medicine is already expanding beyond cancer.[3] Using big data will be the norm for various other diseases, ranging from determining disease risk to understanding the specific mechanisms of a disease in a patient to optimize their individual therapy.[3] It is has been known that premature death from noncancerous illnesses has been related to genetics for almost 30 years—the relative risk of mortality from infection of a biologic offspring is 5.31 as compared with 3.02 for an adoptive child.[10] Thus, investigators have used big data analysis to determine genomic markers of increased mortality in sepsis.[11] For instance, our laboratory has been able to determine within 48 hours of traumatic injury which patients will have a good versus poor outcome based on their leukocyte transcriptome.[12] Other laboratories[13] have used the blood transcriptome to distinguish patients with sepsis requiring antimicrobial support from those with noninfectious systemic inflammatory response syndrome.[13] It is possible to use these kinds of data to alter clinical practice (Fig 1). We can concentrate our efforts on those patients who are likely to have complicated outcomes and to minimize futile interventions on patients who will not benefit from certain treatments such as immunomodulators. This form of personalized medicine will be more cost effective than the currently used, generalized treatment of all critically ill patients. The potential for such predictive tests is not limited to those who arrive in emergency distress; our practice can be tailored before making an incision. Efforts to improve patient outcomes and cost effectiveness in surgery have led to the American College of Surgeons' Clinical Risk calculator (available from: http://riskcalculator.facs.org/), yet another example of a large database, like the National Surgical Quality Improvement Program, being used to give real time guidance to health care. By combining these types of analyses with a patient's preoperative genetic profile, the future surgeon's practice of personalized medicine will be augmented (Fig 2).

As with all advances, there may well be some detrimental aspects to the incorporation of big data in surgery. Social inequality may become amplified in medicine, because the lack of Internet access may limit the ability of poor or rural individuals' to take advantage of big data's use in medicine.[7] Currently, advances in the technology are limited to major academic institutions, although community-oriented health care providers like Geisinger, InterMountain West, and Kaiser-Permanente are leaders in integrating EHRs into their practices. In addition, new issues may arise with confidentiality, privacy, and safety of genomic information in EHRs.[7] Records of extensive biological specimens will be in the EHR, possibly including an individual's cell populations, proteins, metabolites, and the patterns of RNA and DNA[3]; thus, patient privacy and protection will become a mounting priority and technologic challenge.

We must remember that more data is not always better, especially if the data are compromised, inconsistent or inaccurate (ie, "dirty"), or interpreted wrongly.[1] Big data in medicine is no different than what Bill Gates has quoted for business, "The first rule of any technology…is that

automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency."[14] Thus, it is important that the data be consistent and comparable.[1] Patients could be put at risk if there is a high frequency of misinterpretation concerning information such as genomic data, and a substantial effort will be required to ensure that health care providers are prepared for the era of genomic medicine and big data in surgery.[7]

Big data will revolutionize how health care is delivered and has the potential to change patient outcomes dramatically. Many challenges remain as to how big data are interpreted, integrated into physician education and hospital infrastructure, and eventually applied to patient care. Efforts on several fronts are underway currently to facilitate the transition and troubleshoot the many obstacles. In the not-to-distant future, patients with the same disease process may receive vastly different treatments that have been tailor-made to their respective transcriptomics, proteomics, microbiomics, and metabolomics.

REFERENCES

1. Chute CG, Ullman-Cullere M, Wood GM, Lin SM, He M, Pathak J. Some experiences and opportunities for big data in translational research. Genet Med 2013;15:802-9.
2. Kannry JL, Williams MS. Integration of genomics into the electronic health record: mapping terra incognita. Genet Med 2013;15:757-60.
3. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015;372:793-5.
4. Overby CL, Tarczy-Hornoch P, Hoath JI, Kalet IJ, Veenstra DL. Feasibility of incorporating genomic knowledge into electronic medical records for pharmacogenomic clinical decision support. BMC Bioinformatics 2010; 11(Suppl 9):S10.
5. Collins FS. Shattuck lecture–medical and societal consequences of the Human Genome Project. N Engl J Med 1999;341:28-37.
6. Welch BM, Kawamoto K. The need for clinical decision support integrated with the electronic health record for the clinical application of whole genome sequencing information. J Pers Med 2013;3:306-25.
7. Hazin R, Brothers KB, Malin BA, Koenig BA, Sanderson SC, Rothstein MA, et al. Ethical, legal, and social implications of incorporating genomic information into electronic health records. Genet Med 2013;15:810-6.
8. Anantasit N, Boyd JH, Walley KR, Russell JA. Serious adverse events associated with vasopressin and norepinephrine infusion in septic shock. Crit Care Med 2014;42:1812-20.
9. Dancey JE, Bedard PL, Onetto N, Hudson TJ. The genetic basis for cancer treatment decisions. Cell 2012;148:409-20.
10. Sørensen TI, Nielsen GG, Andersen PK, Teasdale TW. Genetic and environmental influences on premature death in adult adoptees. N Engl J Med 1988;318:727-32.
11. Nakada TA, Russell JA, Boyd JH, Thair SA, Walley KR. Identification of a nonsynonymous polymorphism in the SVEP1 gene associated with altered clinical outcomes in septic shock. Crit Care Med 2015;43:101-8.
12. Cuenca AG, Gentile LF, Lopez MC, Ungaro R, Liu H, Xiao W, et al. Development of a genomic metric that can be rapidly used to predict clinical outcome in severely injured trauma patients. Crit Care Med 2013;41:1175-85.
13. Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Sci Transl Med 2015;7:287ra71.
14. Investing Answers. 50 quotes from the wealthiest man in America. 2011 [cited 2015 July]; Available from: www.investinganswers.com/education/famous-investors/50-quotes-wealthiest-man-america-3088.