Big Data, Health Informatics, and the Future of Cardiovascular Medicine

Joonseok Kim, MD

“Medicine is going to become an information science. In 10 years or so, we may have billions of data points on each individual, and the real challenge will be to develop information technology that can reduce that to real hypotheses about that individual.”

—Leroy Hood (1)

A physician acquires and processes data from various sources to deliver the most appropriate treatment. The lack of high-quality data can undermine any phase of delivering care. For example, scarce information regarding the patient’s condition could lead to a wide range of differential diagnoses with high uncertainty, ultimately leading to diagnostic delay. On the other hand, high-quality, accurate, and clinically relevant information is key to making precise and informed decisions, which results in optimal patient care.

Historically, health care providers solely relied upon the patient’s history and a meticulous physical examination. The quality of medical care was determined by each provider’s ability to obtain and process the available information. In the modern medical era, we now have exponentially increasing amounts of objective information originating from complex diagnostic imaging tools (especially pertinent in cardiology), advanced laboratory results (including genomic and other “omic” data), and entirely new sources, such as from wearable, mobile, or other device technologies (2). More recently, electronic health records (EHRs) have assisted health care providers in using the desired data more efficiently and precisely. Facilitated by rapid advances in data science and driven by our everlasting effort to utilize up-to-date information, we are now facing a health informatics and big data revolution. This new wave has the potential to positively transform the way we practice medicine (3).

Unfortunately, the current medical education system during medical school, residency training, and fellowship still largely focuses on traditional methodologies for collecting data and organizing gathered information (4). A transition from these methodologies to a more modern health informatics approach, including the use of “big data,” is necessary. As such, a remarkable opportunity to help lead this effort exists for fellows-in-training (FITs) and early career (EC) cardiologists.

WHAT IS “BIG DATA” IN HEALTH CARE?

There is no uniform definition of “big data” in health care, but it is commonly characterized by the 5 “Vs”: volume, velocity, variety, veracity, and value (5,6). Volume represents the size of a dataset, usually ranging from terabytes (10^{12} bytes) to zetabytes (10^{21} bytes). Velocity pertains to data in motion and the fast speed of the generation of new data. Variety refers to data in various types and forms, and its resultant complexity. Veracity indicates the trustworthiness and inherited ambiguities due to data uncertainty and inconsistency. Value refers to the additional worth that data can bring to generate knowledge (6). These unique characteristics of big data have limited the application of conventional statistics and epidemiological approach. As a result, this new concept of big data has necessitated the development of different analytical tools (7). In the last decade, there has been an exponential growth in health data from sources such as wearable and implantable devices, smartphones, and real-time sensors (8). System capacities for data acquisition, storage, and processing have become more affordable (8). Furthermore, dramatic advances in big data visualization and analysis

From the Division of Cardiovascular Health and Disease, Department of Medicine, University of Cincinnati College of Medicine, Cincinnati, Ohio. Dr. Kim has reported that he has no relationships relevant to the contents of this paper to disclose.
techniques have enabled us to derive meaningful insights from big data (9). These advances have fueled the discussion of the role of big data in clinical practice and research (10). The proper application of these techniques is seen as a novel approach to answering clinical questions that may have been previously impossible to address. In this regard, it is the analytic method and resultant value that is the most beneficial, not the big data itself (10).

Appropriate use of big data in health care possesses tremendous promise and could be applied to multiple stages of research in cardiology, such as very large-scale population health management, cardiovascular disease risk prediction, precision medicine using genomic information, and clinical decision support through machine-learning algorithms (5,11). Papers utilizing big data and new analytics in cardiology have been sparse, but there is a growing number of publications supporting its potential utility in cardiovascular research. One study demonstrated that analysis through machine learning using a large data registry improved cardiovascular outcome prediction in patients with suspected coronary artery disease compared with the conventional risk scoring method (12). Similarly, Loghmanpour et al. (13) examined the Bayesian network algorithm model, showing superior prediction of right ventricular failure following left ventricular assist device therapy over the currently available risk prediction model. More recently, Circulation: Cardiovascular Quality and Outcomes set its theme on big data, highlighting innovative methodological approaches and big data analytics in cardiovascular outcomes research (14–16).

The Big Data to Knowledge initiative at the National Institutes of Health reinforces the importance of this research in health care (17). Despite the excitement and optimism surrounding big data and its potential utility in cardiovascular medicine, there is a significant gap between enthusiasm in the field and day-to-day use in cardiology practice, as well as in clinical research (11). This gap is partly due to the innate challenges in utilizing complex, unorganized, and large-volume medical data. It is difficult to understand the causal relationships established through the machine-learning technique; therefore, generalizability of the outcome is limited (9). Moreover, although data purification technology has improved, the process of separating signal from noise is becoming more complicated (9). In addition, cardiologists and clinical investigators are not yet familiar with the fast-growing concept of big data and its application in cardiovascular research. FITs and EC cardiologists who are prepared to surmount the barriers in the use of big data and capable of unleashing its potential in cardiology are needed.

**HOW SHOULD WE PREPARE AS FITs AND EC CARDIOLOGISTS?**

FITs have substantial advantages in big data and health informatics research. Training cardiologists of the current generation tend to be information technology (IT) savvy and familiar with computer systems and emerging technologies. Trainees are generally “digital natives” and have witnessed the rapid advances in computer science, the Internet, and smartphones. This familiarity in IT allows for a more natural understanding of health informatics and EHRs. In addition, many fellows and young cardiologists have background knowledge in mathematics, biology, computer science, and bioinformatics, which are core skill sets needed to integrate big data analytics into clinical practice and research. Furthermore, fellows actively take care of patients on the front line, which could inspire the most valuable and clinically needed research ideas that can be readily applied to clinical practice.

There are several ways to gain exposure and learn more about health informatics and data science as a cardiologist, from local to international levels. Most health care organizations have at least 1 medical informatics committee to oversee the operational and/or research aspects of the EHRs and associated technologies. Fellows can become members of these groups and participate in the various projects being carried out. Second, there are numerous online courses providing learning opportunities in data science and big data analytics at very reasonable costs or even no cost. As an example, the American Medical Informatics Association offers “10 × 10” virtual courses to train health care professionals in all aspects of health informatics (18). Fellows in a university or program that offers courses in data science and health informatics could take advantage of these classes outside of the cardiology division (19). These lecture series will help FITs and EC cardiologists gain health informatics literacy and skills. Finally, there are more advanced formal training opportunities, including fellowships in clinical informatics, as well as master degree or PhD programs in health informatics and data science. There is no uniform way to approach and learn the methodologies utilizing big data and health informatics in cardiovascular research. Any of these options will provide a solid opportunity to develop a firm foundation in health informatics and big data analytics.

Big data research and health informatics have not yet gained popularity in cardiology, but there is tremendous potential for their utility in cardiology research and clinical cardiology. Cardiovascular medicine has been a leader in the evidence-based medicine era, with advances in epidemiological
cohort studies and randomized controlled trials. As we move on to the next era of data-driven medicine, big data research and health informatics will help open the door to new insights in cardiology and transform our medical practice (20).

Big data in cardiovascular practice and research is at its nascent stage. Moving forward, translating big data for use in clinical practice will take a great effort among many collaborators, not just solitary investigators. This will be a tremendously multifaceted endeavor that requires high-level expertise in different specialties. It is clear that data scientists, computer scientists, statisticians, health informatics experts, and cardiologists need to practice team science in this endeavor. In this context, there is a high demand for cardiologists with the skills and knowledge to collaborate with other specialists in health informatics. Having proficiency in big data from very large-scale populations and a variety of sources will be an invaluable asset as a cardiologist and cardiovascular disease investigator. Cardiology fellows should be perceptive and knowledgeable in big data and health informatics concepts, and be prepared to collaborate to achieve the ultimate goal: prevent cardiovascular disease and improve cardiovascular outcomes.

ACKNOWLEDGMENT The author appreciates the valuable contribution of Dr. Eric Kirkendall for his insights, manuscript review, and comments.

ADDRESS FOR CORRESPONDENCE: Dr. Joonseok Kim, Medical Science Building, University of Cincinnati, 231 Albert Sabin Way, MLC 0542, Cincinnati, Ohio 45267. E-mail: joonseok.kim@uc.edu.

REFERENCES

RESPONSE: Moving Beyond Big Data to Causal Inference and Clinical Implementation

Peter W. Groeneveld, MD, MS
Cardiovascular Outcomes, Quality, and Evaluative Research Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; and the Michael J. Crescenz Veterans Affairs Medical Center, Philadelphia, Pennsylvania
E-mail: petergro@upenn.edu

Medicine is in the midst of a digital revolution, and the practice of cardiovascular medicine and clinical research is being thoroughly transformed by both the unprecedented expansion in the volume and variety of patient data,
as well as the development of advanced information technologies than can organize and analyze these data, providing precise, actionable information to clinicians at the point of care. In his excellent essay, Dr. Kim rightly emphasizes that cardiology fellows-in-training would be wise to prepare themselves for practicing medicine in the “Big Data” era, and he highlights opportunities for informatics-savvy physicians to influence the evolution of 21st-century cardiovascular care.

Although this enthusiasm is justified, it is also essential not to lose sight of one of the great achievements of 20th-century cardiology—namely, the establishment of scientific evidence as the cornerstone of effective clinical practice. One unfortunate aspect of the application of “data science” to medicine is that some data scientists—although truly ingenious in their mastery of data taxonomies, information technology, computing architecture, and advanced analytics—lack a firm grounding in, and are occasionally dismissive of, scientific methodology (1,2). At its core, biomedical science entails hypothesis testing with the goal of understanding causal relationships, while recognizing that chance, bias, and confounding inevitably threaten the validity of causal inference. The scientific approach may be unnecessary in many fields where data science has flourished (e.g., marketing, consumer analytics, finance, insurance, meteorology, professional sports, and so on) (2). But, it remains essential in medicine, because physicians not only need accurate predictions of future adverse events that might befall our patients, but also need a clear understanding of how our clinical actions might change the probability of those future events.

Hence, although it is arguably important that all physicians-in-training have a basic understanding of the diversity of biomedical data and the innovative ways that it can be analyzed to effect clinical care—lest they be surprised when their smartphones start recommending treatments and tests for their patients—it is absolutely vital that physicians be scientifically literate, and thus, suitably discerning in assessments of the clinical value of these Big Data applications. Improving outcomes is clinical medicine’s primary goal; better predictions are subordinate. It is also critical for physicians pursuing advanced informatics training to recognize that neither Big Data nor advanced analytics can unerringly identify causal mechanisms, and transforming predictive information into effective clinical actions that improve patient outcomes is equally important as, and often much harder than, producing precise predictions from petabytes of data.

A comprehensive understanding of the field of biomedical informatics encompasses not only the harnessing of vast and varied data sources and applying complex and adaptive algorithms to generate accurate predictions and actionable information, but also the humble recognition that these elements are not sufficient to improve patient outcomes. A robust definition of informatics includes not only the analysis of data, but also the appropriate use of information in highly complex clinical settings where human decision makers (3)—clinicians and patients—will continue to play vital roles.

REFERENCES

