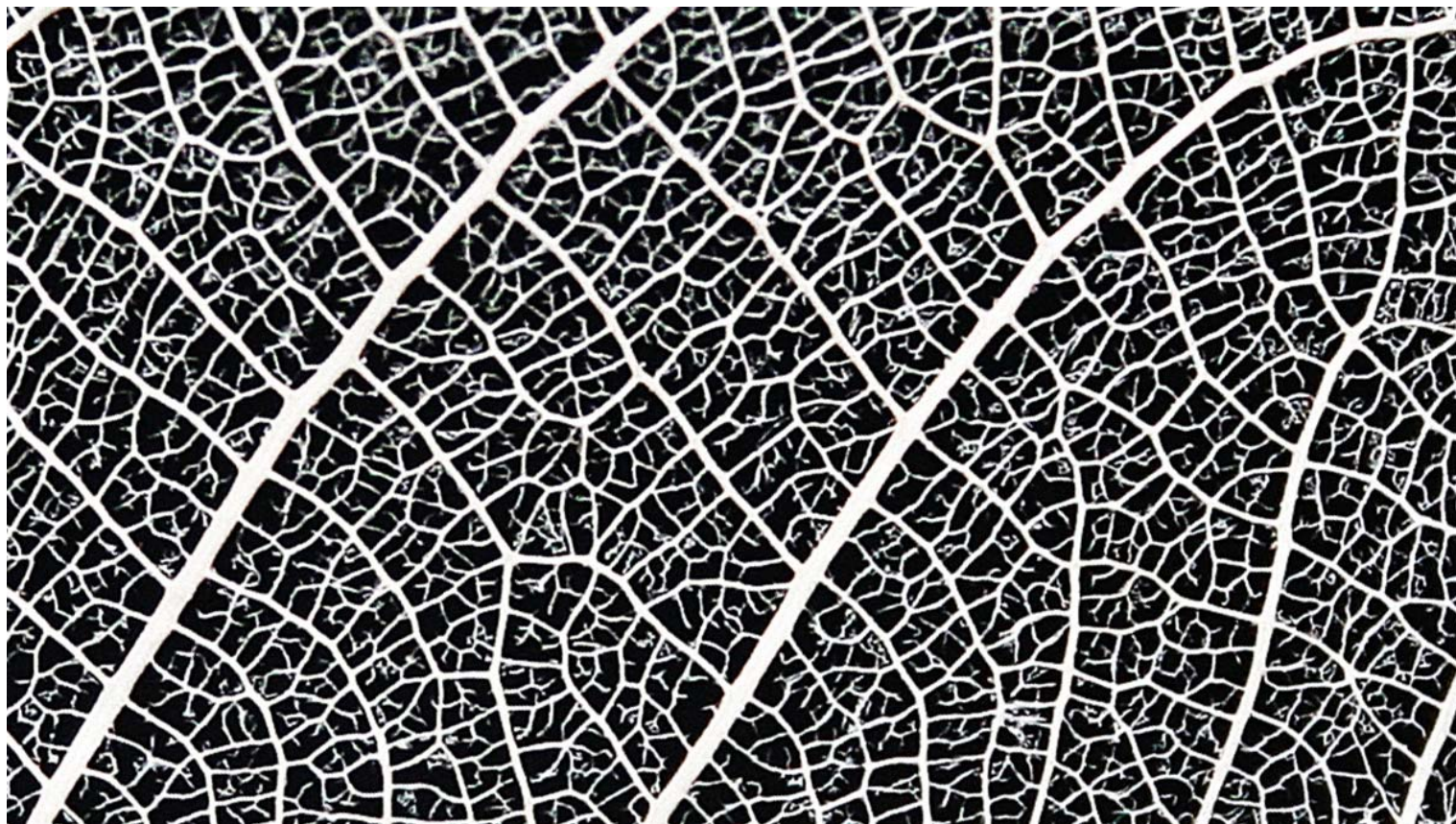


INFORMATION &amp; TECHNOLOGY

# Why Health Care May Finally Be Ready for Big Data

by Nilay D. Shah and Jyotishman Pathak

DECEMBER 3, 2014



There has been a lot of buzz about “big data” over the last few years. This is hardly surprising, given the sheer scale of the data sets that are being produced daily. A total of 2.5 quintillion terabytes of data were generated every day in 2012 alone, and it is estimated that as much data is now generated in just two days as was created from the dawn of civilization until 2003. While other industries have been far more successful at harnessing the value from large-scale integration and analysis of big data, health care is just getting its feet wet. Yes, providers and payers are increasingly investing in

their analytical capabilities to help them make better sense of the changing health care environment, but it is still early days. Here are some key elements that are crucial for health care to truly capture the value of big data.

**Integrating data.** Businesses and even political campaigns have successfully linked disparate data sources to learn “everything possible” about their citizens, customers, and clients, and apply advanced analysis and computation to modify existing strategies or create new ones. Similarly, leveraging heterogeneous datasets and securely linking them has the potential to improve health care by identifying the right treatment for the right individual or subgroup.

One of the first changes we face is the lack of standardization of health care data. The vast amount of data generated and collected by a multitude of agents in health care today comes in so many different forms – from insurance claims to physician notes within the medical record, images from patient scans, conversations about health in social media, and information from wearables and other monitoring devices.

The data-collecting community is equally heterogeneous, making the extraction and integration a real challenge. Providers, payers, employers, disease-management companies, wellness facilities and programs, personalized-genetic-testing companies, social media, and patients themselves all collect data.

---

#### INSIGHT CENTER

#### Innovating for Value in Health Care

SPONSORED BY MEDTRONIC

A collaboration of the editors of *Harvard Business Review* and the *New England Journal of Medicine*, exploring best practices for improving patient outcomes while reducing costs.

---

Integration of data will require collaboration and leadership from the public and private sectors.

The National Institutes of Health recently launched a Big Data to Knowledge Initiative (BD2K) to enable the biomedical research community to better access, manage, and utilize big data. Some early work is also being pursued through large collaborations such as the National Patient-Centered Research Network (PCORnet)

and the consortium Optum Labs, a research collaborative that has brought together academic institutions, health care systems, provider organizations, life sciences companies, and membership and advocacy organizations.

**Generating new knowledge.** One of the earliest uses of big data to generate new insights has been around predictive analytics. In addition to the typical administrative and clinical data, integrating additional data about the patient and his or her environment might provide better predictions and help target interventions to the right patients. These predictions may help identify areas to improve both quality and efficiency in health care in areas such as readmissions, adverse events, treatment optimization, and early identification of worsening health states or highest-need populations.

Equally critical is the focus on new methods. One of the reasons health care is lagging behind other industries is it has relied for too long on standard regression-based methods that have their limits. Many other industries, notably retail, have long been leveraging newer methods such as machine learning and graph analytics to gain new insights. But health care is catching up.

For example, hospitals are starting to use graph analytics to evaluate the relationship across many complex variables such as laboratory results, nursing notes, patient family history, diagnoses, medications, and patient surveys to identify patients who may be at risk of an adverse outcome. Better knowledge and efficient assessment of disparate facts about patients at risk could mean the difference between timely intervention and a missed window for treatment.

Natural language processing and other artificial intelligence methods have also become more mainstream, though they are mostly useful in harvesting unstructured text data that are found in medical records, physician notes, and social media. Mayo Clinic teamed up with the IBM cognitive computer known as Watson, which is being trained to analyze clinical-trial criteria in order to determine appropriate matches for patients. As the artificial-intelligence computer gets more information and learns about matching patients to studies, Watson may also help locate patients for hard-to-fill trials such as those involving rare diseases.

**Translating knowledge into practice.** While standardized data collection and new analytical methods are critical to the big data movement, practical application will be key to its success. This is an important cultural challenge for both those who generate and those who consume the new knowledge. Users such as physicians, patients, and policy makers need to be engaged right at the beginning, and the entire research team should have a clear idea about how the new knowledge might be translated into practice.



The insights from big data have the potential to touch multiple aspects of health care: evidence of safety and effectiveness of different treatments, comparative outcomes achieved with different delivery models, and predictive models for diagnosing, treating, and delivering care. In addition, these data may enhance our understanding of the effects of consumer behavior, which in return may affect the way companies design their benefits packages.

Translating these new insights into practice will necessitate a shift in current practices. Relying on the evidence from randomized controlled trials has been a gold standard for making practice-changing decisions. While in many cases such trials may be necessary and justified, it will be critical to identify where the evidence generated by big data is adequate enough to change practice. In other cases, big data may generate new paradigms for increasing the efficiency of randomized clinical trials.

For example, as new knowledge is gained about the comparative benefits of second-line agents for treatment of diabetes, policy makers and expert groups may consider using this information to develop guidelines or recommendations or to guide future randomized trials.

**Optum labs – a start to achieving these goals.** A proverbial village is required to make sense of the messy medley of data points that is big data in health care today, integrate and manage them in a safe and secure environment, and translate the findings into practice. One such village is Optum Labs, a research collaborative that has brought together data from the administrative claims of more than 100 million patients and the electronic medical records of over 30 million patients, along with researchers, patient advocates, policy makers, providers, payers, and pharmaceuticals and life sciences companies. The vision of Optum Labs is to boost the generation of high-quality comparative-effectiveness evidence, accelerate translation of knowledge into the development of new predictive models and tools, and improve the delivery of care.

For almost two years now, researchers from Optum Labs's 15 partner organizations have been working in this environment. The knowledge generation within Optum Labs is in its infancy, but early signs point to a significant potential to change health care with more than a 100 studies currently under way. Optum Labs's priority is to enable clinicians to connect insights from big data directly to the care of an individual patient. This is particularly important for complex patients whose care requires careful prioritization (e.g., people with "comorbidities," or multiple chronic

conditions). For example, researchers are studying the longitudinal variation in care analysis of hip and knee surgery in patients with diabetes and obesity. By analyzing the data on their treatment outcomes, we may be able to learn something that will help us create better protocols for caring for them. The hope is to gain insights about what works and for whom. Ultimately, what should drive this initiative and others is to address the complexities, unmet needs, and challenges facing patients.



**Nilay D. Shah** is an associate professor in the Division of Health Care Policy and Research at Mayo Clinic, where he focuses on studying and improving the health-care-delivery system.



**Jyotishman Pathak** is director of clinical informatics services at Mayo Clinic and an associate professor in its Division of Biomedical Statistics and Informatics. He also serves as director of clinical informatics at the Mayo Clinic Robert D. and Patricia E. Kern Center for the Science of Health Care Delivery.

## This article is about **INFORMATION & TECHNOLOGY**

 FOLLOW THIS TOPIC

Related Topics: [IT](#) | [HEALTHCARE](#)

## Comments

Leave a Comment

POST

0 COMMENTS

---

▼ **JOIN THE CONVERSATION**

---

**POSTING GUIDELINES**

We hope the conversations that take place on HBR.org will be energetic, constructive, and thought-provoking. To comment, readers must sign in or register. And to ensure the quality of the discussion, our moderating team will review all comments and may edit them for clarity, length, and relevance. Comments that are overly promotional, mean-spirited, or off-topic may be deleted per the moderators' judgment. All postings become the property of Harvard Business Publishing.