Data Mining: Healthcare

Katherine Marconi
Health Care Administration and Health Administration Informatics, University of Maryland, Adelphi, Maryland, U.S.A.

Matt Dobra
Department of Economics, Methodist University, Fayetteville, North Carolina, U.S.A.

Charles Thompson

Abstract
Current practices and today’s information technology investment and strategic decisions can either promote or limit tomorrow’s successes. This entry discusses the types of big health data and its impact on patient, provider, and organizational health decision making. The entry ends by discussing possible future trends and threats to using big data to improve the delivery of health services.

INTRODUCTION

Big data has many attributes that apply to the large electronic sources of health data being created, managed, and analyzed by healthcare providers, health organizations, and patients and their families. Data from genetic mapping, pharmaceutical tracking, public health reporting, digital x-rays, computerized axial tomography scans and laboratory results, payer and provider data, insurance claims data, and consumer online behavior adds up to petabytes of information. What makes this data so exciting is that big data has the potential to improve individual and population health, make the business of healthcare more cost-effective, and lead to new treatments of chronic and infectious diseases. In healthcare, the success of enterprisewide electronic information will be measured by its contributions to improvements in individual and population health.

We are in an era of availability of health data that enables us to transform the data to usable health information and devise better ways to manage individual and population health outcomes. But the ability to combine data into large and useful information remains a significant challenge and will take unexpected twists and turns before its full potential is realized. Current practices and today’s information technology (IT) investment and strategic decisions can either promote or limit tomorrow’s successes. In this entry, we discuss the types of big health data and its impact on patient, provider, and organizational health decision making. The entry ends by discussing possible future trends and threats to using big data to improve the delivery of health services.

Some view data[1] as being “big” because it is just ahead of the culture and time period’s methods of data storage and analysis. Big data combines information from different sources and is analyzed to change our practices; it should improve patient outcomes and improve the nation’s healthcare delivery system. This concept of rethinking health information is not a new one. In 1854, John Snow,[2] a founder of epidemiology, modernized methods of how we investigate and treat epidemics, specifically the transmission of cholera. He collected data in a new way, combined it with nonhealth information, and thought differently about it. Although his information covered slightly less than 200 sick individuals, by mapping their location along with the locations of noninfected individuals and the London water supply, he produced “big data” for that time period. He identified the source of a cholera epidemic, how it could be stopped, and introduced us to population-based health. John Snow pictured commonly available information differently and stopped an epidemic from spreading (Fig. 1).

Sources of today’s big health data can be grouped onto four categories based on American Informatics Management Association informatics domains:

- data associated with the delivery of clinical care
- public health survey and surveillance information
- genetic and medical research-related information
- healthcare-consumer-driven information

Big data is not simply drawn from each of these sources; it relates information among them in new ways. It also links to other available social and economic information. For example, it may involve linking traditional health information with nonhealth information, such as sales volume, to track patient behaviors or health conditions. Health managers, as they plan enterprisewide IT systems, need to consider these external
and internal sources of information that are available for their decision making.

Another modern example can be found in the genetic mapping of the 20,000–25,000 human genes and the underlying billions of DNA pairs. The National Institutes of Health (NIH) 1000 Genomes Project has made the data freely available on the web for research, the equivalent of “30,000 standard DVDs”. Because of the Human Genome Project, we now have screening tests available for a variety of inherited diseases and many potential avenues for advancing treatment. It is a model for shared medical research information that is available to others for further analyses.

Thus, the core principle of big data in health is the ability to combine large amounts of information using different analytic methods to improve clinical and related service delivery decision making. But we should also be aware that because big data influences how we make decisions, it may lead to changes in our organizations’ structures and cultures. Big data necessitates working in clinically led teams, rather than the traditional physician-driven care model. It involves sharing of primary medical information among researchers, public health agencies, patients (consumers), and health services. Big data is changing the way we share health information and deliver healthcare.

**TYPES OF BIG HEALTH DATA**

Combining clinical, public health, research, and consumer health data into meaningful information is challenging. Medical decision making is very complex, and recording it involves textual information, not just coding. While common data definitions for clinical conditions are in place, such as the *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9-CM) for diagnosis coding, there are gray areas that require further clarification and consensus such as definitions of individual characteristics in research studies and different versions of Health Level Seven International messaging standards. The many available software systems to choose from also add to this complexity.

There are policy issues of confidentiality and privacy, where individual information needs to be pooled for analysis without identifying the person. At the same time, there is the need to protect business-sensitive information in a very competitive and regulated medical environment. While these issues exist in other industries, they are magnified in healthcare and have become barriers to realizing the potential of big data. Clinical services, public health, medical research, and consumer-driven information share these common barriers to contributing to care improvement.
CLINICAL SERVICES DATA

For healthcare providers to realize the potential for clinical data to improve their practice and patient outcomes, their organizations must have the technology and capacity to relate information from a number of data sources, including unstructured data and visual information. Not only is this data large, but organizations must be able to acquire it, store it, and analyze it in real time to produce meaningful information for clinical decision making. In this context, meaningful information means results that are easily understood by clinicians, support staff, and administrators (depending on the system).

Clinical Decision Support Systems (CDS) layer on the analytic software to translate clinical data into real-time information for clinical decision making. They apply rules to patient care information to indicate contradictions in care or other outliers. The rules may be a combination of medical expertise and analysis of past illness, diagnoses, and treatment patterns. For CDS to improve care, the system must be acceptable to clinical providers and easily fit into the complex patient—provider workflow of organizations. In one example where the fit was not completely thought through by system implementers and users, information from a CDS, bypassed the nursing information and had the potential to lead to medication errors. As an Agency for Healthcare Research and Quality whitepaper indicates, the timing and ownership of CDS systems are essential to their success.\[5\]

Clinical big data, however, is not only useful for individual patient care; it also makes the individual part of a population. For relatively rare conditions, where previously a specialist might ask one or two colleagues for a second opinion, large clinical data sets give the provider (or clinician) the ability to review treatments for additional patients with similar diagnoses, giving them additional data for clinical decision making. It also provides a base of information for monitoring disease trends, service usage, and quality of care. The National Notifiable Disease Surveillance System operated by the Centers for Disease Control and Prevention (CDC) is a good example. Symptoms are documented in an electronic health record (EHR) at the clinical encounter level, and a diagnosis is coded and entered into a database. The ability to view, aggregate, and analyze this data enables public health practitioners to monitor the occurrence and spread of diseases. As in the John Snow example, clinical data leads to population health management.

To improve quality of care and to change care patterns, big clinical data is impossible without building comprehensive EHRs, longitudinal health records of an individual’s health. Comprehensive EHRs include diagnoses, problem lists, present and past medications, results of tests, and treatments from different units and facilities that are accessed by individuals. They form the basis of CDS and other analytic systems. While the percentage of physicians adopting some form of EHR doubled between 2008 and 2011, this percentage still is only 55%.\[6\] Reports from State Health Information Exchanges also show limited progress in information sharing among hospitals and physician practices, but the information frequently is limited to demographics and pharmaceutical information.

This limited data collection and sharing is apparent in the Beacon Community Program grantees.\[7\] The federal government funded them to provide prototypes of electronic medical record systems. They are important pilot projects for comprehensive EHRs, but most focus on linking information for specific diseases, such as diabetes, heart disease, or asthma or partial health facility functions, rather than the comprehensive data needed to cover patient care that encompasses many different conditions at different health facilities. Fig. 2 summarizes the current clinical uses of electronic clinical information as described by physicians. The figure indicates the variety of functions that EHRs contribute to as they become more common and comprehensive.

If one looks at all of the certified health IT systems approved through the federal EHR technology program, it is a time of experimentation and flowering of platforms to create large clinical health data systems. The Office of the National Coordinator for Health Information Technology’s\[6\] (2012) certified Health IT Product List provides a myriad of systems meeting meaningful use requirements. The systems are needed to promote data standardization that will allow data exchange (interoperability) among organizational entities and their many IT systems. A software company executive states that with any type of acquisitions companies can have “from 50 to 70 business systems alone”.\[8\] Multiple EHRs need to be integrated into a manageable number of systems that are interoperable, thus easily transferring information from one system to another.

Healthcare managers today have many options in planning their enterprisewide EHR solutions. Managers may choose to opt for commonly used systems, such as Wexler Medical Center’s use of the Elderly Pharmaceutical Insurance Coverage system in its four hospitals to link facility functions, including its inpatient system, emergency room system, revenue cycle system, patient scheduling, and operating room system.\[9\] Others have opted to build interoperability among existing systems and on integrating CDS systems within them. A good example of interoperability can be found in the work of Health Information Exchanges that are creating interfaces among different physician and hospital electronic systems, allowing exchange of patient data to facilitate efficient healthcare delivery.

Of importance in choosing EHRs with big data in mind, providers and administrators of clinical services need to decide how to store the large amounts of data available to them in forms that facilitate their real-time
analyses for quality improvement. Some of the key management decisions to enable these systems to produce big data include the availability of standardizing for different order sets; security, multiple clinical services, and clinician teams within facilities; enterprise management; imaging software development; and linking to analysis and knowledge management applications. The National Institute of Standards and Technology’s[4,10] user-centered design provides one process to guide organizations through these decision-making processes.

The federal government has provided both support and constraints for the growth of electronic medical records and big data. Meaningful Use requirements, which are tied to federal Medicare and Medicaid Incentive Payments, are facilitating this development through the definition and required reporting of health measures and usability standards, such as pharmaceutical interaction checks.[11] Meaningful Use has three phases that are being phased in through 2015: Stage 1 is data capture and sharing basic clinical information; Stage 2 focuses on capturing and sharing advanced clinical processes; and Stage 3 captures improvements in quality, safety, and efficiency. Because it is being implemented in different phases, it can be used as a guide for the development of comprehensive EHRs.

Because Meaningful Use focuses on the development of a limited set of common measures that must be reported to the federal government, it may focus providers on meeting federal standards rather than on developing a comprehensive EHR that meets their own needs. Additionally, the construction of usability standards for healthcare data is lagging behind other federal health EHR standards and requires further development, and federal rules for patient consent and information sharing need to be reconsidered, given this emerging era of big data.

The federally driven Accountable Care Organizations (ACOs) have potential to link patient care among organizations through EHRs. They can provide new opportunities for amassing the wealth of health information available into large data sets for clinical decision making. Funding for ACOs is tied to patient care across different services, necessitating EHRs that can be used to analyze populations of patients, rather than just individuals. No matter which direction the organization of clinical health services takes, EHRs are the wave of the future. In What’s Ahead for EHRs: Experts Weigh In, The California Health Foundation[12] discusses the next generation of EHRs: innovative systems that facilitate the use of large pools of information. Their report and a review of current health big data efforts show that we are only in the initial stages of using big data to improve health outcomes.

PUBLIC HEALTH SURVEY AND SURVEILLANCE INFORMATION

Public health information systems range from those that store individual health information for public health surveillance, such as immunization or infectious disease reporting, to real-time alert systems for drug interactions, disseminating research findings, and reporting unlikely clusters of unknown or rare conditions. The focus of this information is on protecting the public’s health, rather than individual health. Information from health surveillance systems, such as CDC’s Notifiable Diseases reporting, and surveys, such as the National Interview Health Survey, has existed for decades. But as
pointed out by experts in this area, their information often is not linked or interoperable among reporting organizations, including the cities and states involved in the surveillance systems. The growth of public health information has outpaced our capacity for storage, interpretation, and use. Similar to clinical care, it is an area where coordinated efforts are needed among health facilities and with city, state, and federal health agencies.

An early 21st century World Health Organization (WHO) surveillance system is a potential model for future worldwide surveillance. During the severe acute respiratory syndrome epidemic, the WHO created a virtual laboratory model using the phone, video, and Internet to monitor and respond to the outbreak. With today’s advanced technology, efforts can go beyond this and allow for even quicker reporting, analyses of information, and responding to unusual health events.

One recent example of analyses of a large set of information for improving population health was published in the September 21, 2012, Morbidity and Mortality Weekly Report. By combining time-trend rat inspection information with census data that covered a population of 770,000 individuals and approximately 35,000 residences in 12 Bronx, New York, neighborhoods, New York City was able to estimate the prevalence of rats, a known health risk.[13] To realize the full potential of this information as big data, the city might consider linking this data with neighborhood clinics and other surveillance information.

The U.S. government is taking some steps to grapple with its diverse health incident disease reporting systems. BioSense 2.0 is an effort by the CDC to reduce the costs and increase the feasibility of state and local data systems that will communicate with each other. The Food and Drug Administration[14] is undertaking a similar effort to monitor product safety—the Sentinel Initiative to pool existing resources so that large amounts of data can be processed to quickly produce needed information. It also is supporting a Virtual Laboratory Environment to produce innovative analytics for using the information currently available throughout the United States.

Both policy and resource issues need to be addressed to make large sets of public health data available for linkage and analysis to improve population health. First is the flow of information between and among cities, states, and the U.S. government. Data sharing and linkage with each other and with other surveillance information is limited because of a lack of data standardization, structured ways to get clinical information into EHRs, nonuse of standard health information exchange protocols, and the privacy and security required for records that identify individuals. Additionally, updating historical surveillance and survey systems requires resources frequently not available to state and local health agencies. These are not insignificant barriers and need to be considered when prioritizing future forays by public health into big data.

**MEDICAL RESEARCH DATA**

Within the medical research community a huge amount of information exists, but it is tied to specific grants and institutions. Tension exists between the pull to share information for quicker development of new treatments and the need to patent information to protect profits. At the national level, The NIH has a number of disease-based initiatives to share information for use in further research. Besides its Genomes Project, for example, the National Heart, Lung, and Blood Institute,[15] NIH has developed the Cardiovascular Research Grid and the Integrating Data for Analysis, Anonymization, and Sharing initiatives to enable researchers to easily store and share information. It not only uses technology to store information but also emphasizes communication and education about the system. The success of the Cardiovascular Research Grid and other efforts will be measured in their ability to advance prevention, diagnosis, and treatment beyond the individual research results and meta-analyses that we see today.

Medical research also is advancing in its use of big data through mathematical modeling. Pharmaceutical companies are using predictive modeling to design new drug formularies and to modify existing ones. From creating mathematical models for neurology clinical trials, to characterizing the genetic determinants of heroin abuse, and to geomapping infectious diseases spread, disease modeling is an important medical research tool.

Once again, common definitions, data standardization, and advanced analytical software will facilitate sharing of huge data sets among researchers. The PhenX project led by RTI International and funded by the National Human Genome Research Institute[16] is one example of how to plan and produce big data for genetic research and ultimately impact public health. Genetics and epidemiological research are being integrated to provide researchers with high quality and low burden measures that can be included in genome-wide association studies and other types of studies. With large population studies producing large amounts of information about exposure to potential carcinogens, weak causal relationships, such as the association of environmental factors with genetic characteristics, now can be studied. But barriers to pooling data and the meta-analysis of existing studies include lack of common exposure measures and associated analytics. The PhenX project is producing a toolkit to solve these barriers; it is stretching the science of medical research analytics.
A 2010 Position Statement on Data Management and Sharing signed by 17 organizations from five countries highlights the policy and political barriers that need to be overcome for big data in medical research to reach its full potential.\(^\text{[17]}\) The agreement points to the current complexity of country policies and procedures for sharing research information and it defines the principles by which such data sets can be used by others to improve the public’s health. Signatories include NIH, Agency for Healthcare Research and Quality, and the Bill and Melinda Gates Foundation. A Call for Action on Health Data from Eight Global Agencies, including the WHO, states similar principles for the timely sharing across countries of “health financing, health workforce, service access and quality, intervention coverage, risk factors, and health status” information. Its principles center on “developing a common data architecture, strengthening performance monitoring and evaluation, and increasing data access and use”.\(^\text{[18]}\)

These agreements show that the major public and foundation funders of medical research agree on the principles of data sharing. However, the mechanics of storing and accessing data sets still are being worked out. In a 2011 conference, participants stressed that the technological systems for research repositories exist; it is the impact of data sharing on research careers, intellectual property, and profits that must be agreed to. This is especially true in the pharmaceutical industry, where big health data is a reality.

**CONSUMER-CENTERED INFORMATION**

Patient-driven care is a commonly used concept in health services. The Institute of Medicine (IOM) defines patient-centered care as: “Health care that establishes a partnership among practitioners, patients, and their families (when appropriate) to ensure that decisions respect patients’ wants, needs, and preferences and that patients have the education and support they need to make decisions and participate in their own care.”\(^\text{[19]}\) Its meaning can vary but its underlying concept is that individuals manage their health status by actively seeking information about their health and that they and providers communicate with each other. These efforts, along with capturing and analyzing consumer-driven health information, are caught up in the lack of comprehensive EHRs and connected surveillance systems. Because most health services IT efforts focus on EHRs and other electronic records attached to organizations, rather than tied to consumers, the development of comprehensive longitudinal health records remains a challenge. While consumer-accessible medical records and information is recognized as important, the business models for organizing and analyses still need development.

In a 2011 survey, CDC reports that less than 50% of people use the Internet to learn about health information. Less than 10% communicate with providers by e-mail.\(^\text{[20]}\) Security and privacy concerns are a major factor in limiting this interaction. But there is a huge potential for big data analyses of consumer-driven information. This potential includes not just Internet information patterns but usage of remote patient monitoring for conditions such as diabetes or asthma and other electronic devices.

One model for the analysis of consumer Internet behavior can be found in a 2003 National Cancer Institute funded study. Researchers Bader and Theofanos\(^\text{[21]}\) partnered with Ask Jeeves to analyze the feasibility of measuring cancer hits on Ask.com. Their analysis showed the types of cancers queried and the types of content queried, such as symptoms or treatment. Their methodology forms a framework for today’s much larger Internet-driven health data analyses. Not addressed in their article is how this information then could be used to improve consumer searches—the purpose of big health data. A more recent article by Socha et al.\(^\text{[22]}\) maps information about users of a library-based phone health literacy service with Census information. The authors found that combining information can identify geographic areas and populations that the phone service is not reaching.

Another source of big data is remote patient monitoring. Remote patient monitoring, which produces real-time information not just for individual behavior but for patterns of behavior and associated treatments, is expected to more than double by 2016—from a $8.9 billion to a $20.9 billion market.\(^\text{[23]}\) The data produced requires systems that can handle large amounts of information, especially if visual imaging is remotely transmitted, but not only is detection of illness made easier for consumers, it also presents opportunities for analyzing areas of business growth.

For healthcare executives, now is the time to lay out strategies for the roles that consumers will play in their organization’s service delivery electronic interactions. At one end of the spectrum, consumers can be made part of a clinical “shared decision-making” process. A paper by Swan\(^\text{[24]}\) shows how consumer involvement might be achieved. They are educated about their options and listened to regarding their wishes for clinical procedures. In the middle, there will at least be opportunities for communication and questioning of medical personnel using mobile phone texting and other electronic devices. At the opposite end of the spectrum, consumers can be viewed more passively as readers of web information on health, with companies then analyzing and shaping where their information comes from. Health organizations need to define how they will interact with their public before they consider their IT systems.
At the least, they should have plans for the analyses of their market’s Internet behaviors that can be used to build new consumer services, attract new patients, and retain existing ones. The websites of large health organizations, such as Kaiser Permanente and the Cleveland Clinic, contain a wealth of medical information waiting to be mined for consumer use patterns. These organizations also allow patients to use the Internet to access their medical information and to interact with physicians, providing opportunities for analyses and improvements in their business processes.

CREATING ANALYTICAL TOOLS THAT DELIVER INFORMATION FOR CLINICAL AND BUSINESS DECISION MAKING

Big data in health must draw from multiple IT platforms and multiple types of information, ranging from text to disease coding and billing information. Health organizations first need to resolve these types of IT issues so that analytics can be created to produce real-time and useful information. A number of software tools are available for large data sets. Some are specific to one area of health, such as CDC software for analyzing specific surveillance and survey data sets; others are not specific.

Choosing one or more analytical tools starts with common definitions. One of the more complex areas in healthcare that must be made manageable before EMRs and big data sets can be built is getting clinical data into the EHR, such as through the use of clinical order sets. In a case study on CDS Systems, Clinovations\(^{25}\) started with approximately 1300 computerized provider order entry sets that physicians used in six hospitals. Through a consensus process with the clinicians (that also could have been augmented with statistical modeling of order set data) all clinicians were given a chance to develop standard order sets. The result was 354 electronic order sets for use in an EMR and CDS system.

Big data also requires a skilled analytic workforce that combines research and statistical skills frequently found at universities, large public health agencies, and consulting organizations in addition to clinical staff involved in the delivery of health services. Thus, an unlikely combination of health data management and software skills, statistical analyses, experienced medical care, and data literacy is needed. In essence, while discussions of the big data workforce frequently concentrate on data scientists or analysts, a team approach is required in healthcare. The combination of medical knowledge, engineering, computer science, and communication is too rare a skill set for an organization to depend on in one person. For the healthcare executive to release teams for this work means considering the time, staffing, and resources that must be devoted not just to information storage but to end uses including analytical and decision-making processes. It requires ensuring that clinical and other patient staff have time built into their schedules for adapting clinical decision-making systems to their institution’s needs.

There are broad trends in data analysis software that are likely to provide lasting value to health analytics. For example, Software for the Statistical Analysis of Correlated Data (SUDAAN\(^{26}\)) is widely used for survey data. MATLAB\(^{26}\) is a powerful tool used for structural modeling, EViews is popular among people interested in analyzing time-series data, MapReduce/Hadoop are a Java-based combination frequently used for data-mining applications, and Statistica and JMP are increasingly used. Other specific applications’ main purpose is the displaying of data, such as geographic information systems software. In smaller practices and specific health clinics, Microsoft Office tools Excel and Access are frequently used for data analysis. While Access is capable of limited data mining and Excel is capable of basic statistical analysis, neither is a robust replacement for a dedicated software package or for storing big data sets.

For clinical and health business data sets, Statistical Analysis System (SAS) and Statistical Product and Service Solutions (IBM/SPSS) often are the analytical software of choice, whereas among researchers the usage of SPSS lags far behind that of Stata and SAS. For example, in a study analyzing the use of statistical packages across three health journals in 2007–2009, Dembe et al.\(^{26}\) found from articles mentioning the statistical programs used that, 46% used Stata and 42.6% used SAS, while only 5.8% used SPSS. Robert Muenchen’s research\(^{27}\) indicates that among academics, a wide variety of biomedically targeted statistical programs, most notably Stata and R, are quickly increasing in market penetration.

SAS, SPSS, Stata, and R are examples of how each analytical package has different costs and advantages. The pricing agreements they have vary with the different software publishers. R, as open-source software, is free. Pricing for Stata 12 varies by the version; for example, one of the cheapest versions that can be purchased allows datasets with up to 2047 variables and models with up to 798 independent variables, with a more expensive version allowing for datasets with up to 32,797 variables and models with up to 10,998 independent variables. The licenses for SPSS and SAS, on the other hand, are annual licenses. The pricing of SPSS is generally such that many of the statistical tools that are included in the full versions of SAS and Stata require the purchase of additional modules that can quickly inflate the purchase cost of SPSS.

In addition to the cost advantage, R and Stata benefit from their easy and relatively rapid extensibility. While the capabilities of each of these software packages has
increased over time, the user bases of both R and Stata contribute extensively to the computational power of these software packages through the authorship of user-written add-ons. As a result, Stata and R users generally do not have to wait for the new, cutting-edge techniques to be incorporated into the base version of the software—many have already been written by users, and those with an understanding of the programming languages can script their own.

While Stata and R have an advantage in cost and extensibility, the relative strengths of SAS and SPSS are in the analysis of big data. Using Stata and R is far more memory intensive than SPSS or SAS. This advantage, however, is quickly disappearing with developments in computing, particularly the move from 32 bit Windows to 64 bit Windows. Recent extensions to R further reduce this limitation, allowing data sets to be analyzed from the cloud. Related to this, SAS and SPSS also have an advantage in the actual modeling of big data, particularly in the realm of data mining. SPSS Modeler (formerly Clementine) and SAS Enterprise Miner offer a full suite of data-mining techniques that are currently being developed by R users and are mostly absent from Stata.

Some of these modules are essential to many health scientists, including modules for dealing with survey data, bootstrapping, exact tests, nonlinear regression, and so on. R is always no more expensive than SPSS and SAS; and in the long run, Stata is usually cheaper than SPSS and SAS. These very different costing structures show the time and expertise needed in choosing analytical software.

User-friendliness is certainly one of the many concerns when considering statistical programs. There are likely to be large differences across purposes of what defines user-friendly, in particular between academic and health business settings. As a result, the criteria for user-friendliness is likely to differ across purposes; while decision makers in a corporate setting are likely to view the quality of the graphical user interface as the most important element of a software’s user-friendliness, academics will typically view the ease of coding as contributing the most to ease of use.

SUCCEEDING IN A BIG DATA CULTURE

As discussed in the beginning of this entry, the success of big data in healthcare will be judged by its ability to integrate health and nonhealth information and produce real-time analyses that improve patient outcomes, overall population health, and related business processes. Big data takes the paper-based quality improvement mantra of Plan, Do, Study, Act and brings it into the electronic age.[28] This will mean continual changes in the way medicine is practiced and services and research projects are managed, and in every aspect of healthcare delivery. Big data has the potential to change the relationship of consumers and the industry.

The McKinsey Institute Big Data Study points out that the U.S. healthcare system is at a crossroad. It must develop comprehensive EHRs, standardize the way information is collected, and turn it into useful information. If information is able to be standardized and shared, it then can influence patient care and health outcomes. One story that shows how pervasive change must be in our health culture is the transforming effect of patient satisfaction data on health services. We often think of the outcomes of healthcare in terms of patient health and illness severity. But another dimension is patient satisfaction with a facility’s services—its cleanliness, the friendliness of staff, and the food that is served. When one hospital set up an ongoing system for measuring and monitoring these dimensions, it was able to make practice changes that raised abysmal patient satisfaction rates. The system led to efforts to instill a culture of service throughout the organization, affecting staff from cleaning crews to surgeons. The facility may not have been able to compete on specialty services with other area faculties, but because it can use data for continuous quality improvement, it can now compete using positive patient experiences as a competitive marketing tool.

As further development occurs in this facility and it is able to link patient satisfaction experiences with patient and care characteristics, it will realize the potential of big data. Similarly, when surveillance data is routinely linked with census and environmental information, the potential for using this information to pinpoint and act upon population health issues greatly increases. Health in present times is a business, with government public health agencies also adopting common business practices. Big data in healthcare, when it is available electronically, has the potential to make healthcare more efficient and effective.

REFERENCES
