By Barry P. Chaiken, MD, FHIMSS

HEALTH & IT QUALITY

Big Data Drives Big Change

Every MBTA bus in Boston carries a networked sensor that broadcasts the location of the bus along its route. This allows smartphone users to know exactly when the next bus will arrive at their corner stop, and MBTA supervisors to monitor the performance of drivers.

Every new General Motors automobile includes an event data recorder (EDR) that captures information about the car's performance during an accident. Law enforcement and insurance companies use this information to help determine the cause of accidents, while car manufactures utilize the information to assist them in designing safer cars.

Every mobile phone captures location using either GPS or cell tower triangulation. Smartphone owners use their phones as navigation devices, while retailers use this information to influence search results on these same smartphones.

Exhaust Data

The digital age is the age of big data where every piece of technology captures data available for later use. The McKinsey Global Institute (MGI) describes data generated in this way as digital "exhaust data," data that are created as a by-product of other activities (Manyika, 2011).

Wikipedia defines big data as:

a term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Big data sizes are a constantly moving target currently ranging from a few dozen terabytes to many petabytes of data in a single data set.

Nicholas Negroponte in his 1995 book *Being Digital* first hinted at the value of big data in this way:

I recently visited the headquarters of one of America's top five integrated circuit manufacturers. I was asked to sign in and, in the process, was asked whether I had a laptop computer with me. Of course I did. The receptionist asked for the model and serial number and for its value. "Roughly, between one and two million dollars" I said. "Oh, that cannot be, sir" she replied. "What do you mean? Let me see it." I showed her my old PowerBook and she estimated its value at $2,000. She wrote down that amount and I was allowed to enter the premises. The point is that while the atoms were not worth that much, the bits were almost priceless.

Big Value

The rapid expansion in the use of EMRs and digitally-driven technology—MRI scanners, body sensors, automated lab tests—brings the era of big data to healthcare. MGI estimates that big data presents a $300 billion potential annual value to the U.S. healthcare system. The five broad areas to deliver that value are: 1) clinical operations, 2) payment/pricing, 3) R&D, 4) new business models, and 5) public health. Sub-areas include comparative effectiveness research (CER), clinical decision support, remote patient monitoring, health economics, and personalized medicine.

Organizations that properly collect, analyze, and utilize big data will achieve a significant competitive advantage over those organizations that fail to recognize the opportunity big data presents.

Healthcare Data Sources

The four large data sources for healthcare include clinical, pharmaceutical, administrative, and consumer (Table 1).

The large investment in EMRs and the increased use of digitally connected medical devices drives the rapid expansion of available clinical data. As these technologies evolve, the data collected becomes more expansive and granular, yet poorly utilized.

Pharmaceutical and medical device companies collect clinical trial data to substantiate the safety and efficacy of their products. Although the clinical trial group represents a subset of the real target population, limited analytics aided by expert opinion provide the basis for a suboptimal product review process.

Both payors and providers utilize administrative data to monitor their business practices. Lacking in clinical meaning, analysis often leads to poor

Table 1

Data Source	Examples
Clinical	Electronic medical records, images
Pharmaceutical, medical devices	Clinical trials
Administrative	Utilization, claims, cost
Consumer	Behaviors, home monitoring, retail purchases

decisions based upon erroneous results gleaned from analysis of incomplete data sources.

As the availability of information from consumers grows with their use of technology, retail entities utilize the data to assist in the management of their businesses. The expanded deployment of remote patient monitoring devices and the collection of data points through social media and consumer monitoring programs (e.g., pharmacy purchases with an affinity card) offer additional data sets unavailable only a short time ago.

These data sources present a valuable area for analysis by researchers striving to find ways to improve care delivery while lowering costs. For example, CER utilizes patient and outcomes data to determine which therapeutic approaches deliver the best results. Such work requires the analysis of disparate data sets covering multiple clinical and administrative information sources collected and controlled by different providers and payors across varied treatment settings.

Linked Data

New analytic tools such as Semantic Web 3.0—linked data—offer ways for machines to analyze these data sets leveraging approaches impossible using standard relational databases and statistical methodologies. These new tools permit researchers to work around the barriers presented by data sets' nonconformance to standards for data collection or storage.

Similar to the use of metadata, Semantic Web techniques allow the assignment of descriptors to each data point, providing a context and meaning to the data. This allows machines, applying powerful statistical techniques, to analyze the disparate data sets in ways not available to humans alone due to the data sets' size and complexity.

The knowledge obtained from big data offer additional benefits to healthcare. CER delivers medical knowledge that can be applied using clinical decision support tools deployed at the point of care.

The analysis of subpopulations allows for the delivery of personalized medicine that accounts for genetic variation between and among ethnic groups.
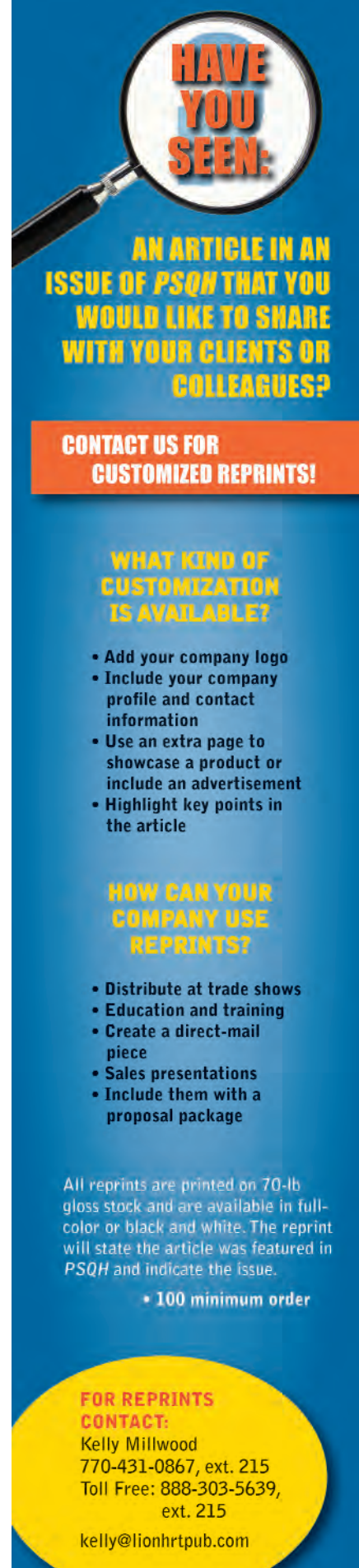
Big data applied to health economics and outcomes research facilitates the development of performance-based pricing plans that reward quality outcomes rather than incentivize utilization. Accountable Care Organizations will derive great value from using big data.

The uses of big data are numerous and far-reaching. Only through innovate analytical techniques will we be able truly to leverage the healthcare data collected and improve the way we deliver care. ▌PSQH

Barry Chaiken is the chief medical officer of DocsNetwork, Ltd. and a member of the Editorial Advisory Board for Patient Safety & Quality Healthcare. *With more than 20 years of experience in medical research, epidemiology, clinical information technology, and patient safety, Chaiken is board certified in general preventive medicine and public health and is a Fellow, and former Board member and Chair of HIMSS. As founder of DocsNetwork, Ltd., he has worked on quality improvement studies, health IT clinical transformation projects, and clinical investigations for the National Institutes of Health, U.K. National Health Service, and Boston University Medical School. He may be contacted at bchaiken@docsnetwork.com.*

REFERENCES

Big data. (2012, January 5). In *Wikipedia, The Free Encyclopedia*. Retrieved 15:41, January 6, 2012, from http://en.wikipedia.org/w/index.php?title=Big _data&oldid=46975701

Chaiken, B. P. (2011) Web 3.0 data-mining for comparative effectiveness and CDS. *Patient Safety and Quality Healthcare, 8*(5), 8-9.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011, May). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.

Negroponte, N. (1995). *Being digital.* New York: Alfred A. Knopf. Available at http://archives.obs-us.com/obs/english/books/nn/ch01c01.htm