**Big Data in Healthcare**  
**Christian Chai**

**Topic Overview and Introduction**

As healthcare transitions from a fee-for-service payment structure to a value-based model, big data will play an increasingly large role in how we treat patients. Big data, or large sets of relevant information, has the ability to revolutionize the industry. With patient records for entire populations only a few keystrokes away, big data will allow healthcare professionals to significantly raise the standard of care.

The raison d’être of big data is to improve data transparency and usability, thus unlocking significant value in large information sets. Proper data collection and analysis will result in more informed decision-making by healthcare professionals and patients alike. This new source of information can give additional insight to caregivers by allowing them to see acute cases within the framework of an entire population. By analyzing patterns amongst millions of prior cases, the big data business provides doctors with an objective tool to effectively treat patients. These analytics will also be valuable to payers and vendors as they progress with the industry toward value-based care. Data-optimized care should then ideally be synonymous with better and more efficient care, which is in the interest of all industry stakeholders.

Big data has already gained traction in other sectors, generating over $30 billion in revenues last year alone (Robertson, 2012). This precedent gives reason to be optimistic: as healthcare begins to adopt this invaluable tool, big data should have a significant transformative effect on the industry.

**Recent Updates**

On June 5, 2012, The Centers for Medicare & Medicaid Services (CMS) announced a new data and information initiative to be administered by the newly created Office of Information Products and Data Analytics (OIPDA). As a new CMS unit, OIPDA will be charged with managing the CMS health data portfolio. By developing, using, and disseminating data and information resources, CMS aims to spearhead healthcare’s evolution into a value-based model of care ("HHS harnesses," 2012).

As part of this initiative, OIPDA announced several new big data and information resources to be available this summer. These products, which include comprehensive data sets, an online peer-review journal, and a search tool for CMS information, are designed to allow data to be leveraged by all stakeholders within the healthcare industry. Through this, OIPDA seeks to foster innovation with big data by defining and promoting higher quality, lower cost care while maintaining patient privacy rights (Ibid).

Other healthcare stakeholders are beginning to utilize big data to their benefit. At the moment, big data’s industry penetration is still relatively minimal.
compared to that of other sectors such as consumer IT. However, many prominent technology companies have developed big data solutions suited for healthcare.

**Analysis**

As the largest insurance payer in the United States, the federal government has unsurpassed ability to stimulate structural change within the industry. The recent establishment of OIPDA is a strong sign that big data is increasingly regarded as the next solution in healthcare.

With further market saturation, big data will confer numerous benefits to the healthcare industry. The result of this shift toward big data will be improved efficiency and quality: the US healthcare sector could create more than $300 billion in value every year by effectively leveraging big data, according to a McKinsey study (Manyika et al., 2011). Even more importantly, patients will receive improved treatment as a result of the research possibilities afforded by big data.

While still in the nascent stages of implementation, big data analytics have already been used to considerable success in healthcare.

One early application of big data can be found in pharmacovigilance (PV), or the assessment and monitoring of drug safety. Partners Healthcare, the largest healthcare provider in Massachusetts, utilized its collection of health records to conduct a pilot program on EHR use in post-market surveillance of drugs. When researcher John Brownstein did a system-wide analysis of EHRs, he discovered a startling trend: the baseline expected level of heart attack admissions to two hospitals jumped 18 percent beginning in 2001 and returned to normal levels in 2004—coinciding with the advent and discontinuation of the prescription pain-reliever Vioxx ("Partners healthcare," 2009).

Currently, Partners Healthcare, as a member of eHealth Initiative and NIH-funded partnerships, is working to build upon this model using recent advancements in health IT ("Pharmacovigilance"). The growing prevalence of EHRs, in conjunction with the ability to de-identify clinical data, may offer a larger-scale, revolutionary approach to post-market drug surveillance.

Analytic consumption of EHRs can also be used to directly improve patient treatment. At New York Presbyterian Hospital, computers are being trained to analyze the risk factors of new patients that are sometimes overlooked by humans. This initiative has already found some initial success: through analyzing patient records with Microsoft technology, New York Presbyterian has reduced the incidence of potentially fatal blood clots by approximately 30 percent (Robertson, 2012).

Furthermore, big data is increasingly able to inform the decisions of not only care providers, but other industry stakeholders as well. For example, Aetna Healthcare, a national diversified healthcare benefits provider, created an integrated health database that it uses to improve evidence-based medicine practices in the most cost-effective manner possible. Aetna also uses big data in its recently synchronized iTriage and CarePass application to give consumers access to a bevy of health resources on a centralized platform ("The power of," 2012). This allows
previously siloed information, whether sourced from providers, payers, or the patients themselves, to be aggregated for better care.

While big data holds significant value for the entirety of healthcare, value-based models of care have perhaps the most to gain from big data. Take, for example, the ACO model, which is hinged upon quality metrics and reduction in total cost of care. Big data can facilitate both of these ends through health trend analysis and comprehensive patient management, making care both more personal and more efficient. This further encourages quality of care over sheer quantity. In fact, ACOs may soon even rely on big data to properly capture and incentivize quality measures.

“What we’re going to need is data on how my patients are doing and how cost-effectively I’ve been able to get that,” according to Dr. Allan Goroll, Harvard Medical School professor and Massachusetts General Hospital primary care physician. "The analytics are going to change. Instead of productivity -- which is really just value -- it’s going to be in terms of how much value was created" (Fluckinger, 2011).

Population health management is another model that could see substantial benefits from big data analytics. By collecting and analyzing data for at-risk groups, we can identify unique needs and interventions for patient subpopulations through predictive modeling—something simply not possible without large-scale informatics.

However, big data only has value if implemented effectively: several issues must be addressed for big data to have a truly transformative effect on the industry.

Of perhaps most fundamental importance is the issue of data quality. What information is valuable, and what can be neglected? What algorithms and IT structures are needed to harness and process the data into a usable form? Unfortunately, there may be no universal answer to this dilemma—objectives vary by stakeholder, and it is up to the individual organization to find and address its data needs.

But even if patient information can successfully be collected and de-identified, this vast expanse of data is meaningless without the talent to interpret it. According to a McKinsey study, there will be a scarcity of qualified workers needed for organizations to realize the advantages of big data. By 2018, the United States alone could face a shortage of 140,000 to 190,000 personnel with the requisite analytical skills to manage big data, as well as 1.5 million managers and analysts equipped to use big data in effective decision-making (Manyika et al., 2011). This highlights the lack of, as well as demand for, data-science education and training needed to foster successful big data use. At the moment, though, companies must largely be self-taught. "Bits of [formal data-science education] do exist in various departments around the country, and also in businesses, but as an integrated discipline it is only just starting to emerge," according to Nigel Shadbolt, professor of A.I. at the University of Southampton and co-director of the Open Data Institute (Rooney, 2012).

Beyond the organizational implementation hurdles, there is an additional ethical component to consider as well: patient medical record privacy. As vast quantities of health data are used for research, previously siloed PHI will be
integrated and disclosed to multiple third parties. Though this may sound concerning, privacy will still be of utmost importance—the large majority of, if not all, PHI will be de-identified as it enters large databases for research consumption (Litan, 2012). Yet, does this still pose a threat to patient confidentiality or security? Naturally, this is a divisive issue amongst the healthcare community. The issue of patient consent is one that must be worked out by all stakeholders to the point of mutual satisfaction. For example, what kind of information disclosure requires consent? How should consent be obtained? Such questions are very important and will have to be addressed sooner rather than later. The reality of the matter is, however, that health data is a public good: the more data available to researchers, the higher the likelihood they will be able to detect the patterns necessary to find new causes, treatments, and cures for diseases (Ibid). This could be a quantum leap of information in healthcare. Perhaps forking over our medical records is not such a high price to pay after all—lives could be at stake.

Conclusion

While hurdles still remain, early successes, as well as government interest, show the promise of big data in healthcare. We are already beginning to see that big data will be a critical piece in healthcare’s transition to value-based care. Yet, the value of big data is often relatively unknown. As industry stakeholders become more familiar with the substantial benefits of big data, we should expect an exponential increase in its deployment in upcoming years.

However, healthcare is already years behind other industries in harnessing and capitalizing upon the data it possesses. Given the stakes at hand, this is simply unacceptable. As the industry progresses toward a model predicated on quality, we must increasingly consider the immense value that big data can have for all healthcare stakeholders.
Resources


