Institute for Health Technology Transformation

Transforming Health Care Through Big Data

Strategies for leveraging big data in the health care industry

iHT²
Institute for Health
Technology Transformation

Acknowledgements

Mike Cottle
Chief Information Officer
Newton Medical Center

Waco Hoover, MBA
Chief Executive Officer
Institute for Health Technology Transformation

Shadaab Kanwal
Executive Director Research & Quality
Kaiser Permanente

Marty Kohn, MS, MD, FACEP, FACPE
Chief Medical Scientist, Care Delivery Systems
IBM Research

Trevor Strome, MSc, PMP
Informatics / Process Improvement Lead, Emergency Program
Winnipeg Regional Health Authority
Assistant Professor, Department of Emergency Medicine, Faculty of Medicine
University of Manitoba

Neil W. Treister, MD, FACC
Medical Informatics Officer
Sharp Community Medical Group

Dear Colleagues:

At the heart of many health care industry debates is what to do about data: how to realize its value for quality care, how to use it to bend the cost curve, how to share it, and how to secure it. Health care providers face significant obstacles in implementing analytics, business intelligence (BI) tools and data warehousing. Health data is diverse, comprising structured and unstructured information in a range of formats and distributed in hard-to-penetrate silos owned by a multitude of stakeholders. To complicate matters, each stakeholder has different interests and business incentives while still being closely intertwined.

While other industries are already leveraging their data assets to improve efficiencies and make more informed decisions, the health care sector lags far behind, much to its detriment. A 2011 McKinsey report estimated that the health care industry can potentially realize $300 billion in annual value by leveraging big data.

To do so, however, the health care industry must identify and establish proven strategies and best practices to manage big data and to conduct the advanced analysis necessary to generate real insights that can benefit the health system. Health care data is rarely standardized, often fragmented, or generated in legacy IT systems with incompatible formats.

The purpose of this paper, Transforming Health Care Through Big Data, is to help executives from hospitals, health systems and other provider organizations to identify and understand models for innovative uses of data assets that can enable them to reduce costs, improve quality, and provide more accessible care. Areas of focus include the storage, processing, analysis, and management of data, while addressing the unique challenges presented by the health care industry, not the least of which is the goal to use data to make better, evidence-based decisions.

Transforming Health Care Through Big Data comprises the contributions of individuals from provider, health system, health information technology, academic, and health policy domains. This group is well-versed in data analysis, patient-centered care, health information technology, decision support systems, and the emerging—and urgent---imperative to transform health care delivery with innovative uses of health data. The leaders who participated in this project include:

- Mike Cottle, Chief Information Officer, Newton Medical Center,
- Shadaab Kanwal, Executive Director Research & Quality, Kaiser Permanente
- Marty Kohn, MS, MD, FACEP, FACPE, Chief Medical Scientist, Care Delivery Systems, IBM Research
- Trevor Strome, MSc, PMP, Informatics/Process Improvement Lead, Emergency Program, Winnipeg Regional Health Authority; Assistant Professor, Department of Emergency Medicine, Faculty of Medicine, University of Manitoba.
- Neil W. Treister, MD, FACC, Medical Informatics Officer, Sharp Community Medical Group

Respectfully,

Waco Hoover, MBA
Chief Executive Officer
Institute for Health Technology Transformation

Table of Contents

Introduction

In 2011 alone, 1.8 zettabytes of data were created globally. To put this into perspective, this volume of data equates to 200 billion, 2-hour-long HD movies, which one person would need 47 million years to watch in their entirety. What's more, this volume of data is expected to double this and every year going forward.[1]

Consider also that:

- Wal-Mart's data warehouses now include some 2.5 petabytes of information, the equivalent of roughly half of all the letters delivered by the U.S. Postal Service in 2010.[2] That volume of data is more than 160 times larger than the holdings of the U.S. Library of Congress.

- U.S. health care data alone reached 150 exabytes in 2011. Five exabytes (10^{18} gigabytes) of data would contain all the words ever spoken by human beings on earth. At this rate, big data for U.S. health care will soon reach zettabyte (10^{21} gigabytes) scale and even yottabytes (10^{24} gigabytes) not long after.[3]

- Kaiser Permanente, the California-based health network which has more than 9 million members, is estimated to have between 26.5 petabytes and 44 petabytes of patient data under management just from electronic health record (EHR) data, including images and annotations. This amounts to the same amount of information contained in 4,400 Libraries of Congress.[4]

Big data is big. And it's getting even bigger in health care. One can hardly pick up any of today's health care publications without coming across a reference to "Big Data" and its growing impact on the industry. The Cleveland Clinic labeled "harnessing big data to improve health care" as one of its Top 10 Medical Innovations of 2012.[5] The Harvard Business Review also recently identified "data scientist" as the "Sexiest Job of the 21st Century."[6] It's true that big data promises to ease the transition to authentic data-driven health care, allowing health care professionals to improve the standard of care based on millions of cases, to define needs for subpopulations, to make more personalized decisions for individual patients, and to identify and intervene for population groups at risk for poor outcomes. But while big data has transformed much of American industry, it's also true that massive information sharing and analysis has yet to generate significant benefits within health care.[7]

The fact is that at the heart of many health care industry debates is: how to realize its value for quality care, how to use it to bend the cost curve, how to share it, and how to secure it.

The need for this paper is critical and urgent. Under the Patient Protection and Affordable Care Act, data volume is expected to grow dramatically in the years ahead. In addition, health care reimbursement models are changing; Meaningful Use and pay for performance are emerging as critical new factors in today's health care environment. It is vitally important for health care organizations to get the tools, infrastructure, and techniques in place now to leverage big data effectively, or risk losing potentially millions of dollars in revenue and profits. In fact, they risk disappearing completely as quality and efficiency become the primary determinants of success.

Big Data

What Exactly Is Big Data?

A report delivered to the U.S. Congress in August 2012 defines big data as "a term that describes large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information."[8]

Big data goes beyond size and volume to encompass such characteristics as variety, velocity, and, with respect specifically to health care, veracity. Big data can be said to comprise five different categories, or streams, of information:

1. Web and social media data: Clickstream and interaction data from social media such as Facebook, Twitter, Linkedin, and blogs. It can also include health plan websites, smartphone apps, etc.
2. Machine-to-machine data: Readings from sensors, meters, and other devices.
3. Big transaction data: Health care claims and other billing records increasingly available in semi-structured and unstructured formats.
4. Biometric data: Fingerprints, genetics, handwriting, retinal scans, and similar types of data. This would also include X-rays and other medical images, blood pressure, pulse and pulse-oximetry readings, and other similar types of data.
5. Human-generated data: Unstructured and semi-structured data such as electronic medical records (EMRs), physicians' notes, email, and paper documents.[9]

In recent years, it has become increasingly apparent that multiple streams of data like these can be leveraged with powerful new collection, aggregation, and analytics technologies and techniques to improve the delivery of health care at the level of individual patients as well as at the levels of disease- and condition-specific populations.

Big data goes beyond size and volume to encompass such characteristics as variety, velocity, and, with respect specifically to health care, veracity.

Big Data

Benefits to Health Care

By digitizing, combining and effectively using big data, health care organizations ranging from single-physician offices to multi-provider groups, large hospital networks, and accountable care organizations stand to realize significant benefits. These can include improving the quality and efficiency of health care delivery; detecting diseases at earlier stages, when they can be treated most successfully: managing specific health populations and individuals; and detecting health care fraud more quickly and efficiently.

Electronic health records, coupled with new analytics tools, open the door to mining information for the most effective outcomes across large populations.

Health Care Quality and Efficiency

As of 2010, national health expenditures represented 17.9% of gross domestic product, up from 13.8% in 2000. Concurrently, the prevalence of chronic diseases like diabetes is growing and consuming a greater percentage of health care resources. Electronic health records (EHRs), coupled with new analytics tools, open the door to mining information for the most effective outcomes across large populations. Using de-identified information, researchers can look for statistically valid trends and provide assessments based upon true quality of care.[10]

Earlier Disease Detection

Electronic sensors are increasingly being employed to monitor key biochemical markers, with real-time analysis taking place as the data streams from individual patients to HIPAA-compliant analysis systems. Analytics like these can alert specific individuals and their providers to potentially adverse events, such as side effects to medications, the early development of infection, and allergic reactions.[11]

Fraud Detection

Big data is widely expected to fundamentally transform medical claims payment systems, resulting in reduced submissions of improper, erroneous or fraudulent claims. For example, a significant challenge confronting the Centers for Medicare and Medicaid Services (CMS) is managing improper payments under the Medicare Fee-For-Service Program (FFS). Until recently, claims were manually reviewed against medical documentation submitted by providers to verify compliance with Medicare policies. Using powerful, new big data tools, techniques and governance processes, CMS is dramatically improving its fraud-detection efforts.[12] In fiscal year 2011, for the second year in a row, CMS anti-fraud activities resulted in more than $4 billion in recoveries, an all-time high, owing in large part to big data-based detection and analytics tools.[13]

Big data also can be used to improve such endeavors as population health management, the identification and measurement of more accurate quality metrics, the management of capitated populations, and treatment protocols for a wide range of chronic conditions.

Big data also can be used to improve such endeavors as population health management, the identification and measurement of more accurate quality metrics, the management of capitated populations, and treatment protocols for a wide range of chronic conditions such as diabetes and congestive heart failure (CHF). Through the use of data mining techniques, big data can be used to identify patients and populations at risk for various conditions and diseases, identify adverse drug events, improve selection of candidates for patient-centered interventions and identify costly procedures, waste and delays. All told, the health care industry can potentially realize as much as $300 billion in annual value by effectively leveraging big data.[14]

Big Data

Early Successes

> All told, the health care industry can potentially realize as much as $300 billion in annual value by effectively leveraging big data.[14]

Some examples of early success capturing value from big data in health care:

- The University of Ontario's Institute of Technology partnered with a prominent technology firm—IBM-- to develop Project Artemis, a highly flexible platform that leverages streaming analytics to monitor newborns in the neonatal intensive care unit of a hospital. Using these technologies, the hospital was able to predict the onset of nosocomial infections 24 hours before symptoms appeared. The hospital also tagged all time-series data that had been modified by software algorithms. In case of a lawsuit or medical inquiry, the hospital felt that it had to produce both the original and modified readings. Plus, the hospital established policies around safeguarding protected health information.[15]
- A recent New Yorker magazine article, by Atul Gawande, MD, described how orthopedic surgeons at Brigham and Women's Hospital in Boston, relying on their own experience combined with data gleaned from research on a host of factors critical to the success of joint-replacement surgery, systematically standardized knee joint-replacement surgery, with a resultant increase in more successful outcomes and reduced costs. Similarly, the University of Michigan Health System standardized the administration of blood transfusions, reducing the need for transfusions by 31 per cent and expenses by $200,000 a month.[16]
- The Department of Veterans Affairs (VA) in the United States has successfully demonstrated several health care information technology (HIT) and remote patient monitoring programs. The VA health system generally outperforms the private sector in following recommended processes for patient care, adhering to clinical guidelines, and achieving greater rates of evidence-based drug therapy. These achievements are largely possible because of the VA's performance-based accountability framework and disease-management practices enabled by electronic medical records (EMR) and HIT.[17]
- The California-based integrated managed-care consortium Kaiser Permanente connected clinical and cost data to provide a crucial dataset that led to the discovery of adverse drug effects and subsequent .withdrawal of the drug Vioxx from the market.[18]
- Researchers at the Johns Hopkins School of Medicine found that they could use data from Google Flu Trends (a free, publicly available aggregator of relevant search terms) to predict surges in flu-related emergency room visits a week before warnings came from the Centers for Disease Control and Prevention. Similarly, Twitter updates were as accurate as official reports at tracking the spread of cholera in Haiti after the January 2010 earthquake; they were also two weeks earlier.[19]
- Researchers at IBM have devised a prototype program that predicts the likely outcomes of diabetes patients, based on the patients' longitudinal health data and association with particular physicians, management protocols, and relationship to population health management averages.[20]

Big Data

- Physicians at Harvard Medical School and Harvard Pilgrim Health Care recently demonstrated the potential for computer algorithms to analyze EHR data to detect and categorize patients with diabetes for public health surveillance. The researchers used algorithms to review 4 years of EHR data from HPHC, a large, multisite, multispecialty ambulatory practice, and flag patients suspected of having diabetes based on lab results, diagnosis codes, and prescriptions. The algorithms successfully identified patients who had been clinically diagnosed with diabetes, and were able to distinguish between patients with Type I and Type II diabetes. The algorithms were more effective at identifying patients with diabetes than reviewing diagnosis codes alone.[21]

The algorithms successfully identified patients who had been clinically diagnosed with diabetes, and were able to distinguish between patients with Type I and Type II diabetes. The algorithms were more effective at identifying patients with diabetes than reviewing diagnosis codes alone.

Big Data

Future Applications: Individual and Population Health Management

The examples cited above represent encouraging signs of big data's nascent but growing impact on health care. Going forward, big data technologies and techniques are expected to drive decision making at the individual patient, group, and population levels.

Individual patient care

One of the goals of health care transformation is to garner the ability to personalize care for an individual patient. Big data is poised to make a significant contribution toward that goal in the coming years. For example, each diabetic patient has a combination of characteristics that can identify her within the larger population. Using inferences from data about patients that are very similar to that patient, as well as published data and genetic information, big data can help make it possible for providers to offer interventions that are more likely to be beneficial to this particular patient.

Current clinical decision support (CDS) systems already include computerized physician order-entry capabilities that analyze entries and compare them against medical guidelines to alert for such potential errors as adverse drug reactions. By deploying these systems, providers can potentially reduce adverse reactions and lower treatment error rates and liability claims, especially those arising from clinical mistakes. While such obstacles as "alert fatigue" that can negate the benefits of drug utilization reviews still need to be overcome, the signs are promising. In one study conducted at a pediatric critical care unit in a major U.S. metropolitan area, a CDS system tool cut adverse drug reactions and events by 40 percent in just two months.[22]

Future big data-powered CDS systems will become substantially more intelligent by including modules that use image analysis and recognition in databases of medical images (X-ray, CT, MRI) for prediagnosis, or modules that automatically mine medical literature to create a medical expertise database capable of suggesting treatment options to physicians based on patients' medical records. In addition, CDS systems of the future will enable a larger portion of work to flow to nurse practitioners and physician assistants by automating and facilitating the physician advisory role and thereby improving the efficiency of patient care.[23]

Population health management

Many experts today believe that big data promises to help health care organizations to combine and analyze data from EHRs, claims records, and other information sources to improve clinical and financial outcomes, not just on the individual patient level, but across condition- and disease-specific patient populations. To cite one current example, the United Kingdom's National Health Service (NHS) has announced plans to sequence the entire genomes of up to 100,000 patients over three to five years to build a national research database. The database will help clinicians and researchers better understand the genetic causes of cancer and other rare conditions. Based on this data, it is believed that new drugs, treatments, and therapies can be developed and patients can receive targeted therapies that may prove more effective. Genetic sequencing will be voluntary, and the information in the database will be kept anonymous.[24]

Yet realizing the promise of population health management requires new IT tools and a willingness on the part of providers to embrace them. A recent report by the Institute for Health Technology Transformation cited the need for improved electronic health records,

Ideas in Practice

Implement workflow-based user interfaces early on to collect data. Make sure you emphasize data warehousing and ad hoc reporting. Get an enterprise-wide commitment to being a culture of information; allow key users and managers to do ad hoc reporting. Identify certain reporting for a production mode, where it is created and disseminated and trended regularly. Master dashboarding. Make clinical decision support relevant and helpful. Before you buy any data management tool, make sure you talk to like-minded organizations who are using it in the live environment.

Neil W. Treister, MD, FACC
Medical Informatics Officer
Sharp Community Medical Group

Big Data

telehealth platforms, electronic registries, data management software, and more powerful analytics systems.[25]

Electronic health records and automation tools already exist to identify and stratify individual patients who need special attention or care; identify care gaps; measure outcomes; and encourage patients to assume more responsibility for their health. However, they cannot store, manage, and distribute comprehensive, timely, and relevant information to the degree needed for public health management (PHM). Electronic health records, for example, often don't contain data about the care that patients have received outside an organization, and most aren't designed for interoperability. Likewise, many EHRs don't generate real-time alerts for preventive and chronic care, and don't generate quality and population reporting.[26,27]

In addtion, many clinical analytics tools currently in use are quite primitive, reporting just a few basic facts and figures about a patient panel. The next generation of BI tools will have to be predictive and prescriptive to make PHM a reality.[28]

Clinicians are more likely to uses guidelines if the clinicians are part of the creation process and the guidelines support worklfow.

Additionally, many of the interventions needed to improve the health of a large population fall more into the realm of education and public safety than they do into medical practice. Getting diabetics to exercise more or eat a heart-healthy diet, for instance, isn't what doctors do best; those tasks are traditionally handled by nurse educators and public health agencies, often with the help of public service campaigns or classroom instruction.[29]

Some argue that until IT systems are significantly upgraded and clinician issues and concerns are addressed, the promise of big data's impact on PHM—perhaps even PHM itself---will not be fully realized. Writing in the January 2013 issue of Health Affairs, Arthur L. Kellermann and Spencer S. Jones maintain that the potential of health IT to transform health care can and will only be fully realized by a comprehensive, three-pronged effort by government, private vendors, and providers.

"Ultimately, there is only so much that the government and vendors can do. Providers must do their part by reengineering existing processes of care to take full advantage of the efficiencies offered by health IT," the authors write.[30]

We have begun formalizing our physician documentation efforts by building more "templates" which result in more discrete data. We are also evaluating tools which claim to be able to "read" free text and extract relevant data and place it into discrete fields.

Mike Cottle
Chief Information Officer
Newton Medical Center

Big Data

Future Applications: Reducing Variations in Care

Many studies have shown that wide variations exist in health care practices, outcomes, and costs across different providers, geographies, and patients.[31] For example, researchers for the Dartmouth Atlas project have documented that some primary care physicians order more than twice as many CT scans as their colleagues in the same practice. They've also determined that the rates of coronary stents are three times higher in Elyria, Ohio, than they are in nearby Cleveland, home of the Cleveland Clinic. The questions in both cases: Why, and what do those differences mean for patients?[32]

Big data promises to help health care providers to answer questions like these on both the macro and micro levels by streamlining outcomes-based research, also known as comparative effectiveness research (CER). Critically analyzing large datasets that include patient characteristics and the cost and outcomes of treatments can help to identify the most clinically and cost-effective treatments to apply.[33] Indeed, CER offers the potential to reduce incidences of overtreatment and undertreatment, both of which result in worse patient outcomes and higher health care costs in the long run. However, information technology is an enabler, not the solution itself. The potential for benefits is predicated on the assumption that the organization/providers are committed to evidence-supported decisions using analytic tools with available information. If that commitment has not been made, analytic tools provide little value. Then again, the tools themselves need to be transparent and evidence-based, have face validity, be embedded in work flows, and allow for variations based on individual findings when important.

Indeed, for CER to achieve systemwide scale, comprehensive and consistent clinical and claims data sets need to be captured, integrated, and made available to researchers. Stakeholders must address issues such as the potential lack of standards and interoperability for EHR systems that could make it difficult to combine data sets.[34]

Another concern is how to ensure patient privacy while still providing sufficiently detailed data to allow effective analyses. Disseminating knowledge about the most effective treatments to medical professionals will require the introduction or upgrading of tools, including clinical decision support systems, so that physicians can receive recommendations of best practices at the point of care.

Ultimately the goal is for health care organizations to be able to gather and analyze as much useful data as possible to generate better health care in a system that currently is driven primarily by regulatory issues. Key to achieving that goal is knowing specifically what metrics are necessary to measure progress. Without clearly defined, concrete metrics in place for analytics, organizations are likely to generate data that is not useful.

> We are currently evaluating a method to encrypt data at rest, not just in transit. We are constantly evaluating our relationships with vendors with which we share data, making sure we have adequate BAAs in place.
>
> Mike Cottle
> Chief Information Officer
> Newton Medical Center

> Big data promises to help health care providers to answer questions on both the macro and micro levels by streamlining outcomes-based research, also known as comparative effectiveness research.

Big Data Challenges

Significant challenges still lie in the way of health care's ability to leverage big data effectively. For one, the industry remains largely unprepared to deal with the flood of data being generated. In addition, health care data is rarely standardized, often fragmented, or generated in legacy IT systems with incompatible formats. Health care data is also diverse and distributed in hard-to-penetrate silos owned by a multitude of stakeholders.

As a result, health care providers face significant challenges in implementing analytics, business intelligence (BI) tools, and data warehousing, as well as an overarching general reluctance among organizations to share their data.

Industry Readiness

While other U.S. industries, such as the banking and consumer retail sectors, are well along in leveraging big data techniques and technologies, health care remains largely unprepared to handle the deluge of data.

Health care organizations are accumulating 85 percent more data than they did two years ago, but 77 percent of health care executives give their organizations a "C" or below for managing their data, according to an Oracle poll of 333 U.S. and Canadian C-level executives from health care and 10 other industries. Of health care executives interviewed, none gave their organizations an "A" for data "preparedness."[35]

Moreover, despite the high priority they place on implementing EHRs, health care leaders report their organizations are struggling to leverage them: while 34 percent reported being able to capture data from EHRs to help patients, 43 percent said they were unable to collect sufficient data to improve care.[36]

Some experts point to the industry's "cultural reliance on paper" as a major reason for health care's reluctance to embrace big data. Although EHR use has grown from about 20 percent of providers to some 60 percent in 2012, U.S. health care remains predominantly a paper-based system. It will take significant effort to shift attitudes and educate providers about available and emerging technologies.[37] This may be evidence that the clinical front end will be critical to driving the requisite collection of data that becomes "big data".

Data Usability/Trustworthiness

Data usability and trustworthiness have been identified as major issues, especially with respect to clinical decision support. Most clinical data is stored in "unstructured" form, especially within EHRs, making it difficult to access for effective analytics. For example, while individual physicians can read narrative text within an EHR, most current analytics applications cannot effectively utilize this unstructured data.

Currently, most clinical analytics rely on claims or administrative data. This data consists largely of more structured data, but is of limited value in evaluating the efficacy of care and treatment outcomes. Indeed, "much of what is currently documented and contained in the health record responds not to clinical needs but, instead, to diverse and increasingly complex and prescriptive medicolegal, reimbursement, accreditation, and regulatory requirements."[38]

Emerging big data technology and techniques show promise in helping organizations to process and evaluate data from EHRs, clinic equipment, telehealth devices, and home health monitors.

Ideas in Practice

We are just starting to capture information from a number of clinical databases. The ambulatory clinical information is not yet effectively integrated with the in-patient. Much of our reporting must be done manually. The hospital-driven IS department is still grappling with how to integrate and use information in more robust ways. Only recently have informaticists and clinicians become recognized as important customers with ever-changing and expanding needs for timely access to clinically relevant information.

But, we are starting to see some progress. We are implementing MedVentive and DBMotion imminently. Although nearly all the data is currently siloed in financial and/or demographic systems or collected in stand-alone ambulatory and in-patient EHR databases, we are working on ways to bring that data together for our ACO initiatives. We are starting to seek more experience with risk adjustment and predictive modeling. The culture is moving toward one that values and nurtures innovation in data mining.

The medical groups focus on bed days, length of stay, and other measures of avoidable hospitalization. We also look at resource allocation by provider, although it doesn't aggregate or link to diagnosis well and we do not yet have standardized dashboards to present the data to clinicians efficiently. We believe our data will ultimately help us understand which providers manage which types of patients better and why.

Neil W. Treister, MD, FACC
Medical Informatics Officer
Sharp Community Medical Group

Big Data Challenges

Data Fragmentation

The separation, or fragmentation, of data among labs, hospital systems, financial IT systems and EHRs, is another significant obstacle to leveraging big data in health care. Each entity serves as a single repository, or silo, for information whose purpose is to provide clinical care, scheduling or billing information, or operational information. This continues to be problematic for organizations seeking to get individual systems to communicate with each other easily. It remains especially challenging in smaller organizations with multiple systems and taxonomies that make extracting useful information difficult for data mapping.

Key to overcoming these fragmentation obstacles is "normalizing" the data and shifting to a "culture of best practices…best experiences…and using data from various components of health IT to improve care and lower costs in a holistic way."

Ideas in Practice

At the Rhode Island Quality Institute (RIQI), we leveraged the Rhode Island Beacon Community Program to build infrastructure that aggregates individual patient data into a broader, meaningful view of health and health care in our state. Providers have recorded data on hundreds of thousands of patients during the course of routine medical care. Assessing and treating a woman's LDL (cholesterol) level or determining whether a man has been screened for depression is important to that individual. When these unique pieces of data are viewed across a community, an entire system of care can be evaluated and improved, which is important to every individual in the state. By creating an interoperable analytics platform, we now take data that originates in different EHRs across different provider practices and view this data as a whole.

Through the Beacon Program, RIQI and RI providers used the aggregated quality data and analytics to demonstrate that a given medical practice can be evaluated and compared to peers and to targets. More compelling is our demonstration that comparative data can be used to improve the care provided by that practice. Providers recognized patterns of less effective treatment for patients with diabetes and sought advice from peer practices that were performing better. This exchange of learning led to improved care by the providers and improved health of their patients.

A key lesson learned was the importance of high quality data gathering and data entry prior to data aggregation. During early efforts to gather quality data on practices, we observed wide variations in how providers documented clinical information. Whether or not a patient smoked cigarettes, how a blood pressure value was recorded, or when a new medication was started appeared differently in different EHRs, or even in the same EHR across different practices. Comparisons across practices or aggregation of values could not be made until data normalization processes were developed. Leveraging these same data aggregation, normalization, and analytic programs, RIQI is developing programs to identify disease patterns across the state, so that effective health care can be targeted at the individuals that need attention most and through new models of care.

Jonathan Leviss, MD
CMO, Rhode Island Quality Institute
Physician, Thundermist Health Center, RI

Big Data Challenges

Architectural and Infrastructure Issues

How to handle existing legacy systems and their compatibility with new technologies and techniques remain a significant obstacle to many organizations' efforts to leverage big data. In many cases, organizations don't have easy options to upgrade or otherwise adapt their technologies to growing data demands. Data is stored in SQL databases, and the organization's vendor has only certain mechanisms in place to exchange data, limiting the health care organization's ability to roll out new tools.

Health care organizations are collecting and storing so much data that improved data governance measures are urgently needed to identify, enter and leverage the most useful data. Some is being offloaded to patients with portals and kiosks. Organizations are struggling with such questions as how best to determine the value of their data, how to store their data, and how and when to delete and/or archive their data.

Health Information Ownership, Use, and Security

In the discussion over how to approach big data in health care, important questions are also being raised over who owns certain forms of health information, how and by whom that information can be used, and for what purposes. Douglas Patterson, co-author with Kord Davis of the recently released book, "Ethics of Big Data," put the spotlight on these thorny questions in a recent interview:

> "[Let's say] you have an accident and you're taken to the hospital unconscious for treatment. Lots of data is generated in the process, and let's suppose it's useful data for developing more effective treatments. Is it obvious that that's your data? It was generated during your treatment, but also with equipment the hospital provided, based on know-how developed over decades in various businesses, universities, and government-linked institutions, all in the course of saving your life. In addition to generating profits, that same data may help save lives down the road. Creating the data was, so to speak, a mutual effort, so it's not obvious that it's your data. But it's also not obvious that the hospital can just do whatever it wants with it. Maybe under the right circumstances, the data could be de-anonymized to reveal what sort of embarrassing thing you were doing when you got hurt, with great damage to your reputation. And giving or selling data down the line to aggregators and businesses that will attempt to profit from it is one thing the hospital might want to do with the data that you might want to prevent — especially if you don't get a percentage. Questions of ownership, questions about who gets to say what may and may not be done with data, are where the real and difficult issues arise."[39]

Davis, in the same interview, touched on the myriad opportunities big data presents for people's personal and financial data to be siphoned off for both beneficent and nefarious purposes:

> "The field of predictive analytics has been around for a long time, but the development of big data technologies has increased accessibility to large datasets and the ability to data mine and correlate data using commodity hardware and software. The potential benefits are massive. A promising example is that longitudinal studies in education can gather and process significantly more minute data characteristics and we have no idea what we might learn. Which is precisely the point. Being able to assess a more refined population of cohorts may well turn out to unlock powerful ways to improve education.

> We are experimenting with ways to leverage proprietary and open-source analytics tools to distill the unstructured data in EHRs into meaningful structures that can then be more quantitatively analyzed. We are starting off small, with a few pilot projects to gain experience in working with this type of data (eg, trying to gain meaningful information off our triage document comments section), but the goal would be to scale up these capabilities to other unstructured data when this can be made reliable and cost-effective on a larger scale.
>
> Trevor Strome, MSc, PMP
> Informatics/Process Improvement Lead
> Emergency Program
> Winnipeg Regional Health Authority

Big Data Challenges

Similar conditions exist for health care, agriculture, and even being able to predict weather more reliably and reducing damage from catastrophic natural weather events.

"On the other hand, the availability of larger datasets and the ability to process and query against them makes it very tempting for organizations to share and cross-correlate to gain deeper insights….Even a simple, singular transaction, such as buying a pair of shoes online touches your bank, the merchant card processor, the retail or wholesale vendor, the shoe manufacturer, the shipping company, your Internet service provider, the company that runs or manages the ecommerce engine that makes it possible, and every technology infrastructure organization that supports them. That's a lot of opportunity for any single bit of your transaction to be stored, shared, or otherwise misused."[40]

It is for reasons like these that current policies may need to be adjusted and/or rewritten to optimize the flow of data across the health care system while protecting the privacy and security of patients and consumers.[41] Ultimately, health care organizations must focus on diligently protecting and securing four types of data in particular:

- Personally identifiable information. The loss of personally identifiable information such as dates of birth, driver's license numbers, and security numbers is among the greatest of privacy threats. While external threats dominate top of mind discussions, information breaches are growing, presenting the potential for significant loss of customers, incurrence of high compensation claims, lawsuits and permanent damage to reputation.
- Clinical data. Electronic health records contain a wide range of patient-specific information, including prescription data, treatment details, and other data. Combined with a policy number, a hacker can use it to receive unauthorized medical care or bill for services never received. The leakage and/or corruption of such information can even result in irrevocable harm to one's personal and professional life.
- Financial data. With banks and individuals getting more proactive about protecting their financial information, the medical industry is becoming an easy target for hackers. The outsourcing of billing activities and increased internet and mobile involvement in health care create more avenues for potential data theft; the resulting legal consequences and loss of patient trust can taint an organization's brand for life.
- Behavioral data. Behavioral data is the newest and possibly fastest-growing in health care, thanks to monitoring devices, GPS tracking, internet site visits, social media, purchasing habits, exercise activity, and self-reporting. Behavioral data is increasingly becoming the 'hot favorite' for cyber thieves as it helps to draw up startlingly accurate representations of human behavior which are of great demand among marketing companies and also others with illicit intentions. With growing usage of tablets, smartphones and other mobile devices, this data is becoming more vulnerable to theft.[42]

IBM recognizes the importance of improving data usability and the wealth of information that is included in free text entries. The ability to use all the data available, both structured and unstructured, is critical. IBM has developed tools, such as "IBM Patient Care and Insight" which, among other things, uses natural language processing to convert unstructured data to structured data so that it is amenable to analysis.

Marty Kohn, MD
Chief Medical Scientist
IBM Research

Strategies to Leverage Big Data

Given the early successes with respect to big data applications, the immense potential for big data applications in the future, and the remaining challenges discussed in this document, how should individual health care providers and organizations move forward to leverage big data most efficiently and effectively?

The following items provide some direction for health care organizations venturing into the realm of big data:

- Implement a data governance framework. A carefully structured framework for enterprise-wide data governance is arguably the first and most critical priority to ensure the success of any effort to leverage big data for health care delivery. The Data Governance Institute, a provider of in-depth, vendor-neutral information relating to tools, techniques, models, and best practices for the governance of data and information, defines such a framework as a "logical structure for classifying, organizing, and communicating complex activities involved in making decisions about and taking action on enterprise data."[43] According to the DGI, all organizations need to be able to make decisions about how to manage and realize value from data; how to minimize cost and complexity; how to manage risk; and how to ensure compliance with ever-growing legal, regulatory, and other requirements.[44] The DGI also describes 10 components that every formal data governance program should include (see Appendix A).[45]
- Engage providers. Engaging providers is critical to changing the culture of resistance to new approaches to data collection and analysis. Health care organizations are highlighting the importance of big data initiatives by rolling them out at department-wide meetings and rewarding their physicians when they meet standards for data collection and improvement of quality metrics.
- Foster competition and transparency. Similarly, health care organizations are attaching monetary incentives to measuring and looking at data; displaying peer and colleague data with respect to patient satisfaction and quality metrics; and using dashboards, all in an effort to leverage competition and improve performance among clinicians. Such tools, for example, are being used to measure patient flow at Cancer Care of Ontario to improve efficiency of the emergency department.
- Bake analytics into training. More institutions are recognizing that physicians and nurses both need training in analytics to understand how big data tools add value to overall health care performance. Even medical schools, like those at the University of North Carolina at Chapel Hill and the University of Washington-Seattle, are revising their curricula to encourage critical thinking and the use of information.
- Provide for flexibility in information transference. There is a growing recognition that work and learning styles vary among clinicians; facilities are demonstrating a growing willingness to deliver data in multiple ways based on clinician preference and style.
- When possible, choose in-house solutions over vendor-generated solutions. At times the inflexibility of some vendor-generated solutions can be a major obstacle to leveraging big data technology in a given organization. Organizations are increasingly recognizing that some of the most successful solutions to their challenges can sometimes be developed with "in-house" input and expertise. In most cases,

Ideas in Practice

We use EHRs to collect certain amounts of discrete data. We are looking at language processing capabilities from vendors to get discrete data out of narrative text.

Neil W. Treister, MD, FACC
Medical Informatics Officer
Sharp Community Medical Group

Keep providers engaged! Let them be a part of the decision-making process. Solicit their feedback. Today, electronic records do not offer them much in the way of streamlining their activities. Rather the opposite is true---they tend to slow them down. So, one of the only effective ways to get them to use the system is to engage them. If they use the system then you might have a chance at capturing quality data more reliably. Have a clear understanding of how you will get your data out of a vendor's system if you ever leave them **before** you sign the original contract. Never assume that your data will simply take care of itself. Staff appropriately to manage your big data initiatives.

Mike Cottle
Chief Information Officer
Newton Medical Center

Strategies to Leverage Big Data

only large organizations currently have the resources to build in-house solutions. However, in the future, even smaller provider groups and companies will need to tap into one or more big data streams. For these groups, vendor-generated solutions are the only options. When looking at commercially available solutions, ensure that they are sufficiently flexible, scalable and configurable to meet the users' present and future needs.

- Create simple, understandable tools such as dashboards for clinicians on the front lines to visualize incoming data. Organizations should strive to update processes and develop capabilities to enable tool use, and focus on real- or near-real time clinical decision support. Traditional analytics use Extract, Transform and Load (ETL) processes that upload data nightly or weekly to a data warehouse, from which it is then extracted for processing elsewhere. Increasingly big data is moving toward real- or near-real time processing, often at the point of care, to derive value from the data far more quickly for clinical decision support.[46]
- Don't scale up, scale out. Some organizations may be prone to lean toward replacing their older servers with bigger and more powerful servers. Today's trend is to "scale out;" ie, to improve performance and scalability of a system by adding nodes for processing and data storage. This approach may be worth considering because it can make systems easier to manage and to expand to accommodate big data solutions.[47]
- Close the quality loop. Achieving health care transformation requires dramatic and sustainable changes to the structure and processes of health care. Data analytics teams must work in lockstep with quality improvement teams so that analytics tools and techniques can be integrated into the various quality-improvement methodologies which, together, can provide a framework that drives the front-line and administrative changes necessary for achieving desired improvements to health care outcomes and efficiency.

Ideas in Practice

Have a big data analytics "strategy" that aligns with the business and quality goals and objectives of the organization so the output of big data analytics is immediately relevant. Without a strategy that aligns to the overall business, analytics resources are likely to be pulled into working on a series of "one-off" type projects rather than building a strategic information resource for the organization.

- **Ensure that the underlying data model conforms well to business processes and that the entire data model is well documented. This will help prevent differences in the interpretation of what the data "means" in the first place.**
- **Continually validate the output of analytics and predictive algorithms. Sometimes the underlying assumptions of the data used for analytics may change (as a result of change in processes, etc.) and these changes may invalidate current algorithms and increase the risk of analytics returning incorrect information.**
- **Find the right mix of people. Health care data is extremely context sensitive, so it is important to have people who understand both the business/processes of health care as well as the necessary statistical and analytical concepts. Advanced analytics applied incorrectly to underlying data has absolutely no value.**
-

Trevor Strome, MSc, PMP
Informatics/Process Improvement Lead, Emergency Program
Winnipeg Regional Health Authority
Assistant Professor, Department of Emergency Medicine, Faculty of Medicine University of Manitoba

Afterword

It must be emphasized that the health care industry remains well within its infancy of leveraging big data for business and clinical use. Although there have been some successes, many are unproven at the outcome level and much work remains to determine whether those strategies and systems that work best at one facility (eg, The Rhode Island Beacon Community Program, see page 9) can work equally well at another due to cultural, technological, and other types of variables.

It must be emphasized that the health care industry remains well within its infancy of leveraging big data for business and clinical use.

Appendix A

Data Governance Institute Framework

Rules of Engagement

1. Mission
2. Focused Goals, Governance Metrics, and Success Measures
3. Data Rules and Definitions (Policies, standards, guidelines, compliance requirements, business rules, quality rules, and data definitions).
4. Collection of Decision Rights
5. Collection of Accountabilities
6. Collection of Control Mechanisms

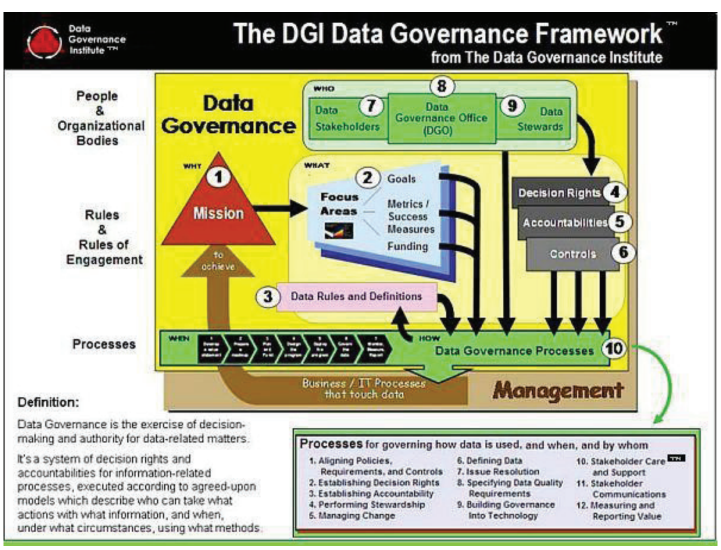
People and Organizational Bodies

7. A Data Governance Office
8. Data Stakeholders
9. Data Stewards

Processes

10. Proactive, Reactive, and Ongoing Data Governance Processes[48]

The DGI Data Governance Framework



References

[1] (TechAmerica Foundation, 2012)

[2] (McKinsey Global Institute, 2011)

[3] (TechAmerica Foundation, 2012)

[4] (Hartzband, 2011)

[5] (The Cleveland Clinic, 2011)

[6] (McAfee, 2012)

[7] (Riskin, 2012)

[8] (TechAmerica Foundation, 2012)

[9] (Soares, 2012)

[10] (TechAmerica Foundation, 2012)

[11] (TechAmerica Foundation, 2012)

[12] (TechAmerica Foundation, 2012)

[13] (Monegain, 2012)

[14] (McKinsey Global Institute, 2011)

[15] (Soares, 2012)

[16] (Gawande, 2012)

[17] (TechAmerica Foundation, 2012)

[18] (McKinsey Global Institute, 2011)

[19] (McAfee, 2012)

[20] (Hani Neuvirth, 2012)

[21] (Klompas M, 2012)

[22] (McKinsey Global Institute, 2011)

[23] (McKinsey Global Institute, 2011)

[24] (DNA Tests to Revolutionise Fight Against Cancer and Help 100,000 NHS Patients, 2012)

[25] (The Institute for Health Technology Transformation, 2011)

[26] (Cerrato, 2012)

[27] (Schiff, 2010)

[28] (Cerrato, 2012)

[29] (Cerrato, 2012)

[30] (Kellerman, 2013)

[31] (McKinsey Global Institute, 2011)

[32] (The Dartmouth Atlas of Health Care, 2012)

[33] (McKinsey Global Institute, 2011)

[34] (McKinsey Global Institute, 2011)

[35] (Oracle, 2012)

[36] (Oracle, 2012)

[37] (Spies, 2012)

[38] (Cusack CM, 2012)

[39] (Wen, 2012)

[40] (Wen, 2012)

[41] (Farnell, 2012)

[42] (Ascenzo, 2012)

[43] (The Data Governance Institute, 2012)

[44] (The Data Governance Institute, 2012)

[45] (The Data Governance Institute, 2012)

[46] (McNickle, 2012)

[47] (McNickle, 2012)

[48] (The Data Governance Institute, 2012)

Bibliography

1. DNA Tests to Revolutionise Fight Against Cancer and Help 100,000 NHS Patients. (2012, December 10). Retrieved from Number 10.gov: http://www.number10.gov.uk/news/dna-tests-to-fight-cancer/
2. Ascenzo, C. (2012, September 20). Retrieved December 12, 2012, from Government Health IT: http://www.govhealthit.com/blog/4-big-data-threats-health-org%E2%80%99s-are-socially-obligated-safeguard-against
3. Cerrato, P. (2012, July 31). Is Population Health Management the Latest Health IT Fad? Retrieved December 12, 2012, from Information Week: http://www.informationweek.com/health care/clinical-systems/is-population-health-management-latest-h/240004578
4. Cusack CM, H. G. (2012). The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. Journal of the American Medical Informatics Association, 1-7.
5. Farnell, J. (2012, November 9). Hot Topics. Retrieved December 12, 2012, from Healthworks Collective: http://healthworkscollective.com/chris-hoffman/62311/four-keys-understanding-big-data-health care
6. Gawande, A. (2012, August 13). Annals of Health Care. Retrieved December 12, 2012, from The New Yorker: http://www.newyorker.com/reporting/2012/08/13/120813fa_fact_gawande
7. Hani Neuvirth, M. O.-F. (2012). Toward Personalized Care Management of Patients at Risk--the Diabetes Case Study.
8. Hartzband, D. D. (2011). Using Ultra-Large Data Sets in Health Care. 2011 Sessions (p. 3). e-healthpolicy.org.
9. Kellerman, A. a. (2013, January). What it will take to achieve the as-yet-unfulfilled promises of health information technology. Health Affairs, 32(1), 63-68.
10. Klompas M, E. E. (2012, November 27). Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. Diabetes Care.
11. McAfee, A. a. (2012, October). Big Data: The Management Revolution. Harvard Business Review, 61-68.
12. McKinsey Global Institute. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey & Company.
13. McNickle, M. (2012, June 13). 5 Basics of Big Data. Retrieved December 12, 2012, from Health care IT News: http://www.health careitnews.com/news/5-basics-big-data
14. Monegain, B. (2012, April 24). Federal Medicare Fraud Detection at an all-time high. Retrieved December 12, 2012, from Health care IT News: http://www.health careitnews.com/news/federal-medicare-fraud-prevention-all-time-high
15. Oracle. (2012, July 17). Oracle Press Release. Retrieved December 12, 2012, from Oracle: http://www.oracle.com/us/corporate/press/1704764
16. Riskin, D. (2012, June 12). Big Data: Opportunity and Challenge. Retrieved December 2, 2012, from Healthcare IT News: http://www.healthcareitnews.com/news/big-data-opportunity-and-challenge
17. Soares, S. (2012, June 13). IBM Data Management. Retrieved December 12, 2012, from IBM Data Mag: http://ibmdatamag.com/2012/06/a-framework-that-focuses-on-the-data-in-big-data-governance/
18. Spies, K. (2012, June 11). Public health's five big data hurdles. Retrieved December 12, 2012, from Government Health IT: http://www.govhealthit.com/news/public-healths-5-big-data-hurdles
19. TechAmerica Foundation. (2012). Demystifying Big Data: A Practical Guide to Transforming the Business of Government. Washington, D.C.: TechAmerica Foundation.
20. The Cleveland Clinic. (2011, October). Cleveland Clinic Unveils Top 10 Medical Innovations for 2012. Retrieved December 12, 2012, from my.clevelandclinic.org: http://my.clevelandclinic.org/media_relations/library/2011/2011-10-6-cleveland-clinic-unveils-top-10-medical-innovations-for-2012.aspx
21. The Dartmouth Atlas of Health Care. (2012). Reflections on Variations. Retrieved from Darmouthatlas.org: http://www.dartmouthatlas.org/keyissues/issue.aspx?con=1338
22. The Data Governance Institute. (2012, December 21). FAQ. Retrieved December 21, 2012, from The Data Governance Institute Web site: http://www.datagovernance.com/fw3_FAQ.html
23. The Institute for Health Technology Transformation. (2011). Population Health Management: A Roadmap for Provider-based Automation in a New Era of Healthcare. Washington, D.C.: The Institute for Health Technology Transformation.
24. Wen, H. (2012, June 11). Big ethics for big data. Retrieved December 12, 2012, from O'Reilly Strata.com: http://strata.oreilly.com/2012/06/ethics-big-data-business-decisions.html

About The Institute for Health Technology Transformation

The Institute for Health Technology Transformation (IHT^2) is the leading organization committed to bringing together private and public sector leaders fostering the growth and effective use of technology across the healthcare industry. Through collaborative efforts the Institute provides programs that drive innovation, educate, and provide a critical understanding of how technology applications, solutions and devices can improve the quality, safety and efficiency of healthcare.

The Institute engages multiple stakeholders:

- Hospitals and other healthcare providers
- Clinical groups
- Academic and research institutions
- Healthcare information technology firms
- Healthcare technology investors
- Health plans
- Consumer and patient groups
- Private sector stakeholders
- Public sector stakeholders

Mission and Vision

The mission of the Institute for Health Technology Transformation: to drive improvement and the effective use of technology throughout the continuum of care through education and collaboration among multiple stakeholders. Technology in-and-of itself will not solve the deep challenges facing our healthcare system nor will it alone ensure more accessible and higher quality care. Realizing the benefits of technology across the healthcare continuum is a complex, under utilized and often misunderstood process. Stakeholder collaboration underscores the Institute's focus working to ensure technology has a transformative effect at all levels of the healthcare sector.

What We Do

The Institute for Health Technology Transformation (iHT^2) provides programs that drive innovation, educate, and provide a critical understanding of how technology applications, solutions and devices can improve the quality, safety and efficiency of healthcare. We do this though a number of vehicles including: educational workshops, access to industry thought leaders, peer reviewed research, high level conferences, webinars, focus groups, topic specific committees, and other unique initiatives allowing individuals and organizations access to resources that will enable them to leverage the full value of healthcare technology.

Institute for Health Technology Transformation
244 5th Ave #2150
New York, NY 10001

iHT²
Institute for Health
Technology Transformation