BIG DATA in Health Care

A Series by REALWORLDHEALTHCARE.org

2016

From predicting disease and identifying targeted therapies and cures, to improving our overall quality of life, big data is transforming the way health care decisions are made and care is delivered.

Big Data in Health Care is a recently published series of articles that brings you the stories behind the research and celebrates researchers and organizations for their commitment to improving health care. Please accept this complimentary copy as our way of thanking you for your commitment to advancing medicine and improving patients' lives.

founding sponsor

HEALTHWELL FOUNDATION®

co-sponsor

DATA SOCIETY

www.RealWorldHealthCare.org

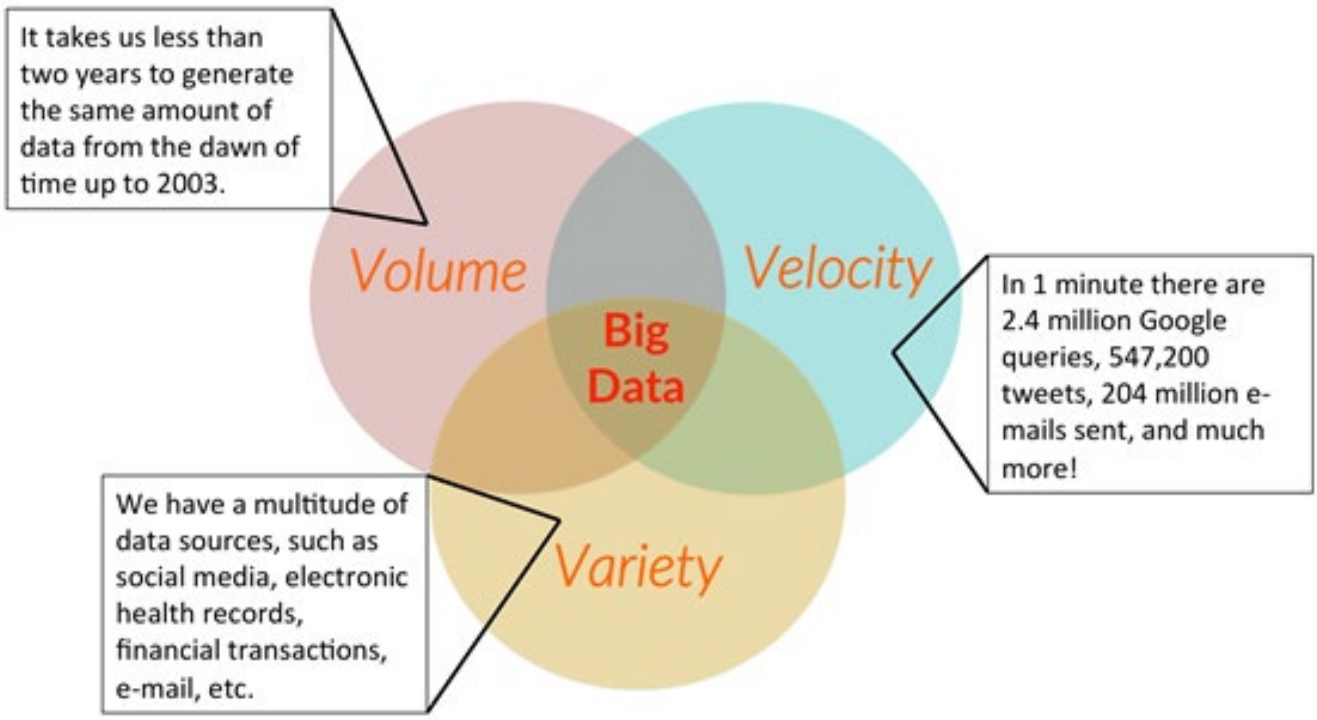
CONTENTS

Big Data in Health Care

Is Big Data Good for our Health? You Bet. Here's Why.

By Cameron Warren and Merav Yuravlivker

The term "Big Data" is increasingly used in our everyday lives. But each mention of it means something different, unique to what we use it for and how we interact with it. Big Data is not information. It's the raw resource that people can use to discover new insights. Just as raw crude needs to be refined to run a car, Big Data needs to be refined to provide useful insights. In 2001, Doug Laney, who currently works for the analyst firm Gartner, defined this raw resource in terms of its three ubiquitous attributes, "the 3 V's" – Volume, Velocity, and Variety.

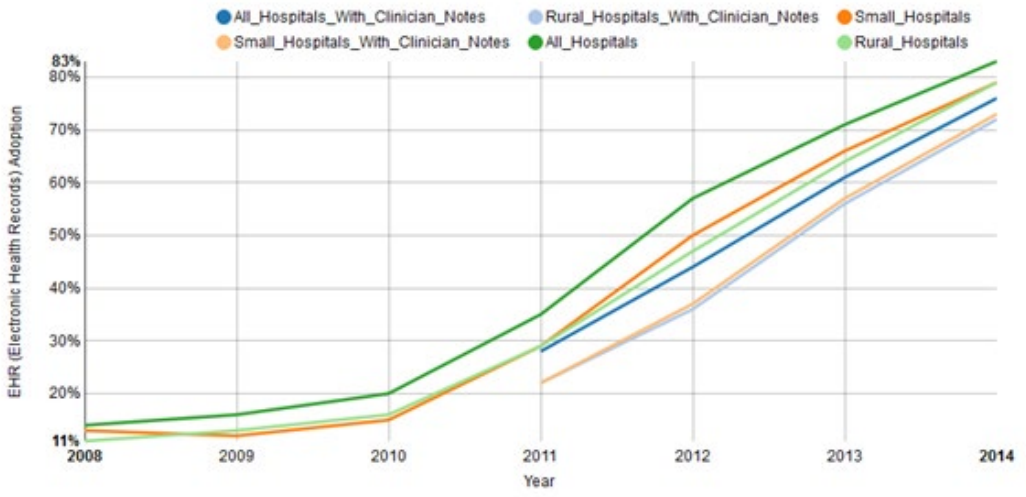


According to Eric Schmidt, the Chairman of Google, today we generate as much data in less than two years than we did from the dawn of civilization up to 2003. That's volume. In 1 minute there are 2.4 million Google queries, 547,200 tweets, 204 million e-mails sent, and that's just 3 categories out of the thousands of ways data is continuously generated. This is velocity and variety.

Making healthcare healthier.

According to the consultancy McKinsey & Co., healthcare represents more than 17% of U.S. GDP, almost $600 billion more than expected for a nation as big and as wealthy as the U.S. There is a lot of waste in the almost $3 trillion dollar U.S. healthcare industry and, for the first time, the new prevalence in well-integrated electronic healthcare records (EHRs) is allowing health insurers and government services such as Medicare and Medicaid to identify fraudulent practices automatically. EHRs have become the norm thanks in part to President George Bush's plan in 2005 to computerize Americans' healthcare information and President Obama's Affordable Care Act in 2009 to incorporate incentives to share healthcare information through health information exchanges. As of 2014, these initiatives have given 76% of hospitals the ability to record and access patient data electronically, which has created a digital health map for millions of people.[1]

Electronic Health Record Adoption 2008 - 2014



Click here for an interactive version

(*note: Clinician notes denote facilities with EHR systems capable of capturing patient-physician interaction via free-form text)

[1] Imler, Dr. Timothy. "Getting Down and Dirty With Big Healthcare Data." The Huffington Post. November 28, 2015

A recently released free mobile phone application by MicroStrategy allows anyone to look at Medicare billings by any physician in the U.S. Information is available based on the number of procedures performed and number of patients treated. Anyone with the technical expertise can analyze this data for patterns and anomalies and identify dubious practices. This is exactly how Medicare found physicians who were inappropriately prescribing well-reimbursed procedures including an ophthalmologist in Florida who billed Medicare more than $21 million in 2012 alone.

Healthcare providers are also generating substantial savings due to the increased quality of the data available. Kaiser Permanente created a new platform to ensure data is shared between all medical facilities. The integrated system has helped the company save over $1 billion from fewer required office visits and tests.

In his book, *Predictive Analytics*, Eric Siegel describes the breadth of uses of Big Data and predictive analysis in the healthcare industry today.

1. Google Flu has shown to forecast an increase in influenza cases at hospitals 7 to 10 days earlier than the Centers for Disease Control and Prevention (CDC) by analyzing online search trends.
2. Stanford University has built a predictive model that diagnoses breast cancer better than human doctors by considering a greater number of risk factors.
3. The University of Pittsburgh Medical Center predicts a patient's risk of readmission within 30 days in order to assist with the decision of release.

McKinsey & Co. estimates that increased integration and sharing of data sources will reduce healthcare costs in the U.S. by $300 billion to $450 billion, and that's not counting the impact of yet undeveloped radical innovations and use cases.

At the individual level, devices are taking patient monitoring to new heights. A new mobile application, Ginger.io, allows physicians to track consenting patients and help them with behavioral-health therapies. Ginger.io collects data about phone calls, texts, location and even motion. Patients also have the ability to complete surveys to better contextualize the data collected about them. The application then combines patient data with research on behavioral health from the NIH to reveal new insights.

Caution: pitfalls ahead

Although Big Data and data science can help the world become a healthier place, the new opportunities are not risk-free. We need to heed the caution signs along the way.

1. ***Privacy:*** data privacy continues to be a problem in healthcare. Medical data can be sent around to third parties as part of administrative processes or prescriptions. In one case, a mother and daughter's medications were mixed up, which led to an unintentional disclosure of a medical condition.
2. ***Data integrity:*** the accuracy of collected data is also a problem. Many patient histories can be subjective and a lot of information concerning prescriptions and patient visits are still entered manually which can be prone to errors. While data can give us many answers, we must also question the source to ensure reliable results.
3. ***Education:*** data analytics are most effective when an industry expert understands the methods and applications of the data. In order for analytics to reach its full potential, healthcare professionals need to be trained to understand the implications behind data analysis.
4. ***Ethics:*** data ethics is still a nebulous area in the data realm. Because much of the data available has come into existence recently, there aren't many standards in place. Even though health insurers cannot use preexisting conditions to reject applicants, they still use prescription data to identify high-risk patients and set rates. Is this ethical behavior? Maybe not, but there are currently no policies in place to prevent this from happening.

So what does the future hold? Perhaps there will be a time where your social media posts about being sad will automatically trigger a notification to your doctor. Or perhaps your Fitbit data will be used to set insurance premiums. Even your diet and medicine could one day be custom tailored to your genetic makeup at the price of a generic drug today. We've already seen the positive impacts that Big Data analytics have had across the healthcare field and, as long as we continue to proceed with caution and foresight, the possibilities are endless for creating a healthier and happier world.

Merav Yuravlivker is co-founder of Data Society, a Washington, D.C.-based organization dedicated to democratizing data literacy by teaching everyone how to turn Big Data into Big Insights. Cameron Warren is a Data Scientist and contributor to Data Society's educational curriculum. Data Society is proud to partner with RealWorldHealthCare.org on its Big Data in Healthcare series.

To read this article on Real World Health Care, click here.

Big Data in Healthcare: Speaking with Dr. Philip Bourne, National Institutes of Health

Our series on Big Data in Health Care continues this week with a conversation with Dr. Philip Bourne, Associate Director for Data Science, National Institutes of Health. Dr. Bourne discusses the goals of the NIH Big Data to Knowledge (BD2K) program and the challenges faced in leveraging big data to improve health outcomes.

Real World Health Care: Why did the NIH establish BD2K?

Philip Bourne: Several years ago, NIH Director Francis Collins set up a data and informatics working group in response to the increasing amounts of digital data being generated in biomedical research. That working group led to the development of BD2K

RWHC: What are the goals of BD2K?

PB: As set out by the working group's report, the goals of BD2K are to promote the "fair" finding, access, sharing, incorporation and re-use of digital content and analytical tools within the entire spectrum of health care. Our goals also include promoting enhanced and diversified training around the process of analyzing large amounts of data and achieving sustainability of the complete digital biomedical ecosystem.

RWHC: "Big data" is a big buzzword these days, and it's being leveraged among a wide range of industries with varying degrees of success. Where does the health care industry currently stand in terms of its overall ability to generate, gather, analyze and share big data toward the goal of improving positive health outcomes?

PB: I think there's a good analogy here between revolutionary changes in other industries and the potential revolutionary change that big data may bring to health care. Take the photography business as an example. When photography went digital, it disrupted the industry and created a completely different business proposition. Today, the photography industry has less to do with pictures and more to do with visual communications platforms like Instagram.

The same kind of disruptor has the potential to happen in health care with the digitization of information. The disruption is happening slowly, even though the growth of digital content has been exponential. Next, we need to get to the "infection point" where big data takes off and becomes disruptive.

Because health care is not a true free market economy like photography, there are more restrictions. Today, the patient really does not have control over his or her health information, but if they had such control, it could be transformative.

RWHC: What are some of the biggest challenges facing the health care industry in terms of its ability to use big data to improve health outcomes?

PB: One of the biggest challenges is the lack of qualified professionals and training for those professionals in conducting analytics. Big data also represents a cultural shift in the industry as we move away from traditional ways of doing research to newer analytical methods. We need more education on the implications of that shift.

RWHC: Where is the health care industry seeing the most success in using big data to improve health outcomes, especially as it relates to health care delivery, treatment optimization and cost containment?

PB: The industry is just beginning to define and use data in different ways, and early stage successes haven't been widely publicized. With that said, our new ability to mine health data records and analyze them to identify changes is statistically significant and provides important predictive tools. For example, researchers at Stanford are studying body mass indexes (BMIs) in specific regions to see if there is any correlation between BMI and the amount of fast food restaurants in those regions.

RWHC: Are there any individual BD2K programs or projects about which you're particularly excited? What sort of initiatives can we expect to see from BD2K during 2016?

PB: We've recently been focusing on privacy, which is a clear issue when it comes to human subjects: How does the industry protect patient privacy? We're also bringing in different types of professionals who have experience in analytics, but in other industries such as digital media and entertainment. We recently had a funding call in this area and are looking forward to seeing how these non-health care professionals apply their skills to biomedical problems — how the entertainment industry can help us visualize large amounts of data in a meaningful way.

We also continue to focus on developing the Commons, which is a shared virtual space where scientists can work with the digital objects of biomedical research. We have a mandate from the federal government and the NIH to promote the sharing and accessibility of research output: outcomes of clinical trials, papers, software and other data. The Commons lets us do that. It's kind of like putting a bunch of Lego blocks in a public square and seeing what people can do with them.

To read this article on Real World Health Care, click here.

Big Data & Health Care: Speaking with Dr. Hallie Prescott

For the latest installment in our series on Big Data & Health Care, we sat down with Dr. Hallie Prescott to discuss the use of structured data and unstructured data in continuously learning health systems. Hallie Prescott, MD, MSc, is Assistant Professor in Internal Medicine in the Division of Pulmonary & Critical Care Medicine at the University of Michigan Health System. She is also a research scientist with the HSR&D Center for Clinical Management Research and staff physician at the VA Ann Arbor Healthcare System.

Real World Health Care: Why do you think Big Data is pervasive in the business world, but not in the health care world?

Hallie Prescott: That's a fairly common observation and one that is difficult to get to the bottom of. There are probably several factors limiting the uptake of big data in health care. First, there is the issue of information privacy. Health care data needs to be highly secure, which can make it difficult to share data across health systems. This type of roadblock tends to limit big data initiatives in health care. The health care systems leading the way in data analytics — the VA and Kaiser Permanente, for example — are successful because they are integrated health care delivery systems.

A second reason why big data initiatives are more widely pursued in the business world is the clear financial incentive to do so. Just look at Netflix. Their use of big data algorithms has given them a competitive advantage. We don't have that sort of free market environment in health care.

Finally, there is the issue of physician and clinician acceptance of big data tools. Physicians still value the art of medicine and like to use their individual decision-making talents to diagnose and manage disease. So, we see some resistance to having a computer tell us what to do.

But even with all these limitations, progress is being made.

RWHC: Health care seems to be moving from the use of structured data to unstructured data. What is the difference between the two when it comes to clinical utility and improving patient outcomes?

HP: Structured data is data that already exists in a spreadsheet format. For example, when vitals signs (temperature, heart rate, blood pressure, etc.) get entered into the electronic medical record, they are stored in a spreadsheet. This data can be examined easily, but does not contain all the necessary information for answering many questions.

There is a vast amount of patient information that's not entered into basic spreadsheets: things like doctors' written notes, radiologists' interpretations of chest x-rays, or pathology reports. This non-spreadsheet data is so-called "unstructured" data, and it often contains very useful information for predicting patients' health outcomes. For example, important lifestyle indicators of health, such as smoking status, are often included within doctors' notes, but not in a structured format.

Traditionally, the only way to learn from unstructured data was to review the medical chart. But, fortunately, we now have automated tools for extracting information from unstructured data sources. For example, natural language processing tools can search for specific words to determine if a patient smokes and how much he smokes.

RWHC: How can big data make positive impacts in a continuously learning health system?

HP: The Institute of Medicine published a report in 2012 on continuously learning health care systems. In a such a system, information is reliably captured, curated, and delivered back to clinicians in order to improve clinical-decision making for individual patients and to improve efficiency and quality of the overall health care system. Learning health care systems require an infrastructure to capture and analyze large amounts of data to inform patient care and system improvement. So, big data is key to a continuously learning health care system.

One way health systems can become better and more efficient is by learning from mistakes at the macro level. As an example, consider what happens to patients in the Emergency Department (ED). As clinicians, we make decisions on where patients should go next: intensive care unit, general medical admission or even sent home. Sometimes, those decisions are wrong, and a patient you send to the hospital ward (or even to home) quickly deteriorates and ends up in the ICU. If we have data-driven models of various factors to consider in making that decision and apply real-time data analytics, we can use them to inform policies and protocols in the ED in order to provide safer care for future patients.

RWHC: Can you give us an example of how you've applied big data in your practice to improve patient outcomes?

HP: At this stage, it's rare to find individual physicians using big data to inform their personal clinical practice. But there are tremendous benefits when you look system-wide. I'm currently studying hospital readmissions after sepsis. We're developing a tool to predict who is at a high risk of coming back to the hospital for specific problems after sepsis, such as for kidney failure, or heart failure, or infection. Because each individual type of hospital readmission happens to only a small portion of the population, we need to identify patterns, and those patterns are only possible when you have huge amounts of data. I'm now looking at the issue within the VA Health System, using over eight years of data to understand these patterns and feed them back to the clinical community to improve patient care.

To read this article on Real World Health Care, click here.

REALWORLDHEALTHCARE.org

Closing the Healthcare Gap: The Critical Role of Non-Identified Information

By Murray Aitken, Executive Director, IMS Institute for Healthcare Informatics

Editor's Note: As part of our series on big data in health care, we are pleased to bring you the following summary of a report by the IMS Institute for Healthcare Informatics, a division of IMS Health that provides objective, relevant insights and research that accelerates the understanding and innovation critical to sound decision making and improved patient care. To request a copy of the full report, click here.

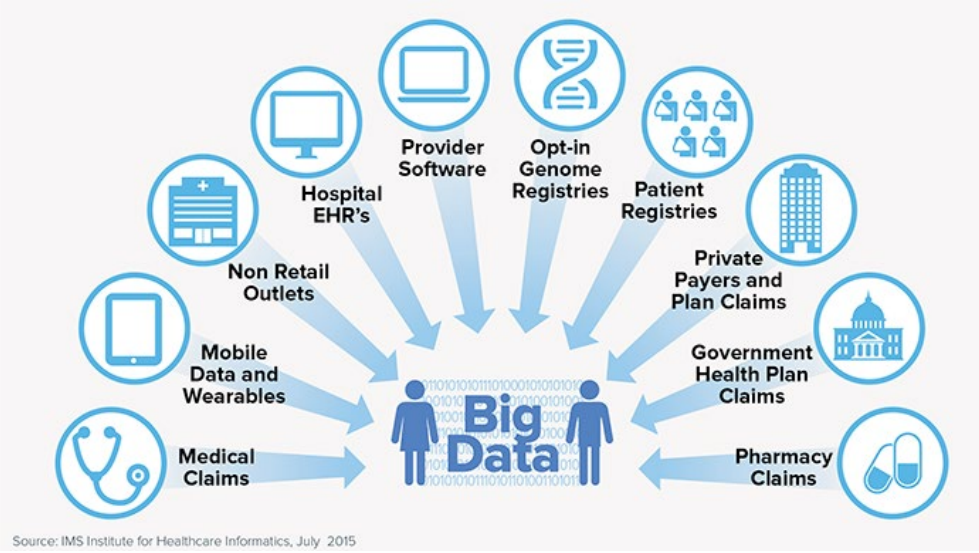
Big data in healthcare brings the opportunity to close the "healthcare gap" — the difference between reality today and what is possible from a clinical, societal and economic perspective. Interactions that patients have with the health system are increasingly being captured digitally, but to maximize the value of this data to advance research and understanding of our connected healthcare system, patient privacy must be protected. Standardized patient privacy and security frameworks for de-identifying this data serve to advance the appropriate use of big data for research across health stakeholders, thereby enabling the data to provide a range of benefits to patients and societies.

"Big Data" Creates New Opportunities to Close the Healthcare Gap by Identifying Issues Guding Solutions and Monitoring Improvements to Patient Care and Access

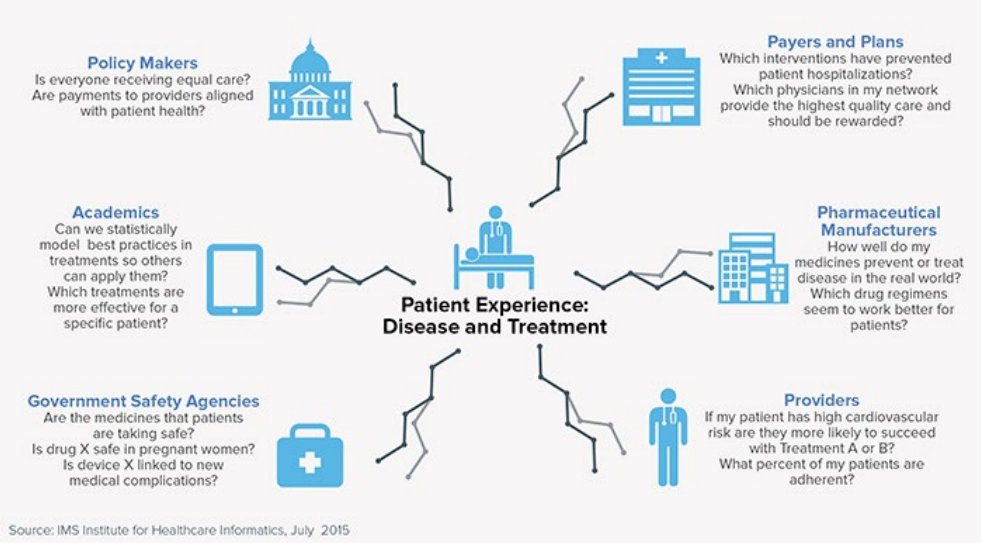
- Interactions that patients have with the health system are increasingly being captured digitally, creating 'big data' that can improve understanding of healthcare usage and care delivery.
- A vast range of research using non-identified longitudinal patient-level data has contributed to efforts that improve health, combat health

disparities and save money.

- Non-identified patient-level data provides real-world medical evidence that accelerates progress and improves performance of health systems around the world.



The most critical benefit of shared information is the ability to enable connected healthcare through data to fully understand what works, what doesn't, at what cost, and with what benefits for individuals and patrons.



- Non-identified patient-level data provides real-world medical evidence that accelerates progress and improves performance of health systems around the world.
- A key benefit of sharing non-identified patient-level information is the ability to create a connected view of patients, critical to supporting a connected healthcare system.

Request a copy of the full report. Readers also may be interested in IMS Health's Real World Evidence Dictionary, an easy-to-use tool that supports a common understanding of Real World Evidence.

To read this article on Real World Health Care, click here.

REALWORLDHEALTHCARE.org

Real World Health Care Interview with Dr. Bonnie Westra

As part of our ongoing series on big data in health care, we spoke with Bonnie Westra, Ph.D., RN, FAAN, FACMI, Associate Professor and Director, Center for Nursing Informatics, at the University of Minnesota School of Nursing. Dr. Westra is also a member of the Institute for Health Informatics and works to improve the exchange and use of electronic health data. Here, she discusses the importance of including nursing data in big data science.

Real World Health Care: Why is the University of Minnesota's School of Nursing spearheading a national action plan to include nursing data in big data science? What are your goals?

Bonnie Westra: Here at the Center for Nursing Informatics, our goal is to lead the discovery, application, and cutting edge thinking for nursing and health informatics scholarship to improve the health of individuals and communities. Nurses and the field of nursing make major contributions to health care. However, as a profession, we need to do a better job of making nursing data more useful for research purposes. We have a good foundation for how to think about nursing data, but we need to move from idea to action and develop standards for capturing, documenting and integrating nursing data with other health care information systems.

Through the Nursing Knowledge Big Data Science initiative, we seek opportunities to standardize and integrate the information nurses gather in electronic health records (EHRs) and other health information technologies. This data is the source of insights and evidence used to prevent, diagnose, treat and evaluate health conditions. The addition of rich contextual data about patients and nursing care will help us develop actionable predictive models that can increase the confidence of nursing leaders' decisions to improve patient outcomes and safety and control costs.

The key element is having nurses involved in health information policy so that nursing data is included in clinical data warehouses for analytics and research.

RWHC: How do you differentiate between the concepts of "big data" and "big data science?"

BW: When thinking about big data, most people consider the volume, or amount of data. But you also need to think about its velocity, or the rate at which data accumulates. Variety is another hallmark of big data. In health care, that can include structured or semi-structured nursing documentation, data from monitoring devices and imaging studies, scheduling and human resource data, and patient-generated data. You also need to consider the veracity, or certainty of the data, in terms of how appropriate it is for either its original purpose — perhaps at the point of use — or for a secondary use in research and analytics.

Big data science has been defined by the National Consortium for Data Science as the systematic study of the organization and use of digital data to accelerate discovery, improve critical decision-making processes and enable a data-driven economy. It encompasses the principled acquisition, curation, exploration, manipulation and interpretation of big data sets.

RWHC: Why is it so important for nursing leaders to understand the value of big data science?

BW: The ability of nurses to make optimal clinical decisions depends on having access to accurate, real-time information, regardless of the care setting. Not only does big data have the potential for improving enterprise operating and financial performance by providing greater visibility to operational issues that support or detract from cost-effective value-based care and services. It also has the potential to improve nursing practice and patient outcomes. It can support improved decision making by offering a comprehensive and synthesized understanding of patients, nurses and organizations. The results of big data analysis enhance confidence in conclusions and can be fed back into systems as clinical or managerial decision support. Plus, it can produce a more robust, timely and valid research agenda.

Unfortunately, many nursing leaders don't come out of school with a background in information science and informatics. They are then put in a spot where they don't necessarily understand how to make sure data is usable.

RWHC: How can big data science help to improve positive patient outcomes?

BW: Big data science can help practitioners comply with evidence-based practice and tailoring treatment to subgroups based on patients' unique characteristics. It also gives us the ability to understand how system characteristics such as staffing models can impact patient outcomes.

As an example, I'm currently working on a study in which we're using EHR data to understand how compliance with the Surviving Sepsis Campaign (SSC) recommendations affects mortality and complications such as kidney or cardiovascular problems. The challenge here is to make sure data is collected and organized in a consistent way so that eventually, we can determine which SSC recommendations work best for which patients.

RWHC: What are some of the main challenges nursing leaders face in terms of accessing and utilizing big data to improve positive patient outcomes?

BW: One of the biggest issues is the lack of nurse informaticians and researchers who know how to create and harness the use of the data. Couple that with competition from other health care priorities such as meaningful use, and the fact that health systems don't receive direct reimbursement for nursing care, and you can see how nursing-related big data becomes a lower priority. We need the health care industry to look beyond financial reimbursement to the overall value of nursing in terms of preventing adverse events and readmissions and improving patient satisfaction.

Nurses are the largest group of health care providers, and it's critical for nurse leaders to have data to demonstrate the impact of their decisions, the value of their practice and how data can facilitate decision-making. We need to prepare nurses for the future with educational programs in informatics and involve nurses in the development of health care informatics technology.

RWHC: What type of support do you think is needed for your efforts from industry?

BW: First and foremost, software vendors in this space need to start collaborating on how to standardize data across disparate systems. A common core is needed. We'll be exploring the topic of standardizing data and processes at the upcoming Big Data Science Conference, June 1-3 in Minneapolis. I invite all Real World Health Care audience members to attend and learn how they can contribute to the future of a national big data science initiative.

To read this article on Real World Health Care, click here.

Big Data Declares a War on Cancer

By Dmitri Adler, Co-Founder and Chief Data Scientist, Data Society

In 1970, cancer was the second-leading cause of death in the United States. President Nixon made fighting this disease a priority in his 1971 State of the Union address: "I will also ask for an appropriation of an extra $100 million to launch an intensive campaign to find a cure for cancer, and I will ask later for whatever additional funds can effectively be used. The time has come in America when the same kind of concentrated effort that split the atom and took man to the moon should be turned toward conquering this dread disease. Let us make a total national commitment to achieve this goal."

Lots of great progress has been made over the past 45 years. Many challenges remain, but the technological capabilities have vastly improved.

In his last State of the Union address, President Obama re-iterated Vice President Joe Biden's plea for a concerted effort to use the brightest minds in the U.S. to cure cancer, and announced the creation of a national cancer moonshot. President Obama asked Vice President Biden to be "in charge of Mission Control." "For the loved ones we've all lost, for the family we can still save, let's make America the country that cures cancer once and for all," Obama said.

The good news is that today there is a massive amount available for cancer researchers to use in their mission. The challenge is that due to the lack of reporting standards and the disparate databases, much of the data is left unanalyzed, which can lead to lots of missed opportunities for breakthroughs.

Since President Obama's declaration, Vice President Biden has met with leaders of the

MD Anderson Cancer Center at the University of Texas which, in 2012, launched the Moon Shots Program aimed at reducing cancer mortality. There are many types of cancers. While they are all driven by gene mutations in various cells, every type of cancer requires a targeted approach. The Moon Shots Program has many mini-projects, or Moon Shots, aimed at treating specific cancers.

The program's innovation is driven by the multitude of specialists involved in the project, from clinicians to biostatisticians and programmers. The Moon Shots include research into B-cell lymphoma, glioblastoma (brain cancer), cancers caused by the human papillomavirus (HPV), high-risk multiple myeloma, colorectal and pancreatic cancers, breast and ovarian cancers, chronic lymphocytic leukemia, lung cancer, melanoma, myelodysplastic syndrome/acute myeloid leukemia and prostate cancer. It covers an unprecedented number of diseases by one effort.

Cancer is a very complicated ailment with very complex treatments. A single tumor can have more than 100 billion cells and each cell can have different genetic mutations. The mutations are not constant over time, which requires an evolving treatment. To understand each cancer, clinicians need to understand the kinds of mutations that are driving it. There are 3 billion code letters, or amino acids, in each cell so understanding the mutations expressed in each tumor is no small task. There are as many as 300 billion opportunities for mutation in just one tumor.

With so much complexity, there are many ways to approach cancer research. For example, scientists at the NIH have used network analysis methods to map out protein interactions to discover new biomarkers and significant players in the cell's architecture. These discoveries help guide clinical studies and other research on gene expression.

Researchers across Moon Shots programs are using machine-learning models to predict whether a patient has various types of cancer based on the expression levels of specific genes. Implementation for thyroid cancer has been especially fruitful. Thyroid cancer usually causes a lump at the base of the neck, and around 5 to 15 percent of these lumps are malignant. By measuring gene expression at the lump, the machine-learning model is able to predict with greater than 90 percent accuracy whether it is malignant or benign. The work was published in *Clinical Cancer Research* in 2012.

Protein data is not the only kind of information used by researchers. Scientists at Case Western University have used machine-learning techniques on Magnetic Resonance Images (MRIs) of breast cancer patients to predict if a patient is suffering from aggressive triple-negative breast cancer, slower-moving cancers or non-cancerous lesions with 95 percent accuracy. Today's capabilities of image analytics can significantly augment the insights gleaned from lab tests. The challenge with cancer is getting a full picture.

Text stored in medical records is another powerful source of relatively untapped data. Modern natural language processing capabilities can analyze massive amounts of unstructured data and combine the results with structured research and clinical information. Combining doctors' notes versus numerical lab tests, for example, can give context to the condition and symptoms of the patient at various stages of different cancers.

Medical records include a treasure trove of data. Factors such as family histories, clinical test results and genomic data are stored in repositories across the world. The challenge is combining all that data in one database.

"Big data is not just big. The term also implies three additional qualities: multiple varieties of data types, the velocity at which the data is generated, and the volume seen within MD Anderson," says Keith Perry, associate vice president and deputy chief information officer.

One of the ambitious objectives of the MD Anderson Cancer Center is to collect and combine patient information including a profile of their genetic makeup, clinical histories, test results, treatment courses and treatment responses. This data will be interpreted by the massive data analytics, which provide real-time decision support to rapidly improve clinical outcomes. This is a much more challenging task than meets the eye.

When the startup Flatiron Health launched with an ambitious goal to improve cancer treatment, one of the largest obstacles they faced is the inconsistency of records from various Electronic Health Record systems (EHRs).

With over $100 million in backing from Google Ventures, Flatiron is facing this basic problem: when measuring the level of a single protein commonly tested in cancer patients, a single EMR from a single cancer clinic showed results in more than 30 different formats. There are over 100 different kinds of protein and genetic tests, biopsies, and other diagnostic methods used in cancer care. And all the various EMR systems out there report these metrics in different ways. This is an incredibly complex data integration problem. So much so that Flatiron purchased Altos Solutions, which makes an EMR service for oncology practices. This allows the company to control the data collection process.

Finding cures and treatments for various types of cancer is truly a Big Data problem. And the ability to collect, store, share and analyze the data cohesively is still in relevant infancy. This isn't a problem you can solve with just one approach. Whether using network analysis, text mining or other machine learning techniques, the task is a true interdisciplinary challenge that requires numerous types of expertise and really Big Data.

Big Data and machine-learning don't hold all the keys, human analysis and contextualization is key. Yet these technologies are starting to shine the light on how humanity will fight one of the most potent killers on the planet. President Nixon's initiative gave us the Frederick Cancer Research and Development Center, an internationally recognized center for cancer research, and has achieved many breakthroughs. President Obama's initiative has the potential to revolutionize the state of cancer treatment. We'll make a comparison in 45 years!

To read this article on Real World Health Care, click here.

REALWORLDHEALTHCARE.org

Big Data in Health Care: Speaking with Dr. Clifford Hudis

Real World Health Care is pleased to bring you the final interview in our series on Big Data and its impact on health care. Here, we spoke with Dr. Clifford Hudis about how Big Data will impact cancer care. Dr. Hudis is Chief, Breast Medicine Service, Department of Medicine, Memorial Sloan Kettering Cancer Center; Vice President for Government Relations and Chief Advocacy Officer for MSKCC; and Professor of Medicine, Department of Medicine, Weill Cornell Medical College. He also serves on the Board of Governors of the American Society of Clinical Oncology's CancerLinQ project.

Real World Health Care: In a recent article, you write that big data represents a new opportunity to increase our understanding of cancer care. How is that so?

Clifford Hudis: The ongoing conversion of medical record keeping in oncology from paper-based records to electronic format means that for the first time in history we have potential access to the treatment and outcomes for the vast majority of adults with cancer who are not treated on prospective clinical trials. This means that we can explore treatment effects including both efficacy and toxicity in patients who might not have participated in the usual, tightly controlled, prospective studies that are used to gain regulatory approval. For example, older (or younger) patients, those with co-morbidities, other malignancies, and so on — all of whom are frequently under-represented in prospective drug-development trials — can be studied.

RWHC: What sort of knowledge gaps do you think big data will be able to identify in the area of cancer care?

CH: Key gaps include toxicities and efficacy in special populations, but also use of drugs "off label" based on either classical histopathologic tumor features or newer genomic testing. Another key area is to study drug-drug interactions or drug-genotype interactions.

RWHC: Can you give us an example of how big data has overcome a known limitation of randomized clinical trials in evidence development?

CH: In other disease areas, such as interventional cardiology, large registries have allowed clinical investigators to refine their understanding of the benefits and harms of specific approaches without the use of conventional prospective randomized trials.

RWHC: What are some of the biggest challenges facing the health care industry in terms of its ability to use big data to improve health care delivery, treatment optimization, and cost containment?

CH: They key challenges may be outside the realm of big data per se. We have a societal challenge in the uniform definition of benefit, efficacy and ultimately value. This is especially true in oncology where drug development costs are high, many diseases are life-threatening, and the pace of innovation has to continue to accelerate. It is possible that big data will allow us to gain deeper and faster insights into some of these issues as new treatments first permeate the treatment arena. At a more mundane level, we would benefit from even greater interoperability and standardization of data storage and access.

RWHC: Much of the literature published on the use of big data in health care focuses on cancer care. Why is cancer care such a ripe area for implementing big data initiatives?

CH: Among the reasons are the myriad diseases — and therefore complexity — that comprise cancer, the acuity of the illness, the broad reach, and the large price we pay in overall public health. In the face of this massive set of challenges, only three percent of adults participate in clinical research that defines and advances the standards of care. To accelerate progress, we need to innovate in the area of data development. Big data is one key opportunity in that regard as it simultaneously offers to provide new insights, broaden the distribution of evolving knowledge, and improve the efficiency of the entire drug development enterprise.

RWHC: How has the use of big data impacted you personally in your practice?

CH: We increasingly have access to patterns of care, treatment decision-making, and patient outcomes across a large and geographically distributed group of clinicians and investigators working in one traditional disease are. All of this can be used to improve patient care in an iterative fashion.