CONTENTS

03: Overview
04: Background
05: Looking Ahead
07: Analysis
08: Correlation
09: Application
10: Conclusion
Introduction: Big Data

“Big Data”, “Artificial Intelligence”, and the “Internet of Things” are terms that are being used more and more frequently in relation to the healthcare industry. While the term “artificial intelligence (AI)” was coined in 1956, its popularity skyrocketed in recent years due to a significant increase in data volume, advanced algorithms enabling machines to “think, work, and react” like humans, and improvements in computing power and storage. Considering the latter, the concept of big data emerged in the 1990s to describe data sets that are too large or complex for traditional databases or data-processing application software to capture, manage, and process in a reasonable amount of time, even with low-latency. Contributing to the accumulation of big data is the Internet of Things (IoT), a term that has evolved since its first use as “Internet for Things” in 1999, when Radio-Frequency Identification (RFID) was essential to the concept. Now, the IoT encompasses everything that is connected to the internet, including sensors, smart phones, medical devices, and wearables, all collecting and uploading in real-time, leading to the rapid accumulation of health-related big data.

The opportunities for big data in healthcare are seemingly endless, but raise a number of questions to consider:

- What types of data are we collecting?
- What are the sources of this data?
- What gaps exist in the data we already have and how can we fill them?
- How is this data currently being used and what are potential other applications for it?
- How do we secure this data to prevent cyber intrusion, loss of data security, and other forms of cybersecurity risk?

This report aims to examine the healthcare industry’s relationship with big data, opportunities to leverage machine learning to support human-centered care, and the hurdles to successfully enabling a joint human/machine intelligence.

---

3. RFID stands for “radio-frequency identification.” This refers to a technology that uses radio waves to scan and read data encoded in RFID tags smart labels. An example of this can be seen at check-out scanners at the grocery store.

"Our cultural definition of healthcare is changing. It is moving away from a brick-and-mortar centric event to a broader, patient-centric continuum encompassing lifestyle, geography, SDOH, and fitness data in addition to traditional health care episodic data. The healthcare industry is on the cusp of learning just how powerful big data is."

James Gaston, Senior Director of Maturity Models, HIMSS Analytics
The Four “V’s” of Big Data

The four principal “v’s” of big data in healthcare include **volume**, **velocity**, **variety**, and **validity**. The sheer **volume** of health and wellness data stems from the growth in sources of data from individuals due to the widespread adoption of the electronic medical record (EMR), the accelerated discovery of determinants of precision medicine (such as genomics), and the rapid advancement in wearable biosensors.

Increasingly, these elements of the digital persona are readily accessible in real time and the **velocity** with which we are bombarded by this digital exhaust is rapidly exceeding the ability of the average clinician to make sense of the information. Beyond these conventional measures of biologic health, there is also a wealth of socio-demographic, geo-locational, and non-medical metadata that all meaningfully interact to determine the state of an individual or population's health and wellness.

Adding to the challenge of creating interoperability across conventional data systems is **variety**; there is now an even greater breadth of data to be integrated across diverse and siloed sources. It is crucial to establish the **validity** and accuracy of all these disparate measurements, their derived inferences, and the actionable conclusions that we infer from this data if big data is to meaningfully shape care decisions.

---

Harnessing Big Data in the Transition to Value-Based Care

While this may seem daunting, the four “v’s” are in fact very well aligned to serve the needs of healthcare transformation from a fee-for-service industry towards an information-enabled and value-driven model of care provision. Access to comprehensive data used to characterize populations as well as precision data to customize decisions to individual needs will critically inform and prioritize healthcare strategies and guide precision care decisions for unique individuals. The increasing focus on preventive intervention, continuous individualization of treatments to evolving patient conditions, and the coordination of care across the patient journey makes accessing, interpreting, and continuously analyzing patient conditions imperative. It also necessitates the timely and contextually relevant processing of high volumes of data. As it becomes increasingly apparent that behavioral health and social determinants of health are important underlying drivers of health outcomes, the diversity and variety of data inputs that need to be integrated to arrive at actionable decisions is rapidly expanding. But how are the relevant signals identified from the noise and how are knowledge and actionable decisions formulated from this data?

Retrieving the Signals from the Noise

Humans cannot begin to scale the existing mountain of data and be able to arrive at meaningful conclusions alone. Lily Peng, MD, PhD, Product Manager in the Google Brain AI Research Group notes that while human intelligence is best suited for integrating small numbers of very “large effect” factors, AI is particularly adept at combing through and identifying patterns in vast numbers of very “small effect”, or obscure, factors. This is the complementary role that machine learning and AI can play as integral partners to human intelligence assisting healthcare providers in grappling with the volume, velocity, and variety of data that is hurtling towards them from all directions. Drawing valid conclusions from these vast sources of data calls for redesigning existing processes for decision-making to incorporate machine learning alongside human intuition and domain expertise to make impactful clinical decisions that enhance value for the patient. Big data and AI can be effective enablers and catalysts for beneficial change and not simply another unwanted layer of complexity in the rhythm of workflow for the practicing clinician, if due consideration is given to its integration into evolving models of care provision and decision-making.

Gone are the days where the patient’s personal clinician served as the all-knowing single source of truth and decisions. Whether talking about care for heart attacks, strokes, trauma, cancer, or complex post-acute care, decisions arise out of shared deliberation amongst a team of individuals, not to mention a patient’s own choices and research about their condition.

As we migrate from applying broad population norms and generalized standards of care towards tailoring care to the customized needs of a specific individual, broad experiential data supported by AI analytics will be required to define the individual norms pertinent to a given patient. While randomized controlled clinical trials conducted upon defined populations serve the purpose of controlling for confounders and isolating the impact of a test intervention in an experimental setting, seldom do the actual test subjects accurately reflect the broad and diverse characteristics of individuals encountered in the real world. From a pragmatic standpoint it will not be possible to design trials to direct care delivery based on all of the nuanced and individualized encounters one experiences in the real world. As such, the parsing and analysis of big data by artificial intelligence will serve an important role in guiding personalized real-world decisions.

5 Lily Peng conference appearances 2018 - Deep Learning for Medical Imaging https://www.youtube.com/watch?v=MXm1EmUJuaU
Leveraging Big Data in Clinical Decision-Making

There are four potential challenges to overcome if big data and AI are to be effective in supporting clinical decision-making. Successful enablement of “intelligent” human and machine learning from the wealth of data to which we have access will require the following conditions:

1. Removing bias in data collection
2. Acknowledging the inherent conflict between anonymity and specificity
3. Meaningful validation of collected data
4. Understanding underlying cause and effect relationships

1. Overcoming Biases in Data Collection

Healthcare data is messy. At the most basic level, notwithstanding efforts to bring order to medical nomenclature, diagnostic coding, etc., there are wide disparities in how individual providers describe, conceptualize, and articulate their observations about patients. As a rule, the exploration, discovery, and analysis performed on any data is only as valid and valuable as the clarity and validity of the underlying data set. These problems are only compounded by the volume and velocity of data that is collected and must then be interpreted. Standardization, semantic classification, and agreed upon conceptual ontologies are some of the necessary steps in the “data cleanup” that is required to bring order to big data sets before they are ready for useful analysis by AI techniques.

In addition to these considerations, the lens that each investigator brings to big data creates inherent biases. Biases can include the categories of data evaluated and how that data was gathered (e.g. what populations were sampled and what sampling tools were used that could selectively include or exclude representation in the data set). It is assumed that the power of high-dimensional data resides in the absence of hidden confounders that remain undisclosed in the data. Unfortunately, this assumption is far from being a forgone conclusion and poses a threat to the validity of conclusions derived from big data by AI techniques. For instance, failing to consider measuring a variable that is a significant driver of a desired outcome, the apparent conclusions may be incorrect and misleading if applied to circumstances where the covariate relationship with the confounder changes.

This hits at the root of the intersection of human domain experience and artificial intelligence and the “large effects” processed by the human brain that would have been missed for small effect factors being examined by the machine.
2. Anonymity at Odds with Specificity

In theory, the process of harnessing the power of big data (or “the wisdom of the crowd”) should permit the protection of personal identity and the security of protected health information through the process of anonymizing the sources of individual data points. This would be true if the output analysis was restricted to population or large population segment conclusions. The paradox is that the value of real-world big data is that it can also be analyzed to provide insights that guide personalized precision medicine decisions for individual patients. The breadth of big data incorporates metadata elements that have the potential (when viewed with sufficient granularity and integrated into the context of other personalized data elements) to permit the deanonymization of personal identities.

At the end of the day there is a trade-off between the value generated through open sharing of big data and the small, finite risk of re-identification of data sources with the potential for intrusion into patient privacy. Due precaution must be taken to structure analyses such that reverse engineering of patient identities does not occur; however, it is worth noting that the benefit of shared open data exceeds the adverse potential for re-identification of the individual. Society will have to come to grips with weighing the ethical trade-offs between the benefits of shared open access to data and the finite, but real, possibility for re-identification of individual people by reverse engineering of the segmented data. Human intelligence, not artificial intelligence, will be required to grapple with these questions.

3. Creating Measurable Impact

It is reasonable to assume that more robust high dimensional representations of patients and their conditions will lead to a better understanding of the circumstances driving a given disease process. However, it remains to be proven that impactful interventions guided by such data and analyses will reduce cost, enhance outcomes, increase satisfaction, and improve the consumer experience. Thus the integration of data, AI-derived knowledge, and informed clinical decisions must be adopted by and tightly interwoven into clinical processes and workflow to drive potential benefit in the care of patients. Appropriately structured clinical trials are needed to demonstrate that the incremental benefits of a data-driven care process justify any costs (and complications) incurred by these decisions.
4. Correlation Does Not Imply Causation

Defining causal relationships is critical to begin transforming observed patterns in the data to informed interventions, where presumed causal variables can be altered to achieve the proposed outcomes. First and foremost in this process is ensuring that the data subjected to analysis does not suffer from the omission of confounders that may be causally related to the measured outcomes. Domain expertise and human intuition will always be required to work in tandem with artificial intelligence to confirm the absence of hidden confounders. High dimensional data, on the other hand, provides the opportunity to identify blind spots that were not contemplated by the human mind as being causally related to outcomes due, perhaps, to the inherent biases and heuristic assumptions embedded in human domain expertise. The use of machines can help humans reveal these undiscovered or unanticipated variables.

Prior to the existence of intelligent software that could handle large datasets and be programmed to think like a human, well-constructed controlled randomized trials were, and continue to be, used to avoid the biased distribution of hidden confounders. However, real world big data is not always dichotomized between intervention and control groups and often suffers from significant gaps. A randomized control trial or cohort study cannot always solve for the missing piece of the puzzle. AI and machine learning can now provide statistical tools for attributing measurements to fill the data gaps and synthetically construct “controls” for comparison to real world experiences. These tools provide a path forward for making comparisons between observed outcomes from a given intervention and the expected outcomes in the absence of that intervention so that we can simulate the testing paradigms that permit hypotheses about deterministic and causal relationships.

Colin Hill, Chair, CEO & Co-Founder of GNS Healthcare, envisions a process for causal machine learning that begins with inferring potential causal mechanisms by examining relationships in high-dimensional data. Using this information, one can then “reverse engineer” probable causal relationships that can be tested in simulated scenario environments. This is called “forward simulation”, and gives researchers the ability to test the validity of causal hypotheses that cannot easily be examined in the real world.

As applied to drug discovery, Mark Murcko, PhD, Chief Scientific Officer & Co-Founder of Relay Therapeutics theorizes how forward simulation of drug-to-target interactions can be used to perform in-silico screening of potential drugs for effectiveness against biologically validated targets. The simulation is based on a data-driven understanding of the changes in protein motion and function when a drug is deployed.

These approaches towards understanding causality represent the combination of human domain expertise and artificial intelligence applied against vast data sets to predict therapeutic interactions between screened compounds and biological targets of disease processes.
Real-World Application of Big Data

While there are many theoretical opportunities to apply big data in research, AI and machine learning are already making waves in care delivery. The following highlight companies that are leveraging big data for patient triage, diagnostic imaging, and predicting drivers of practice variation, adverse outcomes, and treatment impact.

Twiage, for example, is tackling emergency medicine, helping hospitals to track metrics, allocate resources, and improve reaction time to significantly impact the outcomes of stroke, heart attack, sepsis, and trauma patients. Buoy Health is using big data and AI directly geared at patients to triage illness and direct them to the appropriate care setting. A patient may use the online application to chat with a bot, describe their symptoms, and be guided through a set of questions similar to what they would experience in a physical care setting.

As medical imaging becomes more advanced and in-demand, Zebra Medical Vision aims to empower radiologists to more quickly identify irregularities in imaging results. Zebra's AI algorithms can identify medical conditions, serving as a first set of "eyes" and increasing the speed at which radiologists can perform their work.

Putting into practice the aforementioned theories on reverse engineering and forward simulation, GNS Healthcare uses big data and AI to run in silico clinical trials to understand how drugs will work in the real world as well as helping to identify the best intervention and timing of the intervention for individual patients. Agathos provides an analytics platform for health system that gathers insights and gives feedback to providers, allowing them to see individualized and aggregate trends in their patient data, helping to inform staffing, pre-rounding, and coaching needs, among others to improve clinical workflow and patient outcomes. PhysIQ and Pascal Metrics are targeted real-time patient monitoring solutions that use machine learning to detect slight changes in patient condition, either post-acute or while admitted, and alert their care team for any necessary action.
So how do we use artificial intelligence on big data to drive healthcare innovation?

New applications for big data are growing just as rapidly as the data itself. As we continue to develop new ways to incorporate big data into AI, it is critical for humans to be aware of the need to:

- “Clean” any collected data for potential biases
- Standardize the method of collection or unification of data
- Agree upon the proper utilization of anonymized information
- Avoid falling into the trap of correlation versus causation

---

These points are echoed by CEO, President & Co-Founder of NVIDIA, Jensen Huang, who also emphasizes the importance of “data training” or the process of learning from digital experience. At the same time that machine learning algorithms are rapidly improving on their capabilities, humans must learn to work smarter and adapt to a “new normal”, allowing machines to automate commoditized tasks and liberate providers to perform the human tasks of compassionate care. In this way, human and artificial intelligence can work in tandem to achieve new heights in data analytics, clinical decision-making, and innovation in healthcare.

In the past, software engineers crafted programs and meticulously coded algorithms. Now, algorithms learn from tons of real-world examples — software writes itself.

In the past, software engineers crafted programs and meticulously coded algorithms. Now, algorithms learn from tons of real-world examples — software writes itself. Programming is about coding instruction. Deep learning is about creating and training neural networks.

Jensen Huang, CEO, President & Co-Founder, NVIDIA

About Healthbox

Healthbox, a HIMSS Innovation Company, is a healthcare advisory firm that leading organizations trust with innovation and digital strategy development and execution. Healthbox drives innovation from the inside and out, helping organizations build internal innovation programs in addition to assessing the commercial potential of employee-led projects. We also help them look to the market to find solutions to implement or invest in. Through both approaches, Healthbox provides the tools and support needed to improve care and support organizational growth. We are proud to work with industry leaders who share our passion for building, harnessing, and advancing solutions to empower the reinvention of healthcare.

Eric Louie, MD and Jessica Baker | © Healthbox 2019

---
