

# Big data in healthcare: a new frontier in personalized medicine

## Abstract

Big Data has become a technology utilized more frequently in healthcare than ever before. As healthcare data grows ever more heterogenous and unstructured data becomes more and more difficult to process, researchers have discovered new uses for Big Data in healthcare research and in the personalization of medicine based on many factors. Big Data has long been used in the business world to manage substantial amounts of data and predict outcomes. Since the advent of the electronic medical record (EMR), healthcare providers and researchers have instituted Big Data as the means to manage the voluminous amounts of data generated by patients in today's healthcare arena. A new use for Big Data has evolved from these applications: the use of Big Data to streamline and personalize medical treatment based on the patient's previous data and current data, perhaps collected with sensors or through social media. This Mini Case Report outlines the use of Big Data in the personalization and streamlining of treatment based on the individual patient.

**Keywords:** big data, healthcare, personalized medicine, hypertension, diabetes, heart attack, electronic medical record

Volume 1 Issue 1 - 2017

Cheryl Ann Alexander,<sup>1,2</sup> Lidong Wang<sup>3</sup>

<sup>1</sup>Department of Nursing, University of Phoenix, USA

<sup>2</sup>Technology and Healthcare Solutions, USA

<sup>3</sup>Department of Engineering Technology, Mississippi Valley State University, USA

**Correspondence:** Lidong Wang, Department of Engineering Technology, Mississippi Valley State University, Mississippi, USA, Email lwang22@students.ntech.edu

**Received:** May 30, 2017 | **Published:** July 31, 2017

## Introduction

The electronic medical record (EMR) generates information on multiple aspects of patient care. Enormous datasets are now available describing patient behaviors, signs and symptoms of diseases, human-derived information, insurance claims data and pharmacy refill records, providers notes, and imaging. Regular database management systems we have seen in the past are not able to handle these huge datasets. To manage the substantial amounts of data collected about any one given patient, providers have begun to utilize Big Data Analytics as the primary method for handling this copious amount of data.<sup>1</sup> Table 1 shows the characteristics of big data in the healthcare area. Data is collected not only from the EMR, but also frequently the patient uses sensors, social media, and smartphones to generate data. For the healthcare provider, it can be overwhelming to mine the data for any specific piece of information that may be critical to the patient's care. Big Data however, can handle voluminous amounts of heterogenous data; structured, semi-structured, and unstructured. Much of healthcare data is now either unstructured such as medical notes, nursing notes, or other provider-written information, or semi-structured 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".<sup>2</sup>

It is imperative for providers to recognize the implications for the use of Big Data in healthcare through case studies and premier research. Big Data analytics is unquestionably applicable to healthcare data. Volume, velocity, variety, veracity, and variability are a few of the most important qualities of big data. Big data lowers costs, and reduces waste and fraud in insurance claims, contributes to reducing costs associated with long term sequelae of diseases (e.g. comorbidities), and can often predict or plan for necessary data to permit successful treatment prior to or during the course of an illness. It is frequently possible to predict patient disease course, outcomes,

and suggest an optimized treatment approach based on the assessment of collected datasets. This can be through a variety of datasets as mentioned earlier such as sensors, smartphones, the EMR, social media, websites, and research articles. Data often includes imaging data (e.g. X-rays, MRIs, CTs, etc.), provider notes, nursing notes, genomics, research data, insurance and pharmacy records, data from social media and other web-based applications (e.g. IoT, Twitter, Facebook, etc.), telemedicine data (e.g. home-based monitoring, eICU, etc.), Smartphone apps, and sensors, which have become more frequently used in the monitoring of patient data, especially the elderly. This can often result in early disease detection, prediction, and the best evidence-based treatment plans.<sup>3-5</sup>

**Table 1** Six V's of big data (volume, variety, velocity, variability, veracity, variability, and value) that apply to health data

Characteristics	Description
Volume	High-Throughput Technologies, continuous Monitoring of Vital Signs
Variety	Heterogeneous and unstructured data sources, differences in Frequencies and Taxonomies
Velocity	Increasing data generation rate by the Health Infrastructure, high-speed processing for fast Clinical Decision Support
Variability	Seasonal Health Effects and disease Evolution, non-deterministic models of illness and health
Veracity	Data coming from uncontrolled Environments, Data Quality is Unreliable
Value	Clinically Relevant Data, Longitudinal Studies

Apache Hadoop, Map Reduce, HDFS, data mining techniques, and visualization, etc. are invaluable tools of big data. For example, Apache Hadoop can assist providers in predicting heart attacks in victims prior to actual heart damage. Big data can also be used

to detect previously unidentified patients with diabetes. Patients with hypertension can be identified prior to actual diagnosis and comorbidities can be prevented, which can impact healthcare spending and reduce waste, and patient mortality. However, privacy, cyber security, and eliminating hacker attacks is an eminent concern in big data; although data is de-identified, it can be re-identified under certain circumstances. Ultimately, however, all stakeholders benefit from the use of big data analytics.<sup>6,7</sup>

## Big data in the treatment of hypertension

The measurement of blood pressure (BP) is an essential tool in assessing patient status. Most nurses and providers rely on BP as a measurement of the patient's overall health status for a number of disease processes. When the BP is abnormal, there can be many reasons and the provider must sort through those reasons to determine if the patient has hypertension, or if the patient is simply displaying a reaction to some other disease or circumstance. However, the importance of BP cannot be understated as many studies have established a relationship between increased blood pressure and long-term morbidity and mortality. Since most patients are asymptomatic, hypertension is known as the "silent killer" because most patients are unaware that they even suffer from the disease. Research has identified an increased prevalence within the aging and obese. Studies have also shown the control of BP reduces comorbidities, especially cardiovascular disease. The proper technique and cuff size, however, is important as it will increase the efficiency and accuracy of BP measurement. In more than 90% of cases, hypertension (HTN) is essential, or without a known cause but it does have a hereditary component. Secondary hypertension occurs in less than 10% of cases and has an identifiable cause: including chronic kidney disease, renovascular disease, street drugs, prescription drugs, natural products, food, or industrial chemicals. A normal adult BP is <120/80; prehypertension is 120–139/80–89; Stage I hypertension is between 140–159/90–99; and Stage II hypertension is  $\geq 160/\geq 100$ .

A phenomenon identified as "white coat hypertension" occurs when an elevated BP is only noted in the clinical environment; however, patients with white coat HTN may still have an increased risk for cardiovascular disease. Therefore, home measurement should be strongly encouraged, and is now possible with smartphones, sensors, and telemedicine etc. which can deliver results directly to the provider for more specific treatment.<sup>8,9</sup>

Treatment is often based on modification of lifestyle risk factors, but is generally poorly followed; even when medication is warranted, most patients skip their medication entirely when symptoms disappear, do not have insurance to cover the cost of antihypertensive medication, or for some other reason, including cultural reasons, do not take their prescribed medication. Current treatment fails patients at considerable risk, who need more attention, and by diverting the resources; many more are failed because a population-based public health approach which attacks the structural drivers of HTN, such as cheap, empty calories, excess sodium and sugars, tobacco and heavy alcohol use, and a sedentary population is absent.<sup>10</sup>

Due to the prohibitive cost of hospital readmissions, medications, and insurance coverage among patients with HTN, it has become imperative that stakeholders investigate the best way to deal with the data explosion to determine the best methods to evaluate, treat, predict, and prevent HTN. Some characteristics of big data go beyond size and volume to include variety, velocity, and, specifically to health

care, veracity. Big data provides a strong platform for examining the enormous amounts of data from sensors, EHR, insurance claims, and pharmacy refill records, social media, etc. to determine cause, evaluate, predict, and find the best evidence-based treatment plan (personalized medicine). The use of Big Data analytics in predicting, evaluating, and treating hypertension among adult populations is highly important. Big Data can be used to personalize the treatment of patients with HTN through by analyzing datasets and data subsets generated by social media, smartphones, sensors, telemedicine, the EMR and their past medical history, etc. The use of Big Data in predicting patients with HTN can be helpful for any number of providers who wish to review their datasets for patients who remain at risk for the disease.<sup>11</sup> Big Data can be used to target and identify patients with hypertension, and mine or recommend appropriate care, and tailored treatment by providers. Better disease management to reduce long term costs is available through using big data from a wide variety of sources. Big data analytics has shown that home measurement is a good choice to make diagnosis and identify hypertension.

It is imperative that researchers continue to look for ways to optimize treatment for this deadly disease. Although Big Data is only one approach, combined with current treatment modalities, the use of Big Data in personalization of the treatment of HTN is invaluable.

## Big data and the personalization of heart attack treatment

Acute myocardial infarction (AMI) is the number one cause of death in the US. An AMI occurs when blood flow to the heart muscle is disrupted, causing the heart muscle to become damaged and/or die. Most heart attacks are caused by a reduction in oxygen-rich red blood cells; the heart muscle begins to arrest in as few as six to eight minutes without oxygen, potentially leading to death. Atherosclerosis, a build-up of plaque in the arteries over many years, is a common cause. The healthcare costs associated with patients who are diagnosed with cardiovascular disease is extensive and can be reduced using modern methods of dataset handling and more effective, personalized treatment plans such as Big Data analytics.<sup>12</sup> Big Data analytics helps process data from sensors, the Internet of things (IoT), and telecardiology in myocardial infarction prediction; mine valuable information; tailor medical treatment to individuals; and improve the management and service in myocardial infarction.

The EMR has revolutionized dataset management. Data is cheaper, larger, and includes a broader patient population. Data is noisy, heterogeneous, diverse, and longitudinal. Big data is used to harness data as most healthcare organizations discover opportunities to better understand and predict customer behaviors and interests (i.e., personalized medicine). Big data surpasses the processing capacity of traditional systems. Data is too big, moves too fast, or doesn't fit the strictures of conventional database architectures. Big data fosters novel opportunities to predict and/or more rapidly respond to critical clinical events, generating better health outcomes, and more efficient cost management. The key for reducing the mortality and morbidity associated with cardiovascular disease is found in using big data to mine the enormous amounts of data: omics data, phenotype data, social media, insurance claims information, the electronic medical record (EMR), etc.<sup>13</sup>

Apache Hadoop is an open-source software framework written in Java used primarily for distributed processing and storage of enormous datasets on computer clusters. Hadoop Distributed File System

(HDFS) supports cloud computing using Hadoop. The Text Mining based Hadoop platform is used to create more precise information about comorbidities by converting the patient's unstructured generated data to structured data. Disease prediction, prevention, and personalized medicine are a result.<sup>6</sup> Generally, specific applications and data processing need to be developed, implemented, and built on top of the platforms like Apache Hadoop to extract data for further analysis through the platforms and draw values from data. These kinds of applications and processing include data cleaning (such as error checks and correction, missing data handling, and duplicates removing), data integration, and data dimensional reduction through methods such as principal component analysis (PCA) and factor analysis, etc.

## Discussion

For healthcare providers, Big Data provides unlimited opportunity and access to manage voluminous datasets developed by the patient in the form of the EMR, social media, smartphones, sensors, etc. Technology is expanding at an exponential rate and traditional data management techniques are simply ill-equipped to manage the substantial amounts of data produced by a single patient, much less by the thousands of patients a healthcare organization must examine and mine daily. Big Data is rapidly evolving to produce any number of valuable tools to handle the datasets produced by any one given organization, as well as the thousands, even hundreds of thousands of healthcare organizations that must analyze data daily.

Because most EMR systems do not communicate with each other, the process and goal of data sharing among healthcare providers and organizations has been hampered greatly. With the use of Big Data tools, providers and organizations can still formulate trends, treatment plans, and outcome data even by using multiple EMR systems. Big Data can formulate treatment plans based on the best evidence-based research data, past and present patient data, and personalized applications such as sensors, etc. These tools will become invaluable as healthcare becomes even more cost-conscious. Reduction of waste, the appropriate use of data for healthcare providers, and reducing fraud are common threads for all healthcare organizations and providers.

Healthcare providers will also need to learn how to harness the numerous uses of Big Data to personalize the patient's treatment plan, develop and implement a treatment plan based on the best evidence-based research, and incorporate data from a variety of sources. Currently, the provider can utilize big data to predict, manage, and personalize treatment plans using Big Data tools.

The most important threat to the use of Big Data analytics is privacy. Patient privacy has been regarded as the most prominent issue in healthcare since the advent of the Health Insurance Protection and Portability Act of 1996 (HIPPA). All healthcare providers and organizations are mandated to keep personalized information related to patient care private and if information is provided for research or other approved purposes, it should be de-identified. All data can be re-identified in certain circumstances and researchers still need to conduct further research to determine the best methods to prevent this phenomenon. However, despite HIPPA and privacy issues, Big Data is still the best method to manage healthcare datasets; far more efficient than traditional methods.

The challenges of Big Data in various aspects of healthcare are quite different. In comorbidity, correlated factors, and predictive

modeling, challenges include processing huge volumes of medical imaging data; handling the data with ungrammatical phrases, abbreviations, misspellings, and semi-structures; mining potentially beneficial information in real time; and leveraging the patient/data correlations in longitudinal records, etc. In looking at patient pools to identify patients at risk, it is often difficult to capture behavioral data through multiple sensors and process the data in real time. In disease management of individual patients, data collection, data sharing, information security, and privacy, etc. are challenges.

## Conclusion

Big Data in healthcare is the most likely method to manage large datasets generated by patients within and among healthcare organization. Big Data tools such as Hadoop, HDFS, Apache, and Map Reduce, etc. have established themselves as the leaders in efficient data management, far more efficient than the burdensome traditional methods that leave much to be desired. Big Data can reduce costs, reduce waste, prevent fraud, and personalize treatment. Providers and organizations will both benefit from the use of Big Data analytics in healthcare.

## Acknowledgements

This study was supported in part by Technology and Healthcare Solutions, Inc. in Mississippi, USA. The authors have no funding to report.

## Conflict of interest

Author declares no conflict of interest.

## References

1. Andreu-Perez J, Poon CC, Merrifield RD, et al. Big data for health. *IEEE journal of biomedical and health informatics*. 2015;19(4):1193–1208.
2. Cottle M, Hoover W, Kanwal S, et al. *Transforming Health Care Through Big Data Strategies for leveraging big data in the health care industry*. Institute for Health Technology Transformation. 2013.
3. Sun J, Reddy CK. Big data analytics for healthcare. Tutorial presentation at the SIAM International Conference on Data Mining, Austin, Texas, USA; 2013. 327 p.
4. Azimi I, Rahmani AM, Liljeberg P, et al. Internet of things for remote elderly monitoring: a study from user-centered perspective. *Journal of Ambient Intelligence and Humanized Computing*. 2017;8(2):273–289.
5. China's Meridian Medical Networks Uses Trusted Analytics Platform (TAP) to Build Big Data-Driven Hypertension Risk Model.
6. Ghadge P, Girme V, Kokane K, et al. Intelligent Heart Attack Prediction System Using Big Data. *International Journal of Recent Research in Mathematics Computer Science and Information Technology*. 2015;2(2):73–77.
7. Ebenezer JGA, Durga S. Big Data Analytics in Healthcare: A Survey. *ARPN Journal of Engineering and Applied Sciences*. 2015;10(8):3645–3650.
8. Cheryl Ann Alexander, Lidong Wang. Big Data Analytics in Identification, Treatment, and Cost-Reduction of Hypertension. *American Journal of Hypertension Research*. 2017;4(1):1–8.
9. Klimek P, Kautzky-Willer A, Chmiel A, et al. Quantification of diabetes comorbidity risks across life using nation-wide big claims data. *PLoS Comput Biol*. 2015;11(4).

10. Martin SA, Boucher M, Wright JM, et al. Mild hypertension in people at low risk. *BMJ: British Medical Journal (Online)*. 2014;349:1–8.
11. Groves P, Kayyali B, Knott D, et al. The big data revolution in healthcare: Accelerating value and innovation. 2016.
12. Morley SR. Heart Attack Experiences Described in Weblogs: An Analysis of Sex Differences. *CMC Senior Theses*. 2013.
13. Sun J, Reddy CK. Big data analytics for healthcare. In Proceedings of the 19<sup>th</sup> ACM SIGKDD international conference on Knowledge discovery and data mining, USA; 2013. p. 1525–1525.