# Big Data Analytics for Big Outcomes in Healthcare

**Hailey Michael and Shankar Krishnan, PhD**

**Wentworth Institute of Technology, Boston MA**

Abstract:
The National Cancer Institute estimated that there were over 1.7 million new cancer patients in the United States in 2018 and about a third of these patients will not survive. According to the World Health Organization, cancer is the second leading cause of death globally. Traditional cancer treatments are expensive, and the high prices combined with the low survival rates could turn patients away from going through appropriate cancer screenings and detection procedures. A significant issue within the healthcare system is the lack of personalization among treatments. Patients are looked at from a disease cluster consideration, rather than as an individual patient requiring personalized care. Large volumes of complex data, forming the big data, can be reviewed and analyzed to arrive at a personalized plan for patient case management. The objective of this project is to review examples of current applications of big data in healthcare, highlight the corresponding benefits, and make suggestions for future improvements.

Big data analytics refers to the comparison and utilization of high volume, variety, velocity and veracity of relevant data to select a treatment approach for a given patient. Applying the analytics on relevant big data and zooming in on a specific patient's case to determine appropriate regimen and personalized patient care will help to reach the outcomes with better chances at survival, especially for patients with cancer or cardiovascular diseases.

Results based on big data analytics for cancer treatment are presented and discussed. The benefits of big data analytics include improving diagnoses, reducing readmission rates, providing more effective treatments and substantially reducing associated costs and corresponding time. Incorporating user-friendliness in big data analytics is difficult since massive volumes of data must be organized into appropriate simulations to achieve better results.

In conclusion, based on the potential power and preliminary results, it is predicted that big data analytics will play a crucial role in healthcare management of complicated cases and care delivery, leading to big and effective outcomes.

Keywords: Big Data, Healthcare Analytics, Medical Data Analytics, Cancer Treatment

## Introduction:

Heart disease and cancer are currently the leading causes of death in the United States, taking an estimated 12 million lives yearly [1]. Each patient has their own set of medical records, endless hospital visits, treatment plans and medication. That all comes after the diagnosis. Each person has a different genetic make-up, family history, race, gender, age etc. A diagnosis should consider all of the data associated with a patient. Big data analytics can be used within the health care domain to zoom in on an individual case while simultaneously comparing it to similar cases. One weakness in health care delivery is not the types of treatments, but the lack of personalization among treatment plans. Every patient has a different background and deserves to be looked at as an individual. With the efficient use of big data analytics, heart disease and cancer survivors are likely to increase.

Due to the fast-paced work place and the high demand of physicians, it may be difficult for a doctor to take the proper amount of time on every patient. With the implementation of big data analytics software in hospitals, this process could be done in shorter times. The mass amount of patient data within hospital databases could be scanned and compared with the click of a button. Currently, big data is used throughout multiple industries to compare data and aid in predicting outcomes. Using the advanced technology already being used in hospitals today the data could be collected and stored securely, only to be accessed by approved users (i.e. doctors, surgeons, specialists).

It is important for the software to be as user-friendly as possible to avoid possible misinterpretation of the data. The mass amount of data must be organized in a way that will be easily understood by any necessary healthcare provider. The healthcare providers would also have to be educated in the proper use of the software. The documentation of data must be detailed and consistent. Ideally, multiple hospitals (if not all) would be connected through one single database. Larger amounts of data will lead to more accurate diagnoses. It is also crucial that the data is secure and remains protected within the database. The organization of the data will also affect how it is interpreted and therefore it should be clear and concise. The objective of this paper is to outline the benefits of big data analytics, provide examples of the current applications in healthcare and make suggestions for future improvements.

## Background:

Big data analytics is defined as the analyzation of mass amounts of data and determining the patterns that lie within it. IBM uses four keywords to define big data: volume, variety, velocity and veracity, as shown in figure 1 [2]. Volume represents the scale of the data. Variety represents the different types of data. Velocity represents how fast the data is analyzed. Veracity accounts for the uncertainty of data. This is a common business strategy practiced throughout many industries. For example, in retail big data is used to organize stores in a way that will maximize a customer's purchase. It has even recently been extended to online shopping, recommended products show up on a user's screen based on the items they have looked at or purchased. The store or website analyzes the data from their previous similar customers to be

able to hypothesize about what their future customers will purchase. These techniques can be applied to the health care industry in a similar way. Instead of customers, the data that is analyzed would be that of the patients'.

Approximately 125 million people visit hospitals in the US daily and every one of those patients has medical records [3]. They have tests taken, procedures done and they are diagnosed appropriately. Currently, each patient has an individualized path to treatment and recovery. After the diagnosis is presented, the patient will be provided with a treatment plan that is standard for similar patients with the same diagnosis. Blanket treatments may be useful when dealing with minor issues, but life-threatening diseases must be treated differently.

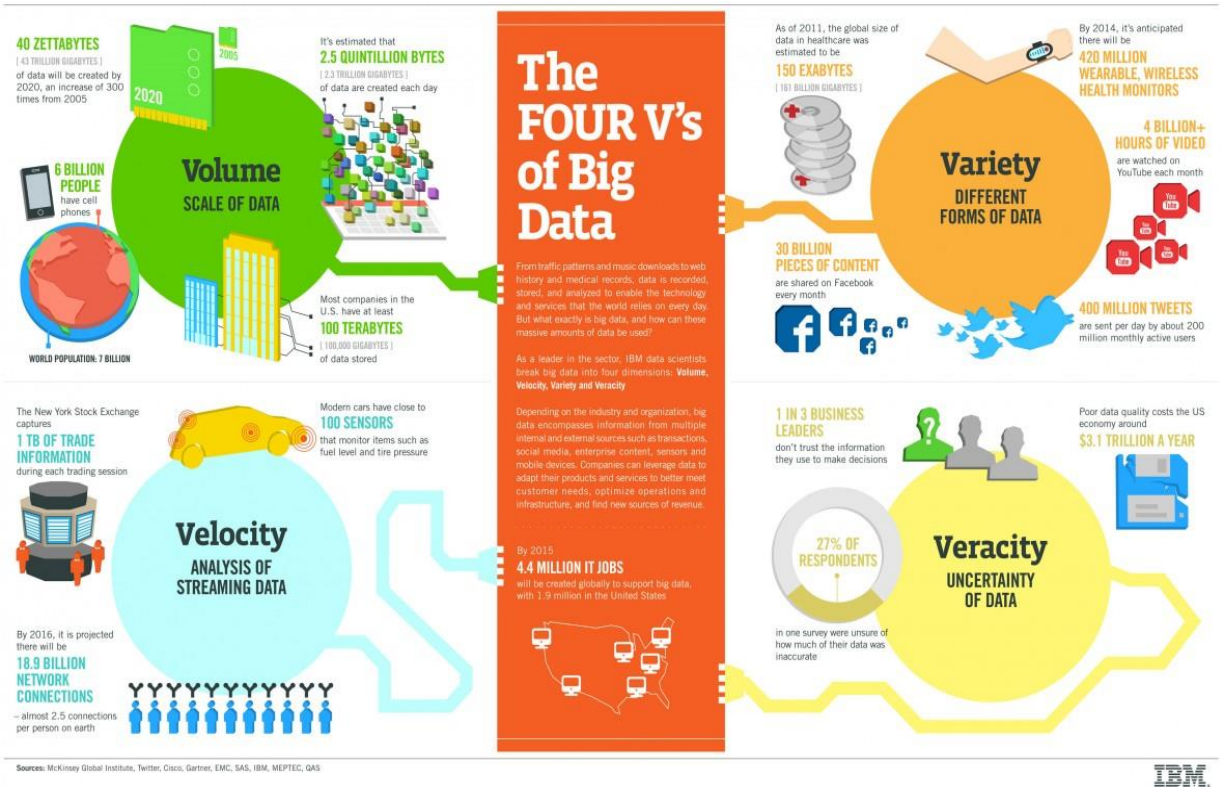


*Figure 1: IBM's definition of Big Data [2]*

Treatments should be personalized to the patient rather than pertaining to the disease. Big data analytics allow doctors to compare patients based on multiple factors; diagnosis, genetic and genomic, physiological and biochemical make-up, age, outcomes, and social data [4]. Considering these factors would make the treatment more personal and in turn- more successful. In 2015, the Chief Information Officer at Beth Israel Deaconess Medical Center in Boston Massachusetts, John Halamka, turned to big data analytics when his wife, Kathy, was diagnosed with stage III breast cancer [5]. Halamka worked on a platform in 2004 called i2b2 (Integrating Biology and the Bedside), which eventually progressed to a big data analytics software called SHRINE (Shared Health Research Information Network). SHRINE was adopted by many Boston hospitals and housed the data of approximately 6.1 million patients.

SHRINE proved to be very useful for the Halamkas, using the software John could compare his wife to patients that were similar to her. He could look at the bigger picture and narrow down which treatments proved to be most efficient and successful. Rather than creating completely new treatments that have never been used before, he manipulated the ones that already existed and chose the best possible treatment plan. Rather than operating right away, which is a common practice, Kathy chose to attack her "estrogen sensitive tumor cells" with a specific chemotherapy regimen first. This shrunk the tumor dramatically before she got a lumpectomy. Following these procedures, she began taking pills to prevent the production of estrogen in her body. This was the best possible outcome for Kathy, without access to SHRINE she probably would have taken a different path with a less ideal outcome. Another issue that John and Kathy faced while dealing with the cancer diagnosis was the difficulty of transferring medical records between hospitals and doctors. With no legal way to communicate about a patient's records at the time, Kathy took matters into her own hands and carried a hard copy of her scans with her on a CD to each of her appointments. This type of situation influenced the development of "Mass HIway" in 2012, which was a secure way to send medical records through a digital database [6]. This would make it significantly easier for patients to take control of their own diagnosis'.

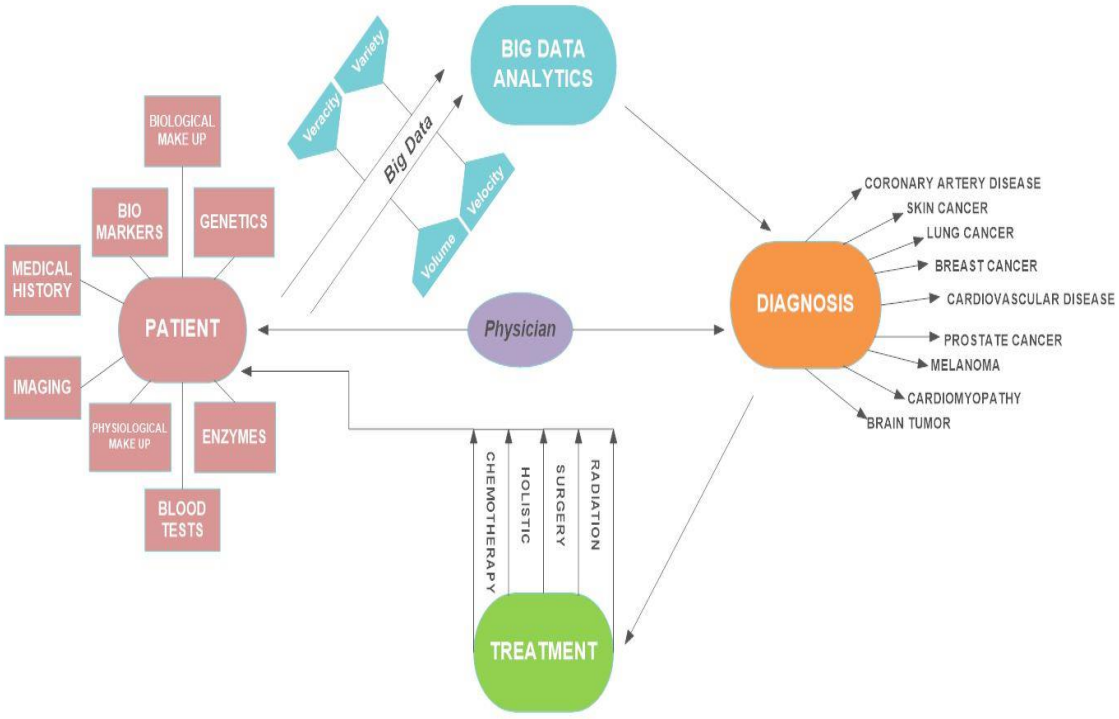


*Figure 2: Flow chart representing big data analytics within healthcare*

John Halamka is a significant advocate of big data analytics within hospitals. While developing both i2b2 and SHRINE he did not receive a lot of support from hospitals and doctors alike. There were many concerns involving patient privacy and competition for data among hospitals. Hospitals did not want to share their data with rival hospitals. Over the last ten years hospitals have started to adapt to the digital world and have opened up to the idea of big data analytics. Today SHRINE is being used by five of Boston's largest and most successful hospitals: Beth

Israel Deaconess Medical Center, Boston Children's Hospital, Brigham and Women's Hospital, Dana- Farber Cancer Institute and Massachusetts General Hospital [7]

Results and Discussion:

Big data analytics benefit both the hospital and the patients. Some of the benefits include improving the accuracy of diagnoses, reducing readmission rates, providing more effective and appropriate treatment plans, reducing costs of healthcare for both the hospital and patient, and decreasing the amount of time spent in a hospital/receiving care. Hospitals across the U.S. have started to adapt big data analytics into their patient care, including Cleveland Clinic, Mass General, Brigham and Women's, Beth Israel Deaconess Medical Center, and Boston Children's Hospital. Big data is still a relatively new concept within the field of healthcare, therefore improvements need to be made for it to work effectively.

One of the biggest battles that big data has to face is patient privacy. The Health Insurance Portability and Accountability Act (HIPAA) ensures data privacy and protects medical information. The act applies to health information from health care providers, health plans, and healthcare clearinghouse and their business associates [8]. It is important to maintain high security around the data and limit the number of personnel that have access to it. Failure to follow the rules and regulations of HIPAA could result in legal and financial consequences for the hospitals. Current big data software, such as SHRINE, is not violate HIPAA. Using "de-identification" to separate medical records from the person that they are attached to make it possible for doctors and analysts to avoid violating patient privacy [8].

SHRINE is now being used at Boston's most influential medical facilities. It is used by hospital faculty to analyze large amounts of data in minimal time. It is ideal for clinical trials, as it can select groups of patients based on specific criteria. It provides a great advantage to doctors while creating research hypotheses. Approved employees of Boston's major hospitals as well Harvard medical students have access to the software through an online database. The database does not display the specific patients, rather the patterns and trends represented by groups of patients. This software makes research more efficient and accurate by providing access to patient data from all five institutes. Apart from SHRINE, there are different variations of big data software throughout the United States. In 2014, google invested $130 million into Flatiron Health, which is a startup company focused on oncology patients specifically. The software is called Flatiron OncologyCloud and the intention is to create the largest database of cancer patients, in order to determine patterns and trends based on specific types of patients or disease. This makes it possible to create ideal treatment that will most effectively treat a patient [9].

Joe Biden is one of the biggest advocates for the use of big data to improve cancer treatments in particular. In December of 2016 congress passed the 21st Century Cures Act, which authorized $1.8 billion of funding to go towards Biden's Cancer Moonshot throughout the next seven years [10]. The goal of Moonshot is to personalize treatments and to find new treatments for various types of cancer. Big data plays a tremendous role in this. Biden's biggest concern is finding a way to most effectively analyze the mass amounts of data that is currently being held in tissue banks, patient registries, electronic medical records and many other databases across the globe

[11]. Over the next seven years there should be tremendous progress made in big data analytics as a result of Moonshot's research and funding.

Memorial Sloan Kettering has implemented IBM's big data program, Watson Health. This software matches Oncology patients with all applicable treatments, ranked in order of successfulness. The patients are compared to like patients, therefore the treatment options are based on the type of person they are combined with their cancer diagnosis. Watson also makes recommendations for treatments or medicines that have not been used yet. Using its knowledge of modern medicine, it matches symptoms with treatments [13]. IBM has expanded Watson Health to over fourteen hospitals nationwide, speeding up treatment processes significantly. The main focus is to reduce the amount of time that physicians have to spend developing a treatment plan, allowing them to see more patients in the same amount of time [14].

Recommendations and Future Work:

Big data analytics software contains endless possibilities to save lives, save money and save time. This is still an up and coming treatment strategy and has not been fully adopted into modern day medical institutions. There is still a lot more ground to cover and a few road blocks to overcome. There are two main issues that need to be addressed: patient privacy and user friendliness. As mentioned earlier HIPAA plays a large role in the protection of patient privacy. This would potentially prevent certain aspects of a patient's data from being put into a database. There are multiple ways around this that still maintain patient privacy. There is one option which involves keeping the data completely anonymous, using characteristics and diagnosis' as the only way of defining each patient. There is also another option which would give each individual patient the opportunity to decide how their medical records are used. Similar to the concept of organ donation- every person would have the right to decide if their records are kept private between themselves and their doctor(s) or if they can be used as part of a database to benefit future patients. With 95% of American adults supporting organ donation, there would most likely be a high percentage of people willing to participate in the use of medical databases [15].

In order for these databases to be most efficient and beneficial it is important to be as user friendly as possible. With millions of patient's data compiled into a single database it will be difficult to sift through all of it. SHRINE uses filters to search for specific characteristics including age, gender, and diagnosis. This is important because it creates a unique assortment of data for every individual patient. SHRINE is mostly focused on using the data for research and clinical trials, rather than creating treatment plans currently. It is vital to make that jump sooner, rather than later. The technology is there, it just needs to be easier to analyze. One possible solution would be searching by "user profile", where a doctor could type in all the characteristics of a patient and see how they compare to previous patients. This would basically create a treatment plan on its own. Using this type of software would enable the user to create a type of portfolio for every individual patient. After all the data is pulled into one place, there should be many different interfaces once the portfolio is created. One possible design is a 3D simulated model of a patient's body, specifically for oncology related cases, that tracks previous patients and where their cancer spread. This would give doctors a heads-up on where to monitor current patients and they would know which areas were high-risk. The software would look similar to the image below. The red areas would represent "high-risk" and the blue areas would represent "low-risk" areas. The other colors would be ranked on a scale. Doctors and researchers would be

able to interact with this body map to see what the most common trends among cancer cases were and how they were most effectively treated.

The most important part of big data analytics is the way that it is utilized. Data is only useful if it is understood properly. By converting it into images, graphs, and tables it becomes easier to interpret. This is a benefit to both the doctors and patients and it will save a tremendous amount of time and money for both. The software could eventually be used as a predictive tool to figure out if/when a patient should be readmitted to the hospital. Readmissions are a huge financial issue for most hospitals. Many times, a patient's hospital stay is too extensive or they are readmitted too many times. This is costly to both the patient and the hospital. With the proper use of big data analytics in a hospital setting this situation can be reduced.

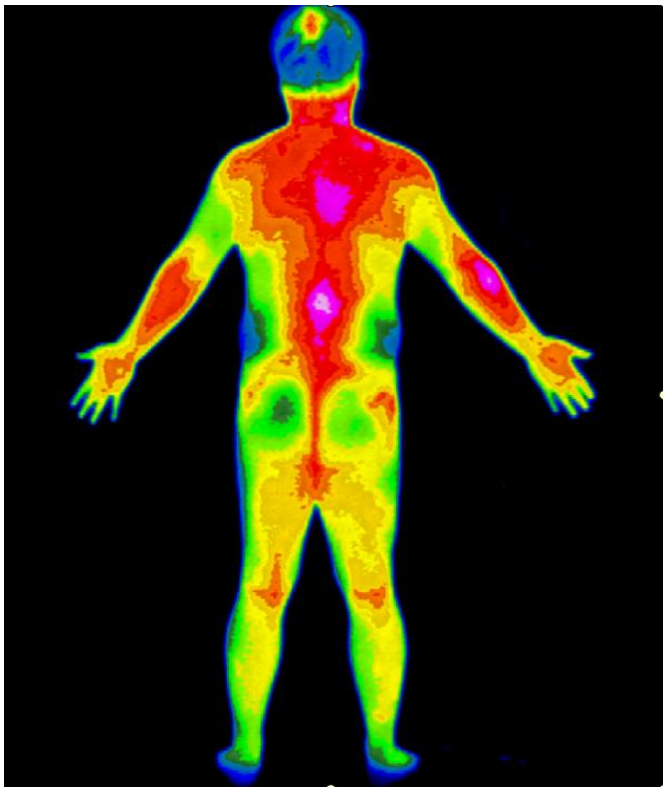

*Figure 3: Example of a potential "heat map" for predicting cancer spread [16]*

Conclusion:

Big data analytics have proved to be very successful throughout a variety of industries. This strategy has recently been implemented into the medical industry and is still not fully adopted into hospitals and medical facilities across the United States. Start-up companies and college research teams make up the bulk of medical big data analytics software right now. Boston is one of the first cities to plunge into the world of big data, by executing SHRINE in their main hospitals.

Over the upcoming years big data analytics will play a more impacting role in medicine and treatment. In the same way that it has changed so many other industries, it will reshape the medical industry. It will allow doctors to treat more patients and it will create stronger patient/doctor relationships. It will give patients the ability to take control of their own treatment and decide which treatment plan will work best for their specific circumstance. Big data analytics will also improve the survival rates for deadly diseases like cancer, heart disease, diabetes, Alzheimer's, etc. Ultimately, this will save money, save time and save lives.

References:


1. Cdc.gov. (2018). “FastStats”. <https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm> (10 Apr. 2018).
2. IBM Big Data & Analytics Hub. (2018). “The Four Vs of Big Data.” <http://www.ibmbigdatahub.com/infographic/four-vs-big-data> (10 Apr. 2018).
3. Coe.neu.edu. (2018). <https://www.coe.neu.edu/healthcare/pdfs/publications/Health%20Aff-2014 Bates_Using%20Predictive%20Analytics.pdf> (10 Apr. 2018).
4. IEEE Spectrum: Technology, Engineering, and Science News. (2018). “Big Data Beats Cancer.” < https://spectrum.ieee.org/biomedical/diagnostics/big-data-beats-cancer> (10 Apr. 2018).
5. 6. Masshiway.net. (2018). “Massachusetts Health Information HIway - Home.” <http://www.masshiway.net/HPP/index.htm> (10 Apr. 2018).
6. Anon, (2018). <https://catalyst.harvard.edu/services/shrine/> (10 Apr. 2018).
7. Agg.com. (2018). “Big Data Analytics Under HIPAA” <http://www.agg.com/Big-Data-Analytics-Under-HIPAA-03-17-2016/> (10 Apr. 2018).
8. Datanami. (2018). “Flatiron Health Archives -Datanami.” <https://www.datanami.com/vendor/flatiron-health/> (10 Apr. 2018).
9. National Cancer Institute. (2018). “Cancer Moonshot” <https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative> (10 Apr. 2018).
10. Forbes.com. (2018). “Forbes Welcome.” <https://www.forbes.com/sites/arleneweintraub/2016/01/26/here-are-the-sticking-points-for-biden-big-data-and-cancer-moonshots/#44df3b7330c3> (10 Apr. 2018).
11. IBM Watson Health. (2018). “IBM Watson Health - IBM Watson for Oncology” <https://www.ibm.com/watson/health/oncology-and-genomics/oncology/> ( 10 Apr. 2018).
12. Organdonor.gov. (2018). “Organ Donation Statistics: Why be an Organ Donor?” <https://www.organdonor.gov/statistics-stories/statistics.html> (10 Apr. 2018).
13. What are areas of the body which lose heat more quickly (2018) <https://outdoors.stackexchange.com/questions/681/what-are-areas-of-the-body-which-lose-heat-more-quickly-and-how-can-i-reduce-th> (10 Apr. 2018).