International Journal of Knowledge Management and Practices
7 (1) 2019, 46-54
http://publishingindia.com/ijkmp/

BIG DATA ANALYTICS IN HEALTHCARE: CHALLENGES AND POSSIBILITIES

Manoj Kumar Jena*, Brajaballav Kar**

**KIIT School of Management, Bhubaneswar, Odisha, India. Email: manoj.jena@ksom.ac.in*
***KIIT School of Management, KIIT University, Bhubaneswar, Odisha, India. Email: braja.kar@ksom.ac.in*

Abstract *With the improvement in technology, the data acquisition capability, its storage and analysis have gone up manifold. However, the costs related to such activities have gone down significantly. This leads to the abundance of data everywhere. Notwithstanding the data omnipresence, users process data selectively, take intuitive decisions, work under both data-overload and data-poverty.*

Arguably, the healthcare sector has enormous challenges in achieving universal basic health services, providing safe, effective, affordable and timely intervention for patients. The sector involves communication between various stakeholders such as government, healthcare institutions, doctors, patients, and insurance companies. This byzantine, recursive, and helical process generates enormous data.

In healthcare practice one of the most important issues is that the big data creation is not purposive; availability of data may not be for the purpose for its use. Secondly, if less data is equally sufficient to make a good quality decision, then big data may bring in confusion. Thirdly, at a conceptual level, statistics relies on effective sampling methods to generalize and predict about population. Big data analytics also uses above principles, indicating that the availability of data alone is not enough. Another daunting challenge is data ownership; data gatherers' claim of exclusivity is unethical which raises questions about privacy, user rights, and public ownership etc. From the business model perspective, the data creation becomes free and thus marginal propensity to consume increases exponentially.

This paper takes a contrarian approach and presents the challenges and critics towards the implementation of big data processes in healthcare.

Keywords: *Big Data, Healthcare, Analytics, Patient-Centric, Governance, Signal to Noise*

Introduction

Data has played a major role in healthcare and related domain. It dates back to 1854 when Florence Nightingale was engaged in providing various nursing and healthcare services to the British troops during the war with Turkey. She observed that improvement in cleanliness and ambiance conditions prior to the admission of soldiers to the hospital played a major role in preventing thousands of deaths. She recorded various information in the form of tables and charts. She named the display a "Coxcomb"; later known as pie-chart (Sheingold & Hahn, 2014).

In the 17th and 18th century, the diagnosis of a disease depended on the patient's narrative of the symptoms and illness as directed by questions from the physician. The physician mainly focussed on the patient's external appearance, expression, breathing manners and the sense of touch to evaluate illness (Reiser & Reiser, 1981). Various necessities and constraints during the diagnostic procedure led to multiple innovations with time. The stethoscope, the thermometer, the microscope, electrocardiograph, and many other electronic devices were invented during the 18th to 19th century. With these innovations, the objective evidence or the judgment based on data from various laboratory procedures and technical devices gradually started to replace the earlier method of judgment based on solely subjective evidence.

Data Generation Technologies, Sources, and Types of Health Care Data

The goal of big data analytics in healthcare system would be to shift the population-based approach to individualized medicine system. Here population-based medicine refers to the "statistically approved evidence-based medicine based on large-scale randomized, double-blind placebo-controlled clinical trials performed under most rigorous conditions". Individualized medicine and practice system is based on individual data, which would play a crucial role for treatment. Topol, in his book, "The creative destruction of medicine" argues that the genome science and digital technologies are being considered as the driving force to draw such transformation (Topol & Hill, 2012). The basis of his argument is the interconnectivity of mobiles and personal computers with web and cloud computing system. Although this is debatable at various stages of application. Research on American use of the internet for health-related purpose indicates that only a few people use digital technologies to get information, communicate with health personnel, or make online medical purchases (Miller & West, 2009). However, the disruption through digital technology applications indicates a data-driven culture.

Type of Data Generated by the Healthcare System (Structured/Unstructured Formats)

Healthcare data are always diverse and complex. From start to end, an individual goes through multiple procedures and departments for getting healthcare services. Therefore, data related to various diagnosis and communications appear to be fragmented.

Healthcare Data Resides in Multiple Places

According to (Belle et al., 2015) healthcare data are spread among different healthcare systems, health insurers, researchers, and government entities. Further, for an individual, the healthcare data comes from all over the organization. Starting from entering demographic details to collecting electronic medical records (EMR) from different departments like radiology, pharmacy, etc., healthcare data represents huge variability. The data occurs in various formats such as texts, numeric, images, digital, video, multimedia, etc. Doctor's prescriptions are in paper format, radiology use image format, and EMRs hold data in form of rows of textual and numeric data.

Primary sources of healthcare data include (a) Claims (Medical claim, prescription claims) (b) Clinical Records (electronic medical records, diagnostics, Registers, admission, discharge and transfers, etc.) (c) Sales and Dispensing (Purchases of drugs and dispensed from hospitals, etc. (d) Clinical research and (e) Patient-generated data through Social media and devices (Szlezák, Evers, Wang, & Pérez, 2014)

Healthcare Data are Unstructured

Although EMR software provides a platform to record data in a consistent manner, in actual scenario data capture shows nowhere any consistency. Healthcare providers would adopt their own formats and are reluctant to follow any standardized documentation. Thus, the data capture is very much unstructured, and would not be suitable for analytics.

A very frequent observation in healthcare systems is that the data have an inconsistent variable definition for same type test or diagnosis. For example, one group of clinicians may define a microalbumin test in certain units and standards than another group. Even two clinicians may define the criteria for diabetics in two different ways. There may not even be a level of consensus about a particular treatment among a group of clinicians. Further, evidence-based practice and new research make the above condition more volatile and varying.

Apart from the genetic information and research data, most of the healthcare data are unstructured. Hon S. Pak (2018) observed that nearly 80% of clinical information in electronic health records (EHR) is unstructured and those are not usable by the health information system. Although these data are converted to some structured format the information is relatively much limited (Ward & Barker, 2013).

Healthcare Data will Only get More Complex

The number of variables for measuring any healthcare related issues are increasing with the progress of technology and innovation. In the case of healthcare there is no finite number of predefined sets or parts for which data are collected for future analysis, rather the data must refer to the complex web of an individual system of the human body. Even the dependencies of every single organ with another remains unpredictable till date. The technological improvement with the exploration of single details makes the existing data redundant. Thus, managing the data related to human systems and turning it to useful information across a population requires far more sophisticated tools and analysis compared to that used in manufacturing industries.

As discussed earlier the data availability and ownership is extremely fragmented among various players starting from healthcare providers such as medical institutions and doctors, patients themselves, health insurers, pharma manufacturers, government, and other technology players. Data are split across multiple formats, systems and are software specific. The structured and unstructured data, inconsistency, variability and complexity of data with frequently changing regulatory environment and standards make it extremely difficult to create a standard solution on a global space. Data integration remains a challenge. Research on exploring analytical opportunities for modeling healthcare data observe the lack of technology for processing large, complex and heterogeneous data (Dinov, 2016).

The Volume of Data Generated-Big Data

The term big data has become ubiquitous. There is no single definition of big data. Various stakeholders give different and often contradictory explanations. The lack of a consistent definition introduces ambiguity in understanding ("Unstructured Data in Healthcare," n.d.), however big data can be thought of as a large set of data which can see a significant application of computational power to interpret and predict further information.

From birth certificate until death certificate, a person accumulates healthcare data. For every visit to a healthcare professional, some information is recorded. A study indicates that for the year 2011, the U.S. healthcare systems generated 150 exabytes of data (One exabyte is equal to one billion gigabytes) ("Where Healthcare's Big Data Actually Comes From" 2018). The human mind cannot, of course, guess or have a representative picture of such huge data. The following section discusses the types of such data and its collection methods.

Starting from various corporates, social media, internets, mobile apps there could be innumerable sources generating and collecting data. Some of the important big data sources for healthcare are worth noting.

The internet of things (IoT): Internet of things represents various computing devices embedded in everyday objects and connected to internet enabling the devices to send and receive data. Nowadays the IoT platforms are the cheapest and cost-effective options to generate data. *Wearables* such as headbands and wristbands, etc. are primary tools, (e.g., FitBit allows individuals to track blood pressure, heart rate, stress level, weight, and other activities). Similarly, *Mobiles apps* in smartphones track user's activities such as steps walked, calories consumed, amount and quality of sleep, and exercise intensities. *Medical devices and sensors* that can send data to the cloud are also used to collect individual information. Some of these medical devices are glucose monitors, pulse oximeters, blood pressure monitors, proximity sensors,

Electronic Medical Records (EMR): EMRs are the digital equivalent of paper records or charts at a clinician office. It typically contains general information such as treatment and medical history of a patient as collected by the individual medical practitioners. It has been proposed that one universal EMR record that can be accessed from any healthcare facility using EMR software instead of having different charts available with different healthcare institutions. EMR plays a major role these days in transforming healthcare quality through the implementation of information technology.

Electronic Health Records (EHR): EHR is also a digital version of a patient's records, while this includes more information about the patient's medical history. EMR provides a snapshot view of a patient's medical history while EHR is a more comprehensive report of the individual's overall health. EHRs are designed to share medical information of the patient with authorized providers and staff across more than one organization. Adoption of EMR and EHRs need significant investment in IT infrastructure, policy and regulations, research and development, and education.

Other clinical data: Other than the digital records, physicians written notes during the process of diagnosis can provide a significant amount of unstructured data. These type data are very much fragmented and needs a lot of hardware infrastructure to process them into information. Prescriptions, laboratory, pharmacy, and other administrative data generated from clinical decision support system can augment to the volume of such data repository.

Insurance providers: There are several insurance providers which record the information about the individual related to her health and claims. These include both private and government health plan claims and pharmacy claims.

Social Media: It is believed that social media posts, such as Twitter feeds, blogs, status updates on Facebook, etc., can reflect and provide evidence about a person's health and state of mind.

Other than above data collection handles, there are other organizations engaged in Genome research studies and other research studies that collect participant data.

Institution and Individual Interactions/Participants

Patient-centeredness is a common approach to improve healthcare. Saha, Beach, & Cooper (2008) studied the "cultural competence and patient centeredness" in development of the healthcare system. As per the study, each approach holds promise for improving the quality of healthcare for individual patients, communities, and population. Patient-centeredness focuses on the interaction between the patient and healthcare providers. An indicative representation of interactions for the patients and other different stakeholders is indicated in Fig. 1.

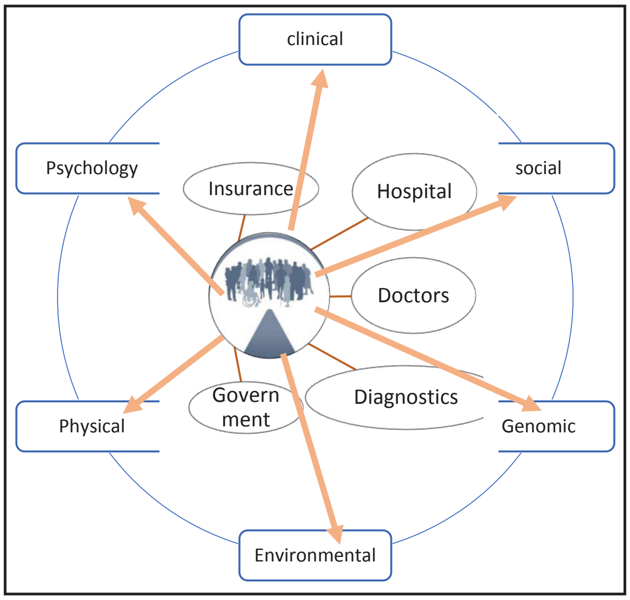

Fig. 1: A Patient-Health System Interaction Model

Patient-Centred care tries to balance the power between patient and care providers. Disparities in clinical decisions could be reduced by increasing patient involvement. The patient's diagnosis may not only be linked to health issues. At the immediate circle, the patient interacts with the stakeholders such as Hospitals, doctors, diagnostic centers, insurance providers, and government for getting healthcare benefits. However, the healthcare need is driven by various peripheral factors such as; a patient's social and psychological condition, physical situation, environmental factors, genomic, and clinical requirements. Therefore, a logical next step in the patient-centric model would be to include individual needs/factors in tandem with institutional communication center. A schematic representation of the said model is shown in Fig. 1. A patient's physical symptoms may be normal, but the psychological and environmental factors may have a significant influence on the overall health. Therefore, the above dimensions and their impact on individual healthcare solutions need to be taken care of while structuring the data for future uses. This indicates a significant challenge in gathering the right information and creating the right sync of repository in place while considering the holistic need of the patient.

Information Usage Pattern

Doctor-Patient Communication and Challenges

It is a common conception that patients themselves are the best agent to communicate the issues to doctors. Study has found that personal collections remain the preferred information resource for doctors, electronic sources rank second (Bryant, 2004). A patient's satisfaction in communicating the scenario depends on the doctor's attention and vice versa. The communication is mostly controlled by doctors through initial questions, intermittent questions between patients' explanations, and neglect of patients' "life world". Research indicates that the transformation of information is related to patients' characteristics such as; sex, age, education, social class, and prognosis; it is also related to doctors' qualities such as; social background, income, and perception of patients' desire for information, and the clinical situation such as the number of patients seen (Waitzkin, 1984).

Multiple types of research have demonstrated that there is a correlation between effective doctor-patient communication and improved patient health outcomes (Stewart, 1995). Studies further find that most of the time complaints about doctors are due to patient-clinician communication problem and not the technical competencies issues. The patient-doctor conversation quality also impacts the patient's adherence to the treatment regimens (Stewart et al., 1999). Poor communication between patient and doctors could be a major challenge in collecting exact information that is expected.

The traditional system of a close, long-term relationship with family-doctors is being replaced with a short-term encounter with specialists. No longer, the personal rapport with the patient is given high importance. A study in the pediatric clinic of a large hospital indicates that the physician often talks jargons or seems not to fully heed the patients' concerns. Mutual dissatisfaction is the outcome (Korsch & Negrete, 1972). This also puts a major challenge in the collection of the right information in the process.

Information Transfer Challenges Across Organizations/Departments

Apart from the doctor-patient communication in a clinic, the data exchange between hospitals, general practitioners, and pharmacy are quite frequent in case of healthcare services. From big data perspective, the information interchange and interpretation are more relevant as the aggregate data has to be ultimately analyzed for generalization and uses.

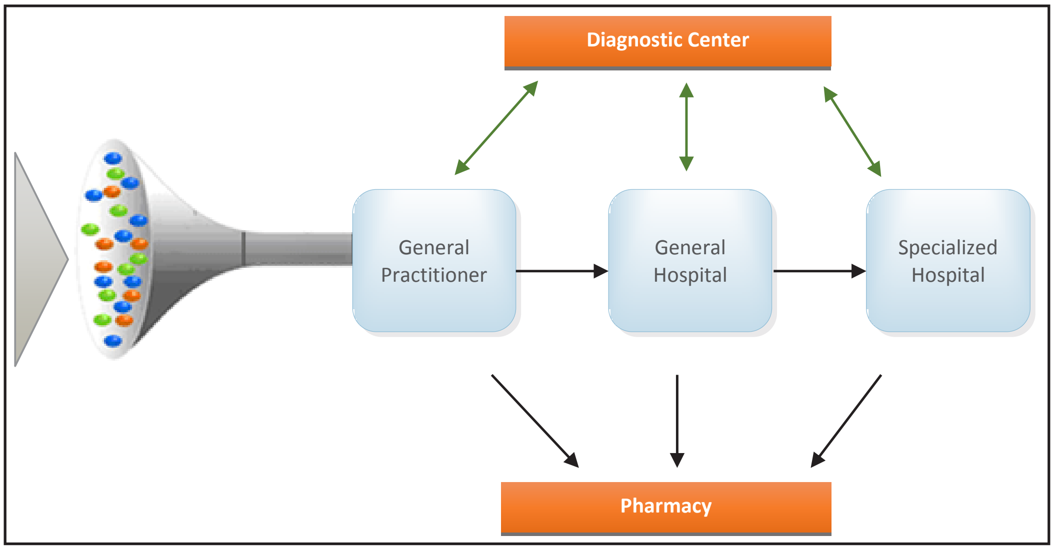

Fig. 2: A Patient-Walk Model from a General Practitioner to a Specialized Clinic

Figure-2 represents a patient-walk model, describing a patient's possible visit to a different institution for treatment and cure. Normally a patient collects ideas from friends, family, and her immediate network before visiting a general practitioner. She goes for diagnosis and finally gets the medicine from pharmacies for minimal treatment. In case this is not adequate, she visits the hospital and repeats the same process. For complicated scenarios, she further gets landed up in specialized healthcare centers and get the treatment done. Therefore, at every subsequent stage, it needs travel and exchange of data. The prescription or the discharge certificate at each stage carries the major information for the next contact person.

Study on discharge summary reveals that the deficit in communication and information transfer at hospital discharge is common (Hellesø, Lorensen, & Sorensen, 2004; Wilson, Ruscoe, Chapman, & Miller, 2001). The availability of such report at the first discharge is low (12%-34%). Discharge summaries often lacked important information such as diagnostic test results (33%-63% missing), treatment or hospital course (7%-22%), discharge medications (2%-40%), test results pending at discharge (65%), patient or family counselling (90%-92%), and follow-up plans (2%-43%) (Kripalani et al., 2007).

Healthcare Challenges in Big Data Implementation

Organizational Barrier

The evolution of big data and its implementation in healthcare revolves around the technological advancement and mutual agreement over organizational standards. Health Information Technology (HIT) consists of innovations such as the electronic health records (EHR), e-prescriptions, computerized provider order entry (CPOE), picture archiving and communication system (PACS), online access of medical journals, video conferencing and feedback system. HIT implementation is expected to provide more responsive care to patients with better quality and safer standards.

However, adoption of HIT has been a question mark due to the organization's inertia and technical ineffectiveness to implement the change. Research finds that technical factors explain 5% of HIT failure; some estimate it as 20% (Westbrook, Braithwaite, Iedema, & Coiera, 2004). Similar to other organizations, the decision support system in healthcare enterprises face significant inertia to adopt the change. In fact, the introduction of HIT should not be viewed as a problem in technology exclusively but rather as a problem in organizational change (Wears & Berg, 2005; Lluch, 2011).

Transition from Paper-Based System to EMRs

As the cost of storage has dramatically decreased over the past decades (hard drive cost per gigabyte reduced from around $1 mn. in 1980s to $0.1 in 2015 (Komorowski, 2014; Mearian, 2014), big data applications are trending more with time. However, healthcare systems are unique in their inherent complexity. Each department within a healthcare industry has different needs, and they follow even different standards pertaining to their needs. Lack of standardization makes it impossible to automate the decision support system and generate a common data hub usable for the entire organization. Some researchers believe that data collected in developing countries are incomplete, inaccurate, unreliable and not timely, which poses a further challenge to the potential of EMR systems (Rovner, 1991).

Due to the above complexity, silo organizational structure, cost, and unavailability of adequate technical competencies, conversion of traditional paper-based records to EMRs is a major challenge in the healthcare system. The transition represents a paradigm shift for the work of the physician and other staffs. This needs extremely systematic activity, and also should be managed from all aspects such as clinical, administrative, cultural, environmental, and organizational need.

Issues of Big Data

Signal to Noise Ratio

A simple laboratory test ordered by a clinician can involve a web of communication channels and this further involves data generation at each stage. Different healthcare professionals are involved in managing a single patient for a simple application. The communication channel gets more complicated by the involvement of insurance players, and other 3rd party healthcare providers indicating the stack of huge fragmented data (Coiera, 2006).

Every communication channel is associated with signal and noise. In case of healthcare; usage of various formats, language, and varying verbal interpretations add noise to the information. Further, the number of possible conversations during each interaction stage between stakeholders increases combinatorially. This depends on the number of individuals engaged in the conversation at a time.

Given there are three players in a one-to-one conversation such as doctors, patient, and nurse there will be three different communication channels and three sets of data. As this increase to 5, the number of channels increases to 10, and this is for a single patient case. For n number of patients/individuals and given r number of people conversing with each other at a time the total number of the possible channel will be given by the following equation.

Number of conversations = $n! / ((r! \ (n - r)!))$

Thus, the channels of communication increase significantly for every addition of a stakeholder. The complexity of these communication networks in case of the healthcare system only augments the noise in the existing limited information.

Current Techniques (Statistical Analysis)

As stated earlier, Florence Nightingale used descriptive statistic to examine the healthcare data through charts and table. Using simple graphical methods, she was able to discover the causes of death in hospitals during the Crimean war (Faltin, Kenett, & Ruggeri, 2012). Starting from pharmaceutical to health economics, development of new drugs, clinical outcomes to EMRs, risk assessment in organ allocation, statistics find its role in every corner of healthcare. The use of statistical methods and concepts in medical journals has increased significantly from 1978-79 to 2004. There has been a shift in the use of descriptive measures such as means, percentage, and variance to the use of contingency tables and statistical power calculations (Bailar & Hoaglin, 2012).

Statistics help through various measurement tools such as data type, average, variability, correlation etc., to gather information from surveys. Quantitative research plays a major role in modeling forecasting details for healthcare needs. Randomize control trials are the backbone of healthcare predictions, which is only possible by statistics. Statistical process control (SPC) charts help to analyze whether any change in processes makes any real difference in outcomes. Repeated measurement of the same parameter often yields different results due to the fluctuations in patients' biological process. SPC helps to interpret the amount of variability in any measurement process and its impact on outcome (Benneyan, Llyod, & Plsek, 2003). Factor analysis is used to construct tests, develop instruments, and check the reliability and structure of existing instruments in healthcare studies (Pett, Lackey, & Sullivan, 2006). Non-parametric statistical tests are quite useful in case of healthcare science where unordered categories and unequal sample size exist in the data (Pett, 2016).

Governance Challenges (Privacy, Security, Public Nature and Usage Who Collects, Who Uses and for What Purpose?

Security and privacy in healthcare: With the explosion of the Internet of Things (IoT) and its ability to deal with real-time monitoring of health data, organizations are gearing up to collect such information in the form of big-data. Gartner estimates 26 billion IoT devices will be functional by 2020 and the amount of traffic generated by such devices will be large enough to place it in the category of big data (Middleton, Kjeldsen, & Tully, 2013). IoT devices through various sensors can monitor and generate health data and can provide real-time data to clinicians over networks. However; data over internet or networks poses risk to an external threat. This significantly increases the concerns regarding security and privacy of patients. Patient privacy is becoming a growing concern as many incidents related to the breaching of privacy are reported over the past decade. A report in Forbes magazine mentioned that Target Corporation sent baby care coupons to a teenage girl before her parents know about the situation (Hill, 2012). A 2018 Verizon data breach investigation report stated that in 2017 out of 53000 reported security incidents, 2216 confirmed data breaches were observed (Gabriel et al., 2018). Further, an annual benchmark study on privacy and security of healthcare data for 2015, reveals that nearly 90% of the organizations in the U.S. had at least one data breach in the past two years (Ponemon.org, 2018). The research also found that criminal attacks are the leading cause of data breaches in healthcare.

Data Infrastructure Challenge

Every successive year different new type threats are emerging, and security systems are being more vulnerable. In the case of healthcare big data, the data sources are extremely diversified and thereby it poses a bigger challenge for the data inflow with respect to the security concerns. A big data healthcare cloud repository would consist of various information from patients' interactions with different peripheral factors. A schematic notation is given in the Fig. 3.

The inherent network design and varied source of data make big data servers entirely different from the existing traditional servers. Big data platforms (also referred to as NoSQL or NewSQL databases) deals with the unstructured information pulled from multiple sources and in multiple formats.

This needs huge hardware cost and strategic spending on IT infrastructure. Thus, the organizations need to estimate the trade-offs between the spending and its probable benefits. As Michael Passe, storage architect for Beth Israel Deaconess Medical Center (BIDMC) based in Boston, says *"We're just starting to figure out how to use it and what makes sense for us, and then trying to figure out how we best posture ourselves from an infrastructure standpoint to support it,"* ("How big data servers are different from regular dedicated hosting servers," 2018).

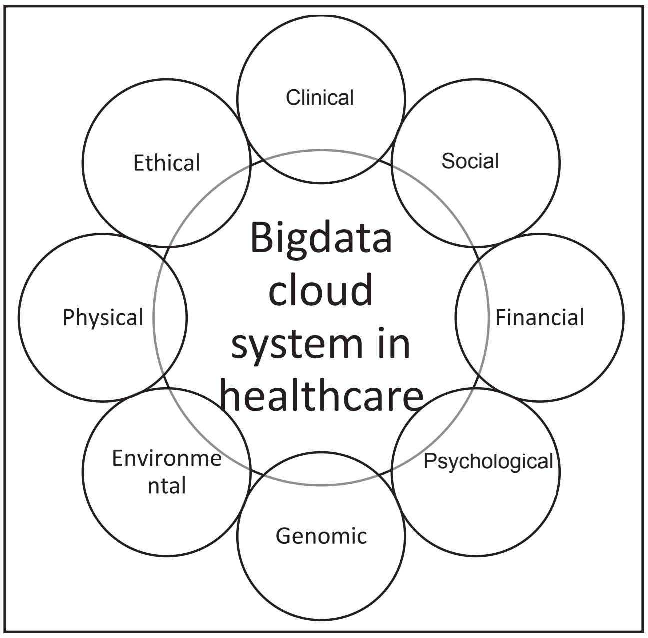

Fig. 3: A Big Data Cloud System in Healthcare

Discussion and Conclusion

Big data literally means more data from different sources but at the same time, it comes with noise. As the data collection and storage has become much cheaper, everyone is interested in data. However, an infinitely large data set will not provide an answer to the questions for which specific data are not collected. Healthcare systems should not confuse more data with more insight. A successful analysis rather tries to answer the questions such as: what is known in specific about the patient? What is known about the population? Most healthcare system can do adequate analytics and reporting with the traditional relational database; using these databases could be more valuable than worrying about big data ("Big Data in Healthcare Made Simple," 2018). An infinite data set or big data can definitely demonstrate the presence of an association between two variables, but the question of cause and effect remains unanswered in the analysis. World Health Organisation defines health as a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." So the concept of health is multi-dimensional and individual oriented. So from a population-based information to individual based medicine through digital technologies and usage of big data to collect population data represents a diametrically opposite viewpoint. Similarly, from the patient perspective, individual patients and their technology usage pattern are unique; large data on a single patient is time-bound and thus may not improve the efficacy of the treatment. There are also challenges related to medical and test practices which differ according to the practitioners and institutions, thus the acceptability of the information is not universal. It is indicated that social media posts may indicate the status of health and mind; however, such posts have cultural nuances and variability. The inferences may not be uniform across different social media users (nationality, culture, gender, age groups, etc.).

The value addition and the cost of implementation for big-data analytics at an organizational level needs to be evaluated. Does generalization indicate better solutions compared to subjective analysis in healthcare domain? Interoperability of EHRs and subsequently the high noise-to-signal ratio in a huge scale of unstructured data in healthcare poses a major challenge for the implementation of prescriptive and predictive analytics. The progress, ubiquity, and interconnectedness of technology pose a serious concern on patient's privacy.

However; given all limitation, it is being observed that machine learning applications can actively help in certain image analysis for better diagnosis. Microsoft's InnerEye initiative (started in 2010) is presently working on image diagnostic tools to interpret the brain images in a more accurate way ("Machine Learning Healthcare Applications – 2018 and Beyond," 2018).

Nevertheless, usage of technology and information modifies our understanding of a process significantly thereby creating new usages. Big Data analytics has such a possibility. Its usage in generating population-based information for policy-making and health governance also has significant implications.

References

Bailar, J. C., & Hoaglin, D. C. (2012). *Medical uses of statistics.* Hoboken, NJ: John Wiley & Sons.

Belle, A., Thiagarajan, R., Soroushmehr, S. M., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big data analytics in healthcare. *BioMed Research International*, 1–16. doi:10.1155/2015/370194

Benneyan, J. C., Lloyd, R. C., & Plsek, P. E. (2003). Statistical process control as a tool for research and healthcare improvement. *BMJ Quality & Safety*, *12*(6), 458–464.

Big Data in Healthcare Made Simple. (2018, October 17). Retrieved December 10, 2018 from https://www.healthcatalyst.com/big-data-in-healthcare-made-simple

Bryant, S. L. (2004). The information needs and information seeking behaviour of family doctors. *Health Information & Libraries Journal*, *21*(2), 84–93. doi:10.1111/j.1471-1842.2004.00490.x

Coiera, E. (2006). Communication systems in healthcare. *Clinical Biochemist Reviews*, *27*(2), 89.

Dinov, I. D. (2016). Methodological challenges and analytic opportunities for modelling and interpreting Big Healthcare Data. *Giga Science*, *5*(1), 12. doi:10.1186/s13742-016-0117-6

Faltin, F., Kenett, R. S., & Ruggeri, F. (Eds.). (2012). *Statistical methods in healthcare.* West Sussex, UK: John Wiley & Sons.

Hellesø, R., Lorensen, M., & Sorensen, L. (2004). Challenging the information gap – The patients transfer from hospital to home health care. *International Journal of Medical Informatics*, *73*(7–8), 569–580. doi:10.1016/j.ijmedinf.2004.04.009

Hill, K. (2012). *How Target figured out a teen girl was pregnant before her father did*. Forbes, Inc.

How big data servers are different from regular dedicated hosting servers. (2018, October 3). Retrieved December 10, 2018 from http://bigdata-madesimple.com/how-big-data-servers-are-different-from-regular-dedicated-hosting-servers/

Komorowski, M. (2014, March 9). Retrieved December 2, 2018 from http://www.mkomo.com/cost-per-gigabyte-update

Korsch, B. M., & Negrete, V. F. (1972). Doctor-patient communication. *Scientific American*, *227*(2), 66–74. doi:10.1038/scientificamerican0872-66

Kripalani, S., LeFevre, F., Phillips, C. O., Williams, M. V., Basaviah, P., & Baker, D. W. (2007). Deficits in communication and information transfer between hospital-based and primary care physicians. *JAMA*, *297*(8), 831–841. doi:10.1001/jama.297.8.831

Lluch, M. (2011). Healthcare professionals' organisational barriers to health information technologies – A literature review. *International Journal of Medical Informatics*, *80*(12), 849–862. doi:10.1016/j.ijmedinf.2011.09.005

Machine Learning Healthcare Applications – 2018 and Beyond. (2018, September 16). Retrieved December 10, 2018 from https://www.techemergence.com/machine-learning-healthcare-applications/

Mearian, L. (2014, March 14). *Data storage – Then and now*. Retrieved December 2, 2018 from https://www.computerworld.com/article/2473980/data-storage-solutions/data-storage-solutions-143723-storage-now-and-then.html

Middleton, B. (2005). Achieving U.S. health information technology adoption: The need for a third hand. *Health Affairs*, *24*(5), 1269–1272. doi:10.1377/hlthaff.24.5.1269

Middleton, P., Kjeldsen, P., & Tully, J. (2013). *Forecast: The internet of things.* Worldwide. Gartner.

Miller, E. A., & West, D. M. (2009). Where's the revolution? Digital technology and health care in the internet age. *Journal of Health Politics, Policy and Law*, *34*(2), 261–284. doi:10.1215/03616878-2008-046

Pett, M. A. (2016). *Nonparametric statistics for health care research: Statistics for small samples and unusual distributions.* Thousand Oaks, CA: Sage.

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2006). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research.* Thousand Oaks, CA: Sage.

Ponemon.org. (2018). Retrieved August 6, 2018 from https://www.ponemon.org/local/upload/file/Sixth%20Annual%20Patient%20Privacy%20%26%20Data%20Security%20Report%20FINAL%206.pdf

Reiser, S. J., & Reiser, S. J. (1981). *Medicine and the reign of technology.* Cambridge: Cambridge University Press.

Rovner, D. R. (1991). Medical informatics, computer applications in health care. Edited by Edward H. Shortliffe, Leslie E. Perreault, Gio Wiederhold, and Lawrence M. Fagan, Reading, MA: Addison-Wesley, 1990. *Medical Decision Making*, *11*(2), 148–148. doi:10.1177/0272989x9101100215

Saha, S., Beach, M. C., & Cooper, L. A. (2008). Patient centeredness, cultural competence and healthcare quality. *Journal of the National Medical Association*, *100*(11), 1275–1285. doi:10.1016/s0027-9684(15)31505-4

Sheingold, B. H., & Hahn, J. A. (2014). The history of healthcare quality: The first 100 years 1860–1960. *International Journal of Africa Nursing Sciences*, *1*, 18–22. doi:10.1016/j.ijans.2014.05.002

Stewart, M. A. (1995). Effective physician-patient communication and health outcomes: A review. *CMAJ: Canadian Medical Association Journal*, *152*(9), 1423.

Stewart, M. A., Brown, J. B., Boon, H., Galajda, J., Meredith, L., & Sangster, M. (1999). Evidence on patient-doctor communication. *Cancer*, *25*(1999), 30.

Szlezák, N., Evers, M., Wang, J., & Pérez, L. (2014). The role of big data and advanced analytics in drug discovery, development, and commercialization. *Clinical Pharmacology & Therapeutics*, *95*(5), 492–495. doi:10.1038/clpt.2014.29

Topol, E. J., & Hill, D. (2012). *The creative destruction of medicine: How the digital revolution will create better health care* (p. 2). New York, NY: Basic Books.

Unstructured Data In Healthcare. (n.d.). Retrieved December 5, 2018 from https://artificial-intelligence.healthcaretechoutlook.com/cxoinsights/unstructured-data-in-healthcare-nid-506.html

Waitzkin, H. (1984). Doctor-Patient Communication. *JAMA*, *252*(17), 2441–2446. doi:10.1001/jama.1984.03350170043017

Ward, J. S., & Barker, A. (2013). *Undefined by data: A survey of big data definitions*. Retrieved December 20, 2018 from https://arxiv.org/pdf/1309.5821.pdf

Wears, R. L., & Berg, M. (2005). Computer technology and clinical work. *JAMA*, *293*(10), 1261–1263. doi:10.1001/jama.293.10.1261

Westbrook, J. I., Braithwaite, J., Iedema, R., & Coiera, E. W. (2004, September). Evaluating the impact of information communication technologies on complex organizational systems: A multi-disciplinary, multi-method framework. In *Medinfo* (1323–1327).

Where Healthcare's Big Data Actually Comes From. (2018, January 11). Retrieved December 20, 2018 from https://www.techemergence.com/where-healthcares-big-data-actually-comes-from/

Widup, S., Spitler, M., Hylender, D., & Bassett, G. (2018). *2018 Verizon Data Breach Investigations Report*.

Wilson, S., Ruscoe, W., Chapman, M., & Miller, R. (2001). General practitioner-hospital communications: A review of discharge summaries. *Journal of Quality in Clinical Practice*, *21*(4), 104–108. doi:10.1046/j.1440-1762.2001.00430.x