The Role of Big Data in Healthcare: The Revolution of African Healthcare.

Achanyi Alemnji Ngufor

Department of Management Information System,

Girne American University, Kyrenia,

North Cyprus, via Mersin 10, Turkey.

Corresponding author.

Email address; achanyialemnji@gmail.com

A B S T R A C T

African healthcare has continuously evolved in quality, diversity and adoption of technology over the years. Over this years, huge volumes of data have been collected from numerous patients in several countries. Most of this data has never been revisited as it exists mostly in written forms. All this data past present and future, encompasses big data in healthcare and harnessing the power of data analytics will uncover patters that will not only improve the quality of health to save lives, but will relatively reduce the cost of implementing this healthcare. Big data through data analytics and its components such as data mining, predictive analysis and technologies such as cloud computing are the future of African healthcare. With good governance and investment in quality infrastructure, the hope of one day seeing expert and advanced healthcare in African economies at international standards is no longer a dream but a reality for all.

Introduction

Big data refers to the large quantities of complex data and information such as the public internet, institutions, telecommunications networks repositories. This type of data cannot be easily managed using the standard tools. Big data is not about large amounts of data but mostly about data analysis and data management for future analysis in order to gain insight into the data. Big data in medical services is of value to several other sectors, such as medical insurance and pharmaceutical companies. The healthcare industry will benefit from big data in that there will be a significant increase in performance that will assist the

healthcare sector to work more efficiently. The health industry will benefit from big data in that it has the ability to cut cost, help disease prediction, improve personalized care, and also help eliminate risk factors involved in the healthcare processes. It also has the ability to empower patients and caregivers. Big data will be uncovered with the rapid digitalization of data which will be highly advanced through eHealth practices put in place.

Healthcare data is big data mainly due to its extensive volumes, but also due to its complexity and the timely need for the interventions. In this sense, big data is special and must be carefully managed in order to reap the potential benefits. (Kaur, Sharma, & Mittal, 2018; Abouelmehdi, Beni-Hssane, Khaloufi & Saadi 2017; Priyanka & Kulennavar (2014), Feldman, Martin, & Skotnes 2012).

Characteristics of Big Data

When looking at the characteristics of big data, we are asking ourselves: what are those things that make big data big data and not something else. What are the dimensions that big data has to uncover for it to be recognized?

Most researchers list three main characteristics of big data;

1. **Volume;** refers to the amount of data generated by the organization, people and sub-sectors that interact. Both new and existing data comprises volume. "cloud" information storage allows for easy access by users on desktop PCs and mobile devices. making it possible for small devices and singular locations to help visualize a world of diverse information.

2. **Velocity;** describes the speed at which data occurs, is shared and captured and the speed at which the data needs to be processed. The

flow of data accumulates as unprecedented rates. Traditionally, most healthcare information has been inflexible: document files, X rays, scripts. But given that healthcare data is a very vital and usually required in real-time, it is important to optimize healthcare data.

Future applications of hospital real-time data, such as sensing infections as soon as possible, quickly identifying them and implementing the appropriate forms of treatment, could help decrease risk of medical errors or even stop occurrences in hospitals. In order to evaluate life-threatening data, real-time streaming data also can monitor neonates in the Hospital. Being able to perform real-time analytics against such high-volume data in motion could revolutionize healthcare.

3. **Variety;** depicts the proliferation of new and re-occurring data types. This data can either be structured or unstructured. Structured data comprises of data suitable for computer processing meanwhile unstructured data such as hand written notes, audios and videos. In universal health care, Big Data 's potential resides in integrating existing data with new data forms, both independently and at the population level.

Priyanka, & Kulennavar, 2014, Toshniwal, Dastidar, & Nath 2015, Inukollu, Arsi, & Ravuri 2014, Feldman, Martin, & Skotnes 2012.).

A fourth characteristic has been discussed in research;

4. **Variability;** depicts a high inconsistency and occasional peaks in data. (Toshniwal, Dastidar, & Nath 2015; Inukollu, Arsi, & Ravuri 2014, Feldman, Martin, & Skotnes 2012)

Inukollu et al 2014 made mention of a fifth dimension;

5. **Complexity;** The complexity of data must be taken into consideration as it directly affects the capacity to uncover the insights hidden within.

Key Healthcare Improvement Goals in Cameroon

1. **Health Promotion;** that will seek the adoption of healthy behaviors by the population; Big data as a tool can help the promotion of healthy behaviors by gathering behavioral data and analyzing patterns of the population on social media.

2. **Disease Prevention;** which on the one hand shall focus on the intensification of the control of priority diseases under surveillance. Big data analytics in the prevention of disease can be a vital tool to uncovering paths to potential outbreaks.

3. **Case Management;** that will prioritize the implementation of integrated high-impact intervention packages; Analyzing the available data can help healthcare professionals to prioritize the intervention packages.

4. **Health System Strengthening;** which will emphasize on implementing a financing strategy financing strategies need consistent data and making the overall cost of healthcare to reduce.

5. **Strengthening Governance, Strategic Steering and Leadership;** at all levels of the health system will be based on a more efficient management of financial resources, the reinforcement of the monitoring/evaluation system, governance, monitoring and evaluation can be easily enhanced by increased communication and availability of necessary data for decision-making.

Healthcare research and interviews have uncovered other specific goals that are universally accepted as future objectives;

- Increasing efficiencies for providers and payers, minimizing costs and errors in implementation

- Enabling research on comparative efficacy for present treatment options and informing Research and development activities

- Migrating towards outcome-oriented and patient-centered medicine to enable a better outcome for the populations

- Empowering consumers and fostering participatory healthcare" (Health 2.0,) prompting Access to personalized medicine for all.

Factors that Will Precipitate the Enhancing of Healthcare Outcomes in Africa

Various global statistics and data tilt in favour of an exponential and even disproportionate increment of the cost of healthcare. Below, the forces that will lead to such disproportionate measures will be discussed (Feldman, Martin, & Skotnes .2012);

1. **Technology;** The 20th century has seen the development of high-tech solutions. Most of these solutions have been developed as a response to real problems encountered by medical personnel and populations all over the world and in Africa. As beautiful, innovative and even more efficient as these new solutions are, they are not cost-effective nor affordable for the masses in Sub-Saharan Africa. Breakthroughs in medicine come as a consequence of the use of huge portions of time,

man-power, material resources and finances. The trend towards the occurrence of new problems and the simultaneous effort needed to mitigate, manage and cure these forecasts a huge rise in healthcare costs.

2. **Demographics;** Several African Countries are made up of vast aging populations. The occurrence of new and resilient chronic illnesses raises the burden of end-of-life care. This is especially an African issue since good governance has not put in place effective schemes to absorb the extra costs.

3. **Quality of Care;** The present issues such as uncoordinated care, the event of highly inefficient workflows, repeated medical errors, high re-admission rates due to poor treatment and early release, hospital acquired infections, nurse and personnel shortages and the lack of skilled labour tend to increase the costs of healthcare. Poor quality of life may lead to casualties, higher costs as re-admissions will be frequent, and even lawsuits against medical malpractice as is so frequently the case.

4. **Structural Issues;** Structural issues here is an umbrella term used to describe institutional inefficiencies that plague the healthcare ecosystem of most African countries. Cases of fraud, waste of vital resources, and even market distortions from third-party payers and failed reimbursement regimes are primary issues plaguing the system. The regulatory overhead caused by more practical issues pushes professionals to practice defensive medicine which in turn spirals in a loop of financial resource and time wastage.

Role of Big Data for Healthcare Ecosystem Stakeholders

The healthcare ecosystem is constituted of various individuals and organizations that interact continuously, forming a network of mutual benefit. These stakeholders are constantly finding for ways to make the system more efficient for all other users and by so doing, it becomes at better and more profitable in the long-run. Big data and the analytics carried out, will provide insights that will be beneficial to all these stakeholders but in very distinct ways;

1. **Patients;** Patients want to have all their medical information to be available at their fingertips and to flow seamlessly across all devices. Some of these patients will fancy having data on various health products available on their screens so they can make wiser decisions about purchases. Big data has the power to enhance customer-friendly service, one-stop shopping, and enable enhanced coordination of care between users in a system. Data Analytics can help achieve the ultimate goal of providing error-free, compassionate and highly efficient care for the patients while decreasing relative cost.

2. **Providers;** Data analytics can enable real-time access to valuable patient, administrative and clinical data. This will facilitate and enhance decision-making and move us one step closer to an error-free and effective care routine.

3. **Researchers;** Researchers need enhanced tools to aid the thought process in order to improve the quality and quantity of workflow. Big data analytics will bring improvements on predictive modelling, statistical

tools and algorithms. The management and outcome of experiments in order to provide innovative results that appeal to the unmet needs of the community and match regulatory measures.

4. **Pharmaceutical Companies**; Pharmaceutical companies are constantly in search for more efficient and cost-effective means to understand the cause and peculiarities of diseases and outbreaks. Drug design and testing is a vital component of modern medicine and pharmaceutical companies are always in need of ways to find more targeted drug candidates. This will lead to more efficient and successful clinical trials in shorter time spans with shorter margin for error. All this will have the effect of providing safer and better solutions at a cheaper rate and to more people. The drug market also uses extensive marketing schemes that include post-marketing surveillance. A free flowing information system will empower pharmaceuticals with faster and more accurate data that will aid decision making.

5. **Medical Device Companies;** Medical devices have been collecting various forms of data from both hospitals and individuals for a very long time. And unless this data is made sense of, it will not aid the healthcare system in anyway. Big data analytics will cut across this data and uncover patterns that will help the medical professionals to make faster, better and better decisions. A problem also poses at the level of combing the old data forms to new ones as the data collected by these machines has changed form over the years.

6. **Payers;** There is a movement from the traditional fee-for-service scheme to a more elaborate pay-for-performance pattern. Big data is a key to unlocking this new potential as it permits the stratification of population according

to risk factors and thereby guiding them to clearer, more elaborate and sustainable business models.

7. **Governments;** Governments all around the globe are trying desperately to reduce the cost of care on a daily basis for its populations. They are also putting in place infrastructure to enforce regulations that will permit them to maximize the social value of data. A common issue and hindrance to the full adoption of technology into African healthcare is the unclear distinction between cost and quality of treatment. Various African populations live on below a dollar a day and the focus of governments is usually a lot less on bringing innovative tools that increase efficiency while increasing cost of care, but their attention is on reducing the cost of care so that most people living below the poverty line can access basic healthcare at all times. Such is the difficulty faced by African governments; Reducing the cost of healthcare without affecting the quality or Increasing the quality and simultaneously bringing costs a little bit up. Big data will help governments make better decisions such as understanding risk factors in different cities, intercepting breakouts before they spread at a nationwide level, understanding the population so that the needy benefit from increased aid, as a whole just making governments more intelligent by the results provided to them. The potential of big data for government decision making on policy and financial distribution is huge and untapped in various parts of Africa.

8. Software Developers; Software and overall improved technology adoption has the

capacity to deliver better and improved care. Sectors such as banking, telecommunications, consumer operation and marketing, travel reservations, and IT are all being revolutionized at a high rate. All these sectors affect healthcare providers and will increasingly affect how it is delivered as time passes. Big data analytics sits in the middle of this evolution as data from all these different sources must be aggregated and analysed to help each one draw out maximum benefits from a future potential outcome.

It must be understood that the main use of big data is to reduce waste and inefficiency. All the above stakeholders accept that those are worthwhile goals. The issue is, the incentives of adopting such methods has conflicting interests all through the system. Research and studies have shown that all these stakeholders are willing to pay a substantial increased amount in order to achieve the ultimate goal of rapid, error-free and compassionate care (Feldman, Martin, & Skotnes .2012).

Data Pools Available in Healthcare Ecosystem

Healthcare Data can be traced back to six distorted pools of data. This data is owned by various stakeholders with different ideas and uses for them. As stated above, the healthcare ecosystem is made up various stakeholders who have competing interests but the same end goals faster, error-free and mindful service (Feldman, Martin, & Skotnes .2012). This means that all stakeholders could benefit from this data if it could be analysed together to uncover patterns that will yield new ways and approaches towards healthcare outcomes. The six major pools are;

1. **Providers;** Healthcare providers possess medical data. Only very few African Economies have managed to implement the use of Electronic health records which are a source of valuable user data.

2. **Payers and Providers;** The relationship between payers and providers in healthcare yields data on claims and the cost of treatments incurred by the payers.

3. **Researchers;** Researchers in healthcare can be grouped into three categories: Academic, government and Independent groups. Academic research yields large volumes of health data. Universities collect large data as their students and lecturers work on health related projects year in year out some of which are attributed grants. Independent researchers such as NGOs and talented individuals usually carry out various forms of research that yields enormous amounts of data. Government invest a lot of money in research that can yield interesting results for their populations.

4. **Developers;** Pharmaceutical companies, Medical device and tech companies invest a lot of resources in their Research and Development (R&D) departments. Such data includes clinical trials.

5. **Consumer and Markets;** Patient behavior, habits and sentimental data is recorded. Such data is collected from their purchase habits and choice of products and brands.

6. **Government;** Governments across the globe possess extensive volumes of population and public health data in various forms that dates back as the creation of most countries. Studying such data extensively has the potential to uncover several patterns.

Data Types Available in Healthcare (Datasets)

This section will expose the various forms in which healthcare data appears in the different data pools. It is important to note that these different data sets mean the process of analyzing

them will not be a straightforward one as most forms are hard to integrate. The various forms are;

1. **Clinical Data;** Statistics show that almost 80% of available healthcare data is stored up as unstructured data. This includes text documents, images, clinical notes and other unstructured forms.

2. **Publications;** In the form of clinical and medical research material.

3. **Clinical References;** Text-based information or guidelines available on packaging such as drugs and devices.

4. **Genomic Data;** A lot of healthcare research produces volumes of new gene sequencing data.

5. **Streamed Data;** The use of home monitoring devices produces large volumes of video data. Telehealth, sensor based devices, wireless and hand held devices constantly produce data to be analysed by healthcare.

6. **Web and Social Networking Data;** Consumers are constantly interacting with their desired product or service providers online. Data from the internet, search engines and social media interactions is widely used to understand individuals.

7. **Business & Organisational data;** Organisational data such as billing details that does not originally link with healthcare. But once integrated with other data, can uncover important details on a patient or user.

8. **Policy, Financial and Administrative;** big data analytics is intended to provide decision makers with intelligent information that aids the decision-making process. Analysing Key performance indicators can also help improve policies related to health (Priyanka & Kulennavar 2014).

Angles for the Harnessing of Big Data

It is clear and undeniable that big data analytics will provide a lot of advantages to any economic sector on whose data it is practiced including healthcare sector. Below, we will attempt to elaborate the exact ways in which this innovative practice can aid the African health sectors make use of their existing data and integrate new forms and volumes of data;

1. **Support Research; Genomics and Beyond**

- Critical research such as that on cancer is carried out by various networks of physicians across continents and across nations. Analytics can streamline a clear approach to integrate solutions from these networks to improve patient outcomes.

- The creation of continuous medical education as it aggregates and integrates data from diverse sources and decision-support systems. This can be a source of educational and professional material to empower patients.

- Permits researchers to interact with and dynamically view the end results of various interventions to aid decision-making. In order to allow our health information systems, interact with external systems, we must ensure our systems are compatible with the HIPAA, CLIA and other regulations. Encryption and firewalls are also used to enforce enterprise security.

- Analytics on data will reduce the ambiguity of medical reports and we could see the production of easy to read reports that will be easier to use by doctors, patients and hospitals.

- Genetic information on its own is irrelevant unless it is integrated with other patient data and put in a particular context. Big data

technology enables scientists to make correlations among the huge volumes of data points from the private and public data sets of genomic data.

2. Transform Data into Information

- Healthcare data is flooding hospital systems all over the world and this data will only increase in volume and complexity as time goes on. This shows that the immediate need for healthcare is to provide better management of this data. Data-driven healthcare is the end goal and this will only be made totally feasible when the bulk of unstructured data is made structured for machine management. This process will convert data into useful information for decision-making.

- Cloud-based predictive software has the potential to uncover patterns in hospital datasets and will lead to a decrease in rates of readmissions and prevent infections acquired at the hospital.

- Data will be pulled from different sources using data mining, algorithms and machine learning techniques to make effective calculated predictions.

- Predictive analytics algorithms can explore patient data with an 86% accuracy and verify the readmission risks.

- Chronic diseases such as diabetes can be avoided through the power of predictive analytics. This is especially useful for non-communicable diseases such as heart diseases and cancer.

- Research based insights can be drawn from data so as to monitor health at a population level and the possibility of early outbreak detection. Predictive analysis can help to identify the sickest members of a population and decrease hospital costs for them.

Four working areas are of high concern for predictive analysis sourced from multiple datasets on a national level to produce even larger datasets. The potential to have real-time information in a shared database;

- Health surveillance and population education on issues such as flu and asthma. This is important as many new cases may be avoided for such illnesses if only more information was spread. Education is also necessary for those already living with the illnesses as they can learn new methods to manage and cope.

- Post market surveillance of drugs: Such data can help pharmaceuticals to detect secondary effects and possible malfunctions in a released product. Patient data and post market analysis may uncover certain patterns of improvement in a certain drug.

- Public Health Research

- Plan of care and best practice Development

3. Support Self-care

- Combining the convenience and usability of the mobile phone with the power of Big data puts valuable medical information within reach of the patients and the community at large. The chronically ill can learn to manage diabetes, heart disease and asthma. This information can help mitigate, prevent and manage disease.

- Device to device communication at point of care between the payers and providers is made easy and data exchange possible.

- Providers can be informed about real-time ongoing care from other providers through an HIPAA dashboard with consent from all concerned parties.

- Machine learning and predictive modelling permit us to recognize subtle changes in normal behaviour and life patterns.

- Consumers possess vital information through the convergence of big data sets,

clinical data sets, government and sensor data. This empowers the community greatly.

4.   Support Providers- Improve Patient Care

Providers face increasing pressure as they have less time and money but are expected to produce better, error-free results. An issue faced with providers is their reluctance to change and the slow adoption of new technologies. The solution is to build real-time behavioral platforms with patient-centered care. Providers will use such systems to share data while preserving privacy.

5.   Increase Awareness

- African markets are plagued with various issues such as counterfeit drugs and big data can help identify such illicit products. Pharmaceutical companies can track drugs to avoid theft.

- Also, environmental issues that trigger asthma can be tracked, the prediction of disease outbreak enhanced, developing countries will get help making policy decisions and public health issues prioritized.

- Helps providers and payers identify high risk patients and special care is attributed to them.

- Real-time sickness forecasting that warns about the relevance of outbreaks in given areas

- Social media tracking as community members update the newsfeed with potentially viable information.

6.   Pool Data to Build a Better Ecosystem

Big data is special in its ability to bring disparate data sets together for the first time in ways that open up for new data uncovering and possibly the answer to different questions. The mixing and matching of data has been known to produce more valuable data than individual pieces.

Stages of Big Data Analytics in Healthcare

It has been mentioned that Big Data analytics has the ability to lower relative costs of healthcare, improve care, and save lives.

But for this analytics to take place effectively and yield the expected results, it must follow clear directives and procedures.

Analytics in our context of big data can be referred to as that process of cutting through large chunks of data present in different formats, from several data sources with the aim of understanding the data and producing insights that have the potential to aid decision-making in real-time and in the near future.

Various analytical concepts such as data mining, natural language processing, artificial intelligence and predictive analytics can be used in healthcare to gather, group, analyse, and visualize data. When vast amounts of data are gathered and analyzed at once, inherent patterns can be recognized, along with correlations and even anomalies which were not obvious in individual data sets.

Big data healthcare analytics has 5 distinct processes relevant to Data value: Acquisition, Storage, Management, Analytics, and finally, visualization and reporting.

Data Acquisition;

Healthcare data can be presented in three forms: structured, semi-structured and unstructured data. This data can be acquired from primary sources such as Clinical decision support systems, electronic health records, hand-written documents etc. Secondary sources of data such as laboratories, insurance companies, government sources and pharmacies. The most important sources of healthcare data are;

- Electronic health records; Patient's digital health details

- Image Processing; Visual data collected within healthcare centers

- Social Media; Updates on newsfeed and personal data

- Smartphones; Phone databases and personal data

- Web base data: From interactions with various websites.

This is the first of five stages because for anything to be processed, it must be gathered and its value is also determined by the source and quality.

Data Storage

The storage mechanism plays a big role in big data studies. This is due to the fact that the volume of data is constantly on a rise and the need for efficient, reliable and large storage platforms that can preserve the value of the data is becoming urgent. The cloud storage technology has been for some time the most promising technology for healthcare. Cloud services provide awareness and scalable analytics options that drive up the value of data. Cloud computing provides powerful options for the storage and complex manipulation on large scales of data. It solves the problem of costly hardware purchasing, maintenance and additional physical space for the use of healthcare institutes.

Data management

Data management in healthcare is a term that encompasses the organization, cleaning, retrieval, mining, and governance of data within healthcare systems. Healthcare management rapidly identifies missing, incomplete or irrelevant data. Data governance involves the overall management of the security, integrity, usability and availability of data present within a structure and an information system (health system).

Data Analytics

Data analytics is a process that seeks to transform raw data into information. Big data analytics in healthcare tends to take one of four forms;

- Descriptive Analytics: Tends to look and analyse past performances based on historic, archived data. This form of analytics gives information on what has been happening within the information system and establishes relevant parameters.

- Diagnostic Analysis: Makes use of historical data to try and come up with the root cause of a problem and discover how it came about.

- Predictive Analysis: Makes use of both real-time and historical data to attempt forecasts of what might happen in the nearest future. It makes probabilistic anticipation of patterns collected from historic data and attempts to make predict future trends.

- Prescriptive Analytics: Such analytics makes use of big data available at a given moment to provide options and ideas on various possible outcomes before a decision is made. Prescriptive analytics focuses on what is the best possible outcome, how to come about it and how to carry it out.

Data Visualization

Data visualization describes how the process data is outlined and presented for the decision makers to understand the results. Complex information is depicted graphically and in pictorial formats. Good visualization enables the uncovering of patterns and correlation among variables.

Limitations and Opportunities of the Use of Healthcare Big Data in Emerging Economies

Several advantages and expert uses of big data have been outlined so far. But various reasons exist and limit the widespread use of big data in African healthcare;

- Privacy and Security: Healthcare data is sensitive and private. Hence, the security measures on creating free-flowing information portals is high and intensive for most African economies.

- Data Ownership: There is a cycle of data ownership issues in healthcare since data originally comes from a patient, is collected by the hospital and is usually stored by

IT companies or independent cloud servers. So to benefit from healthcare data, the issue of ownership must be settled.

- Regulatory Issues; Health issues are very sensitive and the laws governing the movement and storage of such data is very robust.
- Technology; Most African countries have issues that stop the adoption of technology intensive solutions such as poor internet connections, poor IT infrastructure and Power outages which disrupt large systems.

Given all these limitations, various interviews and research have showed that there is hope for better, creative ways for emerging economies in Africa to benefit from healthcare in the near future;

- Working with limited data sets; Big data in the near future will enable us to work with limited data sets to uncover maximum value from the raw data.
- Combining a greater variety of data; With new forms of data appearing in the midst of the large volumes of healthcare data encountered, there is hope for big data to incorporate a greater variety of data.
- Greater Pooling; The increase in data sets gives added opportunities to pool data from bigger, better and richer data sets to uncover new insights of high value to healthcare.

REFERENCES

Abouelmehdi, K., Beni-Hssane, A., Khaloufi, H., & Saadi, M. (2017). Big data security and privacy in healthcare: A Review. *Procedia Computer Science*, *113*, 73-80.

Feldman, B., Martin, E. M., & Skotnes, T. (2012). Big data in healthcare hype and hope. *Dr. Bonnie*, *360*, 122-125.

Inukollu, V. N., Arsi, S., & Ravuri, S. R. (2014). Security issues associated with big data in cloud computing. *International Journal of Network Security & Its Applications*, *6*(3), 45.

Kaur, P., Sharma, M., & Mittal, M. (2018). Big data and machine learning based secure healthcare framework. *Procedia computer science*, *132*, 1049-1059.

Priyanka, K., & Kulennavar, N. (2014). A survey on big data analytics in health care. *International Journal of Computer Science and Information Technologies*, *5*(4), 5865-5868.

Toshniwal, R., Dastidar, K. G., & Nath, A. (2015). Big data security issues and challenges. *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, *2*(2).