International Journal of Big Data and Analytics in Healthcare
Volume 4 • Issue 1 • January-June 2019

The Strengths, Weaknesses, Opportunities, and Threats Analysis of Big Data Analytics in Healthcare

Chaojie Wang, The MITRE Corporation, McLean, USA
https://orcid.org/0000-0001-8521-9420

ABSTRACT

Improving the performance and reducing the cost of healthcare have been a great concern and a huge challenge for healthcare organizations and governments at every level in the US. Measures taken have included laws, regulations, policies, and initiatives that aim to improve quality of care, reduce costs of care, and increase access to care. Central to these measures is the meaningful and effective use of Big Data analytics. To reap the benefits of big data analytics and align expectations with results, researchers, practitioners, and policymakers must have a clear understanding of the unique circumstances of healthcare including the strengths, weaknesses, opportunities, and threats (SWOT) associated with the use of this emerging technology. Through descriptive SWOT analysis, this article helps healthcare stakeholders gain awareness of both success factors and issues, pitfalls, and barriers in the adoption of big data analytics in healthcare.

KEYWORDS

Big Data, Data Analytics, Data Breaches, Data Ethics, Electronic Health Records, Health Information Exchange, Health IT, Healthcare, Machine Learning, SWOT Analysis

1. INTRODUCTION

The US healthcare system has both strengths and weaknesses. It enjoys a large-scale, well-trained, and high-quality workforce of clinicians, nurses, and specialists, robust medical research programs, and the world's best clinical outcomes in select medical services. Yet, it suffers from high expenditure, low performance, and disparity in health status, access to care, and outcomes of care (Barnes, Unruh, Rosenau, & Rice, 2018).

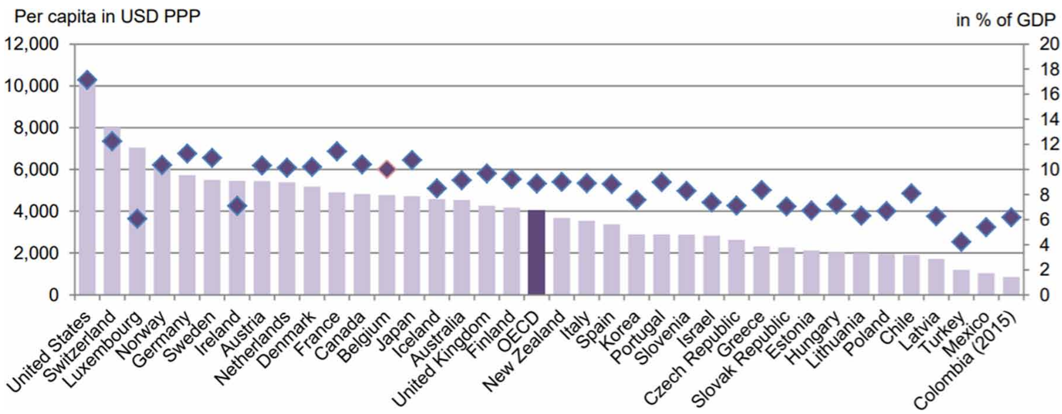
1.1. High Cost of the US Healthcare System

According to a recent report published by The Organization for Economic Co-operation and Development (2018), in 2017 the US spending on healthcare was the largest, measured by both the spending per capita and the percentage of the gross domestic product (GDP) among its 37 member nations. Figure 1 shows that the US spent over $10,000 per capita on healthcare that year, or about 17% of GDP.

Even more alarming is the rapid growth in US healthcare spending. According to the Centers for Medicare and Medicaid Services (CMS), healthcare spending is projected to grow at an average rate of 5.8 percent from 2012-2022, 1.0 percentage point faster than the expected average annual growth in the GDP. By 2022, US healthcare spending is projected to be nearly 20% of GDP (Centers for Medicare and Medicaid Services, 2012).

DOI: 10.4018/IJBDAH.2019010101

Figure 1. 2017 health spending per capita as share of GDP (Organization for Economic Co-operation and Development, 2018, p. 2)



1.2. Low Performance of the US Healthcare System

This extremely high spending is sharply contrasted with the low performance in the US healthcare system. In 2000, the World Health Organization (WHO) published a report that measured and ranked the health system performance of 191 countries. According to this report, the US healthcare system was unimpressively ranked at 37, below most industrialized countries including France, the UK, and Canada, and even below some less developed countries such as Colombia and Chile (Tandon, Murray, Lauer, & Evans, 2000). Almost two decades later, there has not been much improvement in the performance of the US healthcare system. According to a 2017 report from the Commonwealth Fund, the US is ranked last out of 11 high-income industrialized countries based on measures including care process, access to care, administrative efficiency, equity, and health outcomes (Schneider, Sarnak, Squires, Shah, & Doty, 2017). Figure 2 shows that the US healthcare system performs at the bottom on four of the five measures.

1.3. Efforts to Improve the US Healthcare System

In 2007, the Institute for Healthcare Improvement (IHI) launched the Triple Aim initiative to improve the patient experience of care (including quality and satisfaction), improve the health of populations, and reduce the per capita cost of health care. The Triple Aim initiative directly targets the critical measures of healthcare performance including both quality of care and efficiency of care (Institute for Healthcare Improvement, 2007).

In 2008, Congress enacted the Medicare Improvements for Patients and Providers Act (MIPPA). As part of the implementation of MIPPA, CMS introduced the value-based purchasing (VBP) plan linking payments directly to the quality outcomes of the care provided. This pay-for-quality or pay-for-performance plan aims to move away from the traditional fee-for-service plan which provides no incentives for the providers to improve care quality and contributes to the high cost of healthcare (Terhaar, 2018).

In 2010, to address the disparity and inequity in healthcare and to increase access to care for tens of millions of uninsured and underinsured Americans, Congress enacted the Patient Protection and Affordable Care Act of 2010, also known as "Obamacare" ("Patient Protection and Affordable Care Act of 2010," 2010).

While healthcare initiatives, regulations, and policies may help drive the quality and performance improvement at the macro level, effective implementations require concerted efforts at the micro level by all stakeholders including policymakers, providers, payer, patients, and the public. In addition, these diverse stakeholders must be empowered and enabled by innovative solutions and technologies

Figure 2. 2017 healthcare systems performance ranking (Schneider et al., 2017)

	AUS	CAN	FRA	GER	NETH	NZ	NOR	SWE	SWIZ	UK	US
OVERALL RANKING	2	9	10	8	3	4	4	6	6	1	11
Care Process	2	6	9	8	4	3	10	11	7	1	5
Access	4	10	9	2	1	7	5	6	8	3	11
Administrative Efficiency	1	6	11	6	9	2	4	5	8	3	10
Equity	7	9	10	6	2	8	5	3	4	1	11
Health Care Outcomes	1	9	5	8	6	7	3	2	4	10	11

to deal with the complex problems in healthcare. Big data analytics as an emerging technology is one of many indispensable tools in the toolbox and can play an important role in improving healthcare. However, as with any new technology, benefits often come with limitations, opportunities with risks, and hopes with hypes. It is important to understand the full spectrum of this new technology so that it can be efficiently and effectively adopted and applied.

Motivated by the tremendous challenges facing the US healthcare system and the great potential big data analytics has in improving its performance, this paper presents a high-level analysis of the strengths, weaknesses, opportunities, and threats (SWOT) associated with the use of big data analytics in healthcare. The goal of this paper is to help policy makers, care providers, researchers, and the public gain broader and deeper understanding of the many dimensions of the use of this emerging technology in healthcare so that they can make informed and sensible decisions in their efforts to improve healthcare quality and performance.

2. BIG DATA ANALYTICS AS ENABLER TO IMPROVE HEALTHCARE

2.1. Big Data Analytics

During the past decade, big data emerged as a new technology trend thanks to the rapid innovation and advancement in information and communication technology (ICT). Big data was initially defined with three essential characteristics known as the Three V's – Volume, Velocity, and Variety. Since then, more V's have been added. A Six V's model widely used in healthcare adds three additional V's – Veracity, Variability, and Value (Senthilkumar, Rai, Meshram, Gunasekaran, & Chandrakumarmangalam, 2018). Data analytics is an umbrella term commonly used to reference business intelligence, business analytics, data mining, knowledge discovery in databases, and data science. As large volumes of data in wide varieties are collected and made available, the demand increases for extracting values from big data to improve business performance and drive organizational changes. Data analytics answers the challenge by integrating computer science, information technology, statistics, domain knowledge, and human collaboration in a streamlined process of knowledge discovery, creation, and application (Wang, 2018).

2.2. Applications of Big Data Analytics in Healthcare

Healthcare data analytics is the application of big data and data analytics to healthcare. Similar terms with minor differences also appear both in academia and industry. Among them are healthcare analytics, health analytics, healthcare big data analytics, big data analytics in health, and health informatics. Knowledge can be discovered, and insights can be gained from big data analytics to improve healthcare quality and performance. Based on the content analysis of 26 case studies, Wang, Kung, and Byrd (2016) identified five potential benefits of healthcare data analytics: IT infrastructure, operational, organizational, managerial, and strategic benefits. Lebied (2018) provided detailed accounts of twelve applications of big data analytics in healthcare such as prevention of opioid abuse, telemedicine, prevention of ER visits, and fraud detection. These serve as exemplars of the many applications currently recognized and many more innovative applications continuing to emerge.

The University of Pennsylvania Health System (UPHS)'s use of machine learning and Electronic Health Records (EHR) data to predict the onset of severe sepsis is an example of using big data analytics in healthcare. Sepsis is a life-threatening illness due to blood infection. According to the Center for Disease Control & Prevention (CDC), each year, 1.7 million American adults develop sepsis, 270,000 Americans die of sepsis, and one in three patients who die in hospitals has sepsis (Dantes & Epstein, 2018). Sepsis costs American hospitals $27 billion annually, making it the number one cost driver for hospitals (Reinhart, 2018). UPHS developed a predictive model by training a machine learning algorithm using historic patient data stored in its EHR including labs, clinical, and demographic data. The trained machine learning model was then used to predict severe sepsis 12 hours prior to the clinical onset by continuously pulling real-time patient data from the EHR. Alerts were sent to clinicians and nurses to enable additional monitoring and early intervention (Giannini et al., 2017).

2.3. The Adoption Model for Analytics Maturity (AMAM)

When it comes to the adoption and utilization of big data analytics, each organization is constrained by its unique organizational characteristics including size, financial resources, technical capability, and leadership and environmental characteristics such as geographical location, community, and patient populations. To help assess the maturity of healthcare organizations in adopting big data analytics, HIMSS Analytics, a wholly owned subsidiary of the Health Information and Management System Society (HIMSS), developed the Adoption Model for Analytics Maturity (AMAM) (HIMSS Analytics, n.d.). AMAM describes eight stages representing increasing levels of maturity. An organization typically starts out at stage zero in which it only uses analytics in sparse and siloed point solutions and gradually climbs up to stage seven, where analytics is used for prescriptive and personalized care for individual patients.

3. THE SWOT FRAMEWORK AND SUMMARY OF ANALYSIS

3.1. The SWOT Framework

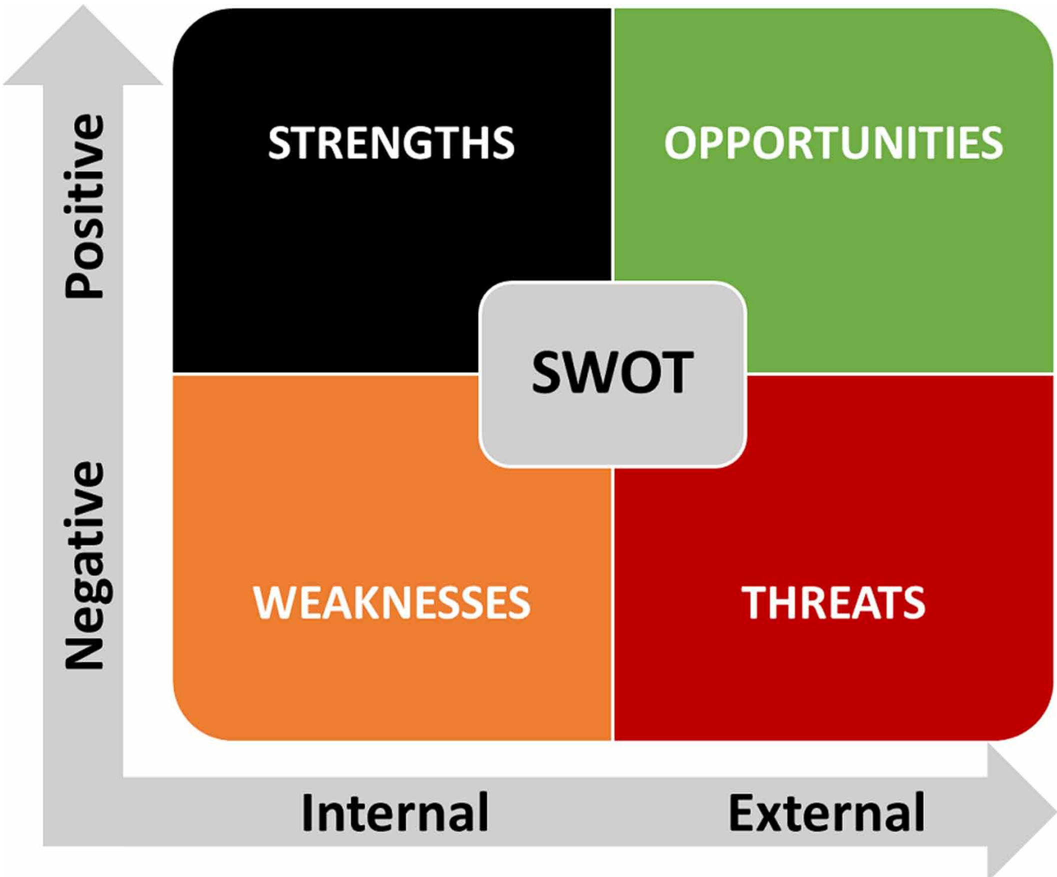
Strengths-Weaknesses-Opportunities-Threats (SWOT) is a management science framework originally used for corporate strategic planning dating back to the 1960s. It is an analysis tool to assess a business's fitness as measured by how well its internal qualities (the strengths and weaknesses) match up with external factors (the opportunities and threats) (Hill & Westbrook, 1997).

Strengths and opportunities are considered positive, favorable, and synonymous to success factors while weaknesses and threats are considered negative, unfavorable, and synonymous to pitfalls, challenges or barriers. The objective of strategic planning is to mitigate or minimize the unfavorable factors and utilize or maximize the favorable factors. The SWOT framework has been applied to the use of big data analytics in general (Wang, Wang, & Alexander, 2015) and in a specific industry or domain (Collins, 2016). This paper applies SWOT analysis to the use of big data analytics in healthcare. Figure 3 shows the framework in four quadrants, known as a SWOT matrix.

3.2. Strengths and Weaknesses of the SWOT Framework

SWOT is a simple yet powerful model for analyzing a complex situation. It reflects the complexity of the reality and presents a dialectically balanced perspective for analyzing and dealing with complex

Figure 3. The SWOT matrix



problems. However, SWOT analysis is not considered a rigorous empirical research method. It is different from the qualitative method aiming at the generation of new theories or the quantitative method aiming at the confirmation or falsification of existing or proposed theories. SWOT provides a way of thinking akin to systems thinking or critical thinking. The definition and classification of what constitutes strengths, weaknesses, opportunities, and threats are subjective and reflect the researchers and practitioners' personal experience and understanding of the complex socioeconomic, cultural, and managerial problems. SWOT's power lies in its simplicity in dealing with complexity. It provides a simple frame of reference for analyzing and dealing with complex situations so that potential opportunities can be explored, potential risks can be assessed, and potential outcomes (intended or unintended, desirable or undesirable) can be illuminated before interventions can be developed and actions can be taken.

3.3. SWOT Analysis of Big Data Analytics in Healthcare

This SWOT analysis was performed based on review of literature, industry reports, and expert opinions. The author also draws upon his professional experiences in systems engineering, health IT, data analytics, and healthcare quality management. Aligning with the four quadrants of the SWOT matrix, this paper sets out to answer the following four questions:

1. What are the strengths of healthcare that make it favorable for the adoption of big data analytics?

2. What are the weaknesses of healthcare that may hinder the adoption of big data analytics?
3. What are the opportunities that favor the adoption of big data analytics in healthcare?
4. What are the threats associated with the adoption of big data analytics in healthcare?

The extant literature on similar subjects tends to be at a lower level or with technical implementation details. Collins (2016) performed a SWOT analysis of big data and health economics in the UK with a focus on costs and included coverage of drugs, biomonitoring, and data repositories. Yang et al. (2016) performed a SWOT analysis of wearable devices in healthcare. While the flexibility of the SWOT framework allows for different levels of analysis, this paper focuses on the big picture (the forest) instead of low-level details (the trees). The results of the SWOT analysis are summarized in Figure 4. Detailed descriptions are provided in the following four sections.

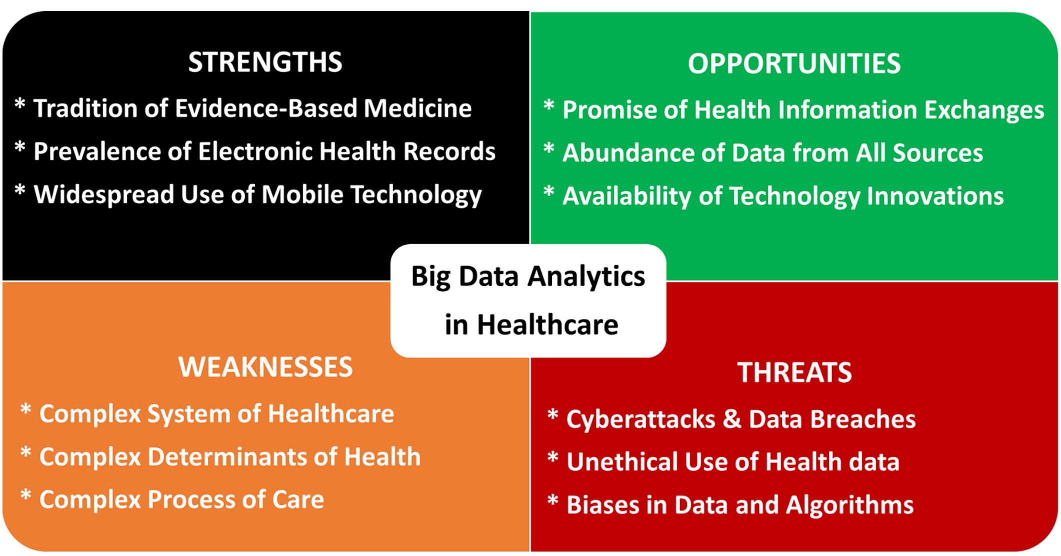
4. STRENGTHS

4.1. Tradition of Evidence-Based Medicine

Western medicine has the tradition of applying science and technology in the prevention, control, diagnosis and treatment of diseases. From the development of medical devices such as Magnetic Resonance Imaging (MRI) and the Automated External Defibrillator (AED) to the use of experimental design and statistical inference in clinical trials, science and technology remain at the heart of modern medicine.

A spirited movement called Evidence-Based Medicine (EBM) began within the healthcare community in the early 1990s. EBM is defined as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research” (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). EBM has become an essential code in the DNA of healthcare ever since. Its principles, processes, and practices are well aligned with those of big data and analytics. The increasing volume, variety, and availability of data along with the advancement of data storage, computing power, and analytics tools and techniques will make EBM more practical and powerful.

Figure 4. Summary of the SWOT analysis on big data analytics in healthcare



4.2. Prevalence of Electronic Health Records

Electronic Health Records (EHR) evolved from traditional Electronic Medical Records (EMR). While EMR was limited to medical records, EHR expands the scope to all records germane to health including medical, administrative, claim, and socioeconomic records. In 2009, the US Congress enacted the Health Information Technology for Economic and Clinical Health (HITECH) Act as part of the American Recovery and Reinvestment Act (ARRA) after the 2008 global financial crisis. HITECH was created to promote the adoption of health IT and implementation of EHR to modernize technology infrastructure and capabilities of healthcare. Federal investment was provided to healthcare providers to incentivize health IT adoption and a federal certification process was put in place to ensure IT systems are developed according to technological, functional, and security requirements.

Since the enactment of the HITECH Act, more healthcare organizations have adopted health IT and implemented EHRs. According to the Office of the National Coordinator for Health Information Technology (2018), “In 2017, 96 percent of all non-federal acute care hospitals possessed certified health IT.” The penetration of health IT and EHRs provides the technical foundation and data sources for the adoption of big data analytics to improve healthcare. At the same time, big data analytics aims to deliver the right amount of information to care providers at the right time in the right place and help realize the intended benefits of EHRs and alleviate the unintended burden associated with EHRs such as administrative overhead and information overload.

4.3. Widespread Use of Mobile Technology

The ubiquitous Internet coupled with Global Positioning Systems (GPS), mobile telecommunication networks, and cloud computing lead to the proliferation of mobile devices (including wearable fitness trackers) and mobile apps that are becoming an integral part of healthcare and our daily life. According to Statista (2018), as of January 2018 there were 3.7 billion mobile users worldwide. While the global mobile broadband subscription penetration rate is around 50%, the Americas and Europe have the highest rates, around 78.2 percent and 76.6 percent, respectively.

Among millions of mobile apps, health and wellness apps are becoming increasingly popular. There are close to 48,000 mobile health and wellness apps available from the Apple App Store alone (Statista, 2018). Mobile health and wellness apps not only change the way healthcare is delivered and how health is monitored and managed, but also provide additional biometrics and lifestyle data to the clinicians, researchers, and policymakers in their efforts to improve care for individual patients and populations at large.

5. WEAKNESSES

5.1. Complex System of Healthcare

The US healthcare delivery and payment system is complex with many interdependent, interacting stakeholders. As a complex adaptive system, it exhibits non-linear, dynamic, and indeterministic behaviors that are unpredictable and difficult to manage and control (Rouse, 2008). Figure 5 shows the many stakeholders and how each of them plays a different role in a complex relationship to deliver healthcare.

Improving healthcare requires the collaboration and concerted efforts of all stakeholders through consensus building. Government policymaking should be conducted by taking inputs from all stakeholders and any potential unintended adverse consequences should be evaluated and remedied. For example, CMS’s bundled payment initiative intended to reduce cost of care may lead to care providers’ discrimination in patient selection commonly known as “lemon dropping” and “cherry picking”, where high risk patients are turned away in favor of low risk patients. Big data analytics can provide evidence and insights to help policymakers assess the intended benefits and unintended harm of health policies.

Figure 5. Stakeholders and interests in health care (Rouse, 2008, p. 19)

Stakeholder	Risk Management	Prevention	Detection	Treatment
Public	e.g., buy insurance	e.g., stop smoking	e.g., get screened	
Delivery System			Clinicians[a]	Clinicians and providers[b]
Government	Medicare, Medicaid, Congress	NIH, Government CDC, DoD, et al.	NIH, Government CDC, DoD, et al.	NIH, Government CDC, DoD, et al.
Non-Profits		American Cancer Society, American Heart Association, et al.	American Cancer Society, American Heart Association, et al.	American Cancer Society, American Heart Association, et al.
Academia	Business schools	Basic science disciplines	Technology and medical schools	Medical schools
Business	Employers, insurance companies, HMOs		Guidant, Medtronic, et al.	Lilly, Merck, Pfizer, et al.

[a]The category of clinicians includes physicians, nurses, and other health care professionals.
[b]The category of providers includes hospitals, clinics, nursing homes, and many other types of testing and treatment facilities.

5.2. Complex Determinants of Health

While modern medicine has great success at treating symptoms and managing both acute and chronic conditions, it is not as successful in preventing and curing diseases. This is due to the complex nature of diseases and the multiple contributing factors including clinical, behavioral, and socioeconomic factors. There is a growing body of research linking socioeconomic factors, such as an individual's life condition and social standing, to his or her health status under the general scheme of social determinants of health (Barr, 2014; Marmot, 2005; McGovern, Miller, & Hughes-Cromwick, 2014). There is a myriad of factors that affect health and influence healthcare. Some are natural while others are cultural; some are measurable while others are not. Machine learning and artificial intelligence may be effective in analyzing natural or measurable factors but must rely on human intelligence and human judgement to deal with the cultural or unmeasurable factors.

5.3. Complex Process of Care

Healthcare is a team sport where multiple professionals must work together to deliver quality care to patients (Nancarrow et al., 2013). For example, the dialysis care for patients suffering from End-Stage Renal Disease (ESRD) requires coordinated efforts by administrative, medical, and social professionals including facility administrators, medical directors, nephrologists, nurses, dialysis technicians, dietitians, and social workers (Maryland Department of Health, n.d.). In addition, ESRD patients tend to have comorbidities such as diabetes and hypertension and are vulnerable to hospitalizations and ER visits. Their quality of life depends on the coordinated care by dialysis facilities, hospitals, communities and their families. This process of care is complex and makes the quality of care much harder to quantitatively measure and analyze.

In addition, patient engagement plays a critical role in achieving optimal health outcomes in the patient-centered care process. Patients must be informed of how their treatment is being influenced by data, evidence, and analytics and be engaged in the healthcare decision making process. This requires ongoing patient education to improve both health literacy and data literacy.

The complexity articulated in the above three interrelated areas poses great challenges to the effectiveness of big data analytics. As much as we would like to use the insights gained from big data analytics to improve the performance of healthcare, actions cannot be taken solely based on the outputs of nondeterministic computer algorithms. For example, IBM Watson generated considerable buzz for its purported utility in helping doctors more rapidly and accurately diagnose illness, but a recent article revealed that Watson failed to live up to those expectations (Hernandez & Greenwald, 2018).

6. OPPORTUNITIES

6.1. Promise of Health Information Exchanges

While EHRs enable healthcare organizations to centralize the management of patient health data, the Health Information Exchanges (HIEs) go one step further by providing the technical infrastructure to connect these disparate and diverse EHRs so that patient health data can be exchanged, aggregated, and shared across the whole healthcare delivery system. “Electronic exchange of clinical information allows doctors, nurses, pharmacists, other health care providers, and patients to access and securely share a patient’s vital medical information electronically - improving the speed, quality, safety, coordination, and cost of patient care” (The Office of the National Coordinator for Health IT, n.d.). HIEs make it possible to establish comprehensive, inclusive, and longitudinal patient and population health data for the effective application of big data analytics.

Since 2009, the Social Security Administration (SSA) has been using HIEs to automatically and timely request and obtain medical evidence records (MERs) from healthcare providers nationwide to support disability claims adjudication (Feldman & Horan, 2011). As of December 2018, over 18,000 healthcare providers represented by 150 healthcare organizations participated in the exchange of MERs with SSA. Compared to a paper process, the use of HIEs greatly speeds up medical evidence acquisition to support disability determination. Faster availability of benefits resulting from the expedited determination helps disability claimants pay for much needed healthcare services and supports healthcare providers in the delivery of healthcare (Social Security Administration, n.d.a). In addition, SSA uses the vast amount of electronic health data along with machine learning to provide data-driven decision support to its disability adjudicators. This use of big data analytics further speeds up the disability adjudication process and helps increase the accuracy and consistency of disability determination decisions. SSA established interoperability guidelines and a certification process to ensure participating organizations comply with industry interoperability standards (Social Security Administration, n.d.b).

6.2. Abundance of Data from All Sources

One of the many V’s of big data is variety. For big data analytics to be effective, many kinds of data that are related and relevant to health and healthcare should be included, especially the socioeconomic data that measure the important social determinants of health. EHRs may contain individual patient’s demographic data but lacks the macro socioeconomic data that can be obtained from government census and survey data.

For example, the American Community Survey data from the US Census Bureau contain rich sets of socioeconomic data about communities in the US and could be used in conjunction with clinical, administrative, and claims data to gain a better understanding of the state of health and healthcare. In addition, data from law enforcement, non-governmental organizations, and social media networks can also be leveraged for big data analytics in healthcare.

In January 14, 2019, President Trump signed into law the Foundations for Evidence-Based Policymaking (FEBP) Act. As part of the FEBP Act, the Open, Public, Electronic and Necessary (OPEN) Government Data Act requires all non-sensitive government data to be made available in open and machine-readable formats. The successful implementation of this federal law will help propel the adoption of big data analytics.

6.3. Availability of Technology Innovations

The past decade has seen a rapid advancement in computer science and information technology. The confluence of cloud computing, mobile computing, machine learning, artificial intelligence, and Internet of Things has given rise to a plethora of nascent tools, techniques, and platforms for performing productive and effective big data analytics. There are many readily available choices of both commercial-off-the-shelf (COTS) products and free open source software (FOSS) products. There are many virtual communities, blogs, forums, and tutorials available on the web. Data science has emerged as a burgeoning vibrant profession, and hundreds of data science programs have sprouted up in colleges around the world.

7. THREATS

7.1. Cyberattacks and Data Breaches

While big data analytics has the potential to help improve the quality and performance of healthcare, the risks of cyberattacks and data breaches are high and should not to be overlooked. According to a recent study over a five-year period from 2012 to 2017, there were 1,512 reported data breaches of protected health information affecting a total of more than 154 million patient records (Ronquillo, Erik Winterholler, Cwikla, Szymanski, & Levy, 2018). These stolen data can be used for "identity theft, criminal impersonation, tax fraud, health insurance scams, and a host of other criminal offenses" (Goodman, 2016, p. 109).

The cyberattacks on healthcare IT systems not only put patients' privacy and security at risk, they also pose economic harms to healthcare providers, insurers, and tax payers. A 2016 study by the Ponemon Insfitute (2016) estimated that about 90% of the healthcare organizations represented in the study suffered data breaches in the past two years and the average cost of data breaches for covered entities surveyed was more than \$2.2 million, while the average cost to business associates in the study was more than \$1 million.

7.2. Unethical Use of Health Data

External malicious cyberattacks are certainly grave concerns, however insider threats and unethical use of health data to discriminate health insurance coverage and to boost corporate revenues and profits should not be overlooked. Big data analytics is a double-edged sword. When applied properly and ethically, it has the potential to do good; otherwise, it may cause harm to patients, providers, and tax payers. One emerging application is the use of predictive analytics to draw insights for profits without regard to privacy laws and protection of patients' personal and health information.

A 2017 report by The Century Foundation (Tanner, 2017) painted a gloomy picture of big data in healthcare. There exists a multi-billion-dollar industry that collects, mines, buys, and sells anonymized patient health data. The patient health data are traded routinely for profit. While the sharing and use of de-identified patient health data for secondary use in medical and policy research are allowed by the Health Insurance Portability and Accountability Act (HIPAA) (Cohen & Mello, 2018), there is a real danger for the patient health data to be re-identified through the process of data linking with additional data sources from social media networks and mobile health and wellness apps and the use of machine learning algorithms. This unethical for-profit data mining along with the upsurge in malicious hacking and data breaches can result in devastating impacts.

7.3. Biases in Data and Algorithms

Reality is complex and unknowable. As stated by Laozi in the 2500-year old Taoist text *Tao Te Ching*, "The tao that can be told is not the eternal Tao. The name that can be named is not the eternal Name. The unnamable is the eternally real" (Laozi, Mitchell, Roig, & Little, 1989). Echoing Laozi, statistician George Box (1976) was famously quoted for the maxim that "all models are wrong" because they are only approximations of true reality. Big data analytics is inherently biased since it relies on data as input, and algorithms as the enablers. Limitations and biases exist in both data and algorithms which can be traced back to the cognitive limitations and biases in human minds (Gianfrancesco, Tamang, Yazdany, & Schmajuk, 2018).

Whether collected through human observations or sensory devices, data are approximate measures of the actual properties of the observed entities. Even correctly measured and curated data cannot escape from the biases of those who define the measurement and those who design the data collection instruments. In addition, incomplete, inaccurate, and missing data are typical and can distort the outcome of the analytics (affectionately known as "garbage in, garbage out").

8. CONCLUSION

This paper presented a high-level analysis of the strengths, weaknesses, opportunities, and threats (SWOT) in the application of big data analysis in healthcare. While the strengths and opportunities are positive, desirable, and easier to comprehend, the weaknesses and threats are negative, undesirable, and demand diligence and vigilance. Although this paper provides even coverage of all four factors, it should be stressed that more attention must be paid to the weaknesses and threats. This is especially true in the era of rapid innovations in information technology which often give rise to an unrealistic expectation of a panacea with quick, autonomous, and magical cures for a complex social problem.

Technology such as machine learning and artificial intelligence have been effective in dealing with "hard" problems such as winning "Go" games, recognizing human voices and faces, and even translating languages. These problems have well-defined boundaries, well-understood rules of game or causal mechanisms, and can be simulated or programmed using software with high degrees of accuracy and certainty. However, social problems are much more complex with ill-defined boundaries and poorly-understood causal mechanisms (Kirk, 1995). Healthcare is a perfect example of a complex system with non-linear, dynamic, and indeterministic behaviors involving multiple stakeholders with conflicts of interests and under the influence of multitudes of intertwined natural and social forces. These weaknesses are inherent in healthcare and should be kept in mind when applying technologies such as health IT and big data analytics. There is no simple and straight-forward solution to the high cost and low-quality challenges facing the US healthcare system. Big data analytics is promising and can help uncover partial and limited knowledge from big data to inform clinical and policy decision making, but it does not provide the whole truth about healthcare and is not a silver bullet; overcoming the weaknesses requires the conscious and judicial use of political skills, business acumen, and collaborative spirit in addition to technical and analytical competency.

While this paper provides only cursory exposure to the subject of big data analytics in healthcare, one of its objectives is to draw attention to some of the many interrelated factors that must be considered when seeking to leverage big data analytics to improve healthcare quality and performance. A more detailed and deeper analysis on any of the factors mentioned in the above high-level analysis would represent a worthwhile research subject. For example, healthcare is a highly personalized process involving the coordination of multiple providers such as acute care hospitals, primary care physicians, and nurse practitioners. Care coordination is a known factor influencing such healthcare outcomes as unplanned hospital readmission (Fluitman et al., 2016). Quantitatively measuring health care coordination is a challenging task. Compared to readily-quantifiable clinical measures such as blood pressure or body mass index, individual human behaviors are much harder to quantify, and care coordination involves multiple parties. Moreover, fair and just attribution of quality measure scores to multiple providers is also a difficult task. This is further complicated by patient behaviors since patients are also part of the coordination process.

DISCLAIMER

The author's affiliation with The MITRE Corporation is provided for identification purposes only and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions or viewpoints expressed by the author. Approved for Public Release; Distribution Unlimited. Case Number 19-0249.

REFERENCES

Analytics, H. I. M. S. S. (n.d.). Adoption Model for Analytics Maturity. Retrieved from https://www.himssanalytics.org/amam

Barnes, A. J., Unruh, L. Y., Rosenau, P., & Rice, T. (2018). Health System in the US. In E. van Ginneken & R. Busse (Eds.), *Health Care Systems and Policies*. New York, NY: Springer. doi:10.1007/978-1-4614-6419-8_18-2

Barr, D. A. (2014). *Health disparities in the United States: Social class, race, ethnicity, and health*. JHU Press.

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799. doi:10.1080/01621459.1976.10480949

Centers for Medicare and Medicaid Services. (2012). *National health expenditure projections 2012-2020*. Retrieved from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/Proj2012.pdf

Cohen, I. G., & Mello, M. M. (2018). HIPAA and Protecting Health Information in the 21st Century. *Journal of the American Medical Association*, *320*(3), 231. doi:10.1001/jama.2018.5630 PMID:29800120

Collins, B. (2016). Big data and health economics: Strengths, weaknesses, opportunities and threats. *PharmacoEconomics*, *34*(2), 101–106. doi:10.1007/s40273-015-0306-7 PMID:26093888

Dantes, R. B., & Epstein, L. (2018). Combatting Sepsis: A Public Health Perspective. *Clinical Infectious Diseases*, *67*(8), 1300–1302. doi:10.1093/cid/ciy342 PMID:29846544

Feldman, S. S., & Horan, T. A. (2011). Collaboration in electronic medical evidence development: A case study of the Social Security Administration's MEGAHIT System. *International Journal of Medical Informatics*, *80*(8), e127–e140. doi:10.1016/j.ijmedinf.2011.01.012 PMID:21333588

Fluitman, K. S., van Galen, L. S., Merten, H., Rombach, S. M., Brabrand, M., Cooksley, T., & Nanayakkara, P. W. B. et al. (2016). Exploring the preventable causes of unplanned readmissions using root cause analysis: Coordination of care is the weakest link. *European Journal of Internal Medicine*, *30*, 18–24. doi:10.1016/j.ejim.2015.12.021 PMID:26775179

Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, *178*(11), 1544–1547. doi:10.1001/jamainternmed.2018.3763 PMID:30128552

Giannini, H., Chivers, C., Draugelis, M., Hanish, A., Fuchs, B., Donnelly, P., & Schweickert, W. et al. (2017). D15 Critical Care: Do We Have A Crystal Ball? Predicting Clinical Deterioration and Outcome in Critically Ill Patients: Development And Implementation Of A Machine-Learning Algorithm For Early Identification Of Sepsis In A Multi-Hospital Academic Healthcare System. *American Journal of Respiratory and Critical Care Medicine*, 195.

Goodman, M. (2016). *Future crimes: Inside the digital underground and the battle for our connected world*. Random House.

Hernandez, D., & Greenwald, T. (2018). IBM Has a Watson Dilemma. *The Wall Street Journal*. Retrieved from https://www.wsj.com/articles/ibm-bet-billions-that-watson-could-improve-cancer-treatment-it-hasnt-worked-1533961147

Hill, T., & Westbrook, R. (1997). SWOT analysis: It's time for a product recall. *Long Range Planning*, *30*(1), 46–52. doi:10.1016/S0024-6301(96)00095-7

Institute for Healthcare Improvement. (2007). Triple Aim - The Best Care for the Whole Population at the Lowest Cost. Retrieved from http://www.ihi.org/Engage/Initiatives/TripleAim/Pages/default.aspx

Kirk, D. (1995). Hard and soft systems: A common paradigm for operations management. *International Journal of Contemporary Hospitality Management*, *7*(5), 13–16. doi:10.1108/09596119510090708

Laozi, M.S., Roig, J. V., & Little, S. (1989). *Tao teaching*. Kyle Cathie.

Lebied, M. (2018). 12 examples of big data analytics in healthcare that can save people. *Datapine*. Retrieved from https://www.datapine.com/blog/big-data-examples-in-healthcare/

Marmot, M. (2005). Social determinants of health inequalities. *Lancet*, *365*(9464), 1099–1104. doi:10.1016/S0140-6736(05)74234-3 PMID:15781105

Maryland Department of Health. (n.d.). Code of Maryland regulations: Sec. 10.30.02.04. freestanding dialysis facilities – staffing. Retrieved from http://mdrules.elaws.us/comar/10.30.02.04

McGovern, L., Miller, G., & Hughes-Cromwick, P. (2014). *The relative contribution of multiple determinants to health outcomes*. Project HOPE.

Nancarrow, S. A., Booth, A., Ariss, S., Smith, T., Enderby, P., & Roots, A. (2013). Ten principles of good interdisciplinary team work. *Human Resources for Health*, *11*(1), 19–19. doi:10.1186/1478-4491-11-19 PMID:23663329

Office of the National Coordinator for Health Information Technology. (2018). Percent of hospitals, by Type, that possess certified health IT. Retrieved from https://dashboard.healthit.gov/quickstats/pages/certified-electronic-health-record-technology-in-hospitals.php

Patient Protection and Affordable Care Act of 2010, 42 U.S. C § 18031. (2010). US Congress.

Ponemon Institute. (2016). *Six annual benchmark study on privacy & security of healthcare data*. Retrieved from https://www.ponemon.org/local/upload/file/Sixth%20Annual%20Patient%20Privacy%20%26%20Data%20Security%20Report%20FINAL%206.pdf

Reinhart, K. (2018). Sepsis: A global public health burden for hospitals. *Becker's Hospital Review*. Retrieved from https://www.beckershospitalreview.com/quality/sepsis-a-global-public-health-burden-for-hospitals.html

Ronquillo, J. G., Erik Winterholler, J., Cwikla, K., Szymanski, R., & Levy, C. (2018). *Health IT, hacking, and cybersecurity: national trends in data breaches of protected health information*. JAMIA Open.

Rouse, W. B. (2008). Health care as a complex adaptive system: Implications for design and management. *Bridge-Washington-National Academy of Engineering*, *38*(1), 17.

Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ: British Medical Journal*, *312*(7023), 71–72. doi:10.1136/bmj.312.7023.71 PMID:8555924

Schneider, E. C., Sarnak, D. O., Squires, D., Shah, A., & Doty, M. M. (2017). Mirror Mirror 2017: International Comparison Reflects Flaws and Opportunities for Better US Health Care. Retrieved from https://interactives.commonwealthfund.org/2017/july/mirror-mirror/

Senthilkumar, S., Rai, B. K., Meshram, A. A., Gunasekaran, A., & Chandrakumarmangalam, S. (2018). Big Data in Healthcare Management: A Review of Literature. *American Journal of Theoretical and Applied Business*, *4*(2), 57–69. doi:10.11648/j.ajtab.20180402.14

Social Security Administration. (n.d.a). Health IT Partners. Retrieved from https://www.ssa.gov/hit/partners.html

Social Security Administration. (n.d.b). SSAgov/HealthIT. Retrieved from https://github.com/ssagov/healthit

Statista. (2018). Mobile Internet - Statistics & Facts. Retrieved from https://www.statista.com/topics/779/mobile-internet/

Tandon, A., Murray, C. J., Lauer, J. A., & Evans, D. B. (2000). *Measuring overall health system performance for 191 countries*. Geneva: World Health Organization.

Tanner, A. (2017). *Our bodies, our data: How companies make billions selling our medical records*. Beacon Press.

Terhaar, M. F. (2018). Value-Based Purchasing. *Clinical Analytics and Data Management for the DNP*, *1*, 27.

The Office of the National Coordinator for Health IT. (n.d.). What is health information exchange? Retrieved from https://www.healthit.gov/faq/what-health-information-exchange

The Organization for Economic Co-operation and Development. (2018). *Spending on Health: Latest Trends*. Retrieved from http://www.oecd.org/health/health-systems/Health-Spending-Latest-Trends-Brief.pdf

Wang, C. (2018). Integrating data analytics & knowledge management: A conceptual model. *Issues in Information Systems*, *19*(2), 208–216.

Wang, L., Wang, G., & Alexander, C. A. (2015). Big data and visualization: Methods, challenges and technology progress. *Digital Technologies*, *1*(1), 33–38.

Wang, Y., Kung, L., & Byrd, T. A. (2016). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*. doi:10.1016/j.techfore.2015.12.019

Yang, Y., Yu, S., Hao, Y., Xu, X., & Liu, H. (2016). The SWOT Analysis of the Wearable Devices in Smart Health Applications. *Paper presented at the International Conference on Smart Health*.

Chaojie Wang is a seasoned software engineer, systems architect, data scientist, researcher, and project manager with over thirty-year experience in industry, government, and academia. Dr. Wang currently works for The MITRE Corporation, an international thinktank and operator of several Federally Funded Research and Development Centers (FFRDC). In his capacity as a principal systems engineer, Dr. Wang advises the federal government on IT Acquisition & Modernization, Data Analytics & Knowledge Management, and Emerging Technology Evaluation & Adoption. Previously, Dr. Wang worked for Lockheed Martin Corporation and participated in multiple large-scale IT projects for the federal government. Dr. Wang is a certified Project Management Professional (PMP), a certified SAFe Agilist (SA), and currently serves on the Editorial Review Board for the International Journal of Patient-Centered Healthcare (IJPCH) by IGI Global and the Issues in Information System (IIS) by the International Association for Computer Information Systems (IACIS). Dr. Wang holds a Bachelor of Engineering in MIS from Tsinghua University, a Master of Art in Economics and a Master of Science in Statistics both from the University of Toledo, an MBA in Finance from Loyola University Maryland, and a Doctor of Science in Information Systems and Communications from Robert Morris University.