Thomas Jefferson University

Jefferson Digital Commons

School of Continuing and Professional Studies Student Papers

School of Continuing and Professional Studies Papers

8-12-2020

Big Data Analytics Applied to Healthcare

Xuejuan Zhang
Thomas Jefferson University

Boris Vishnevsky
Thomas Jefferson University

BIG DATA ANALYTICS APPLIED TO HEALTHCARE

Xuejuan Zhang[1] and Boris Vishnevsky[2]

Doctor of Management in Strategic Leadership

School of Continuing and Professional Studies

Thomas Jefferson University

Philadelphia, PA 19144

In this paper, we review the recent (primarily from 2018 - 2020) literature related to Big Data Analytics (BDA). We also discuss ways of applying BDA in Healthcare. In Section 1, we discuss the definition of Big Data Analytics and its characteristics. In section 2, we discuss the healthcare ecosystem's main stakeholders and the data of each main stakeholder (i.e., healthcare provider, healthcare payer, and patient/consumer). Section 3 discusses the challenges and opportunities of leveraging Big Data Analytics by healthcare stakeholders.

1. Definition of Big Data Analytics and Its Characteristics

Several Big Data definitions have been suggested in the literature as efforts evolve to understand this new field. In this review, we use the consensus definition proposed by the National Institute of Standards and Technology (NIST, https://www.nist.gov/). NIST has been leading collaboration between industry, academia, and government since 2013, and launched the NIST Big Data Public Working Group (NBD-PWG).

NIST (2019) defines Big Data as a large amount of data in the networked, digitized, sensor-laden, information-driven world. Moreover, the characteristics of Big Data that force new architectures can be described in terms of "V"s as follows.

- Volume denotes the size of the dataset
- Velocity is the rate of the flow of data
- Variety refers to the multitude of different repositories of data
- Variability concerns the change in speed or structure

NIST (2019) uses these characteristics as fundamental drivers that define the overall design of a Big Data system. This foundation supports several different data system architectures with multiple analytics life cycles that enable the system's outcomes, including performance-effectiveness, function-appropriateness, and cost-efficiency.

In addition to the 4"V" BDA characteristics, the Big Data may carry additional features also referred to as "V" attributes:

---

[1] This paper is based on an advanced independent study by Doctoral Candidate Xuejuan Zhang. Contact: Xuejuan.Joyce.Zhang@Jefferson.edu
[2] Supervised by Doctoral Faculty Boris Vishnevsky. Contact: Boris.Vishnevsky@Jefferson.edu.

- Veracity is the accuracy, conformity of the data, or description for data quality issues
- Validity is the appropriateness of the data for its intended use
- Volatility is the tendency for data structures to change over time
- Visualization is the usage of charts and graphs to visualize large amounts of complex data as an effective way to convey the meaning of data
- Value is the monetary value or business benefits achieved by an organization or individual who acquires, analyzes, or distributes the data

BDA concerns the application of advanced analytic techniques to big data sets, which has led some researchers to use the words big data and big data analytics interchangeably. The development of digital technologies and advanced analytical tools, such as the Internet of Things (IoT) or Networks of Things (NoT), Artificial Intelligence (A.I.), and Machine Learning (ML), drive the growth of big data and big data analytics. Based on the purpose or the roles in the BDA, new technologies may be categorized into three main groups: techniques for acquiring and analyzing data, technologies for storing and sharing data, and technologies for data communication (Table 1).

Table 1: Digital Technologies Related to Big Data Analytics

Main Purpose	Technologies
Data Acquisition and Analysis	A.I., IoT/NoT
Data Storage / Data Sharing	Blockchain; Cloud computing
Data Communication	Mobile devices with apps; social network (Facebook, Twitter) / messengers (WhatsApp); Video and Image platforms (YouTube, Instagram)

Internet of Things (IoT) commonly refers to a connected world in which every element or 'thing' sends and receives information through sensors. NIST (2016) uses the term Network of Things (NoT) as a synonym of IoT and argues that IoT is an instantiation of NoT, e.g., IoT has its 'things' tethered to the Internet. NIST (2016) provides an NoT model based on IoT-based technologies' underlying and foundational science because the realization of IoT involves sensing, computing, communication, and actuation. In the NoT model, NIST uses the term *primitive* to represent smaller pieces from which larger blocks or systems can be built rather than focus on a definition for what is or is not a 'thing.' The primitives of NoT are called the sensor, aggregator; communication channel; external utility (e-Utility), and decision trigger.

Artificial Intelligence (A.I.) and A.I. technologies are the software and/or hardware systems that can learn to solve complex problems and make predictions. A.I. systems can also undertake tasks that require human-like sensings such as vision, speech, and touch, perception, cognition, planning, learning, communication, or physical action (NIST, 2019b). Examples include A.I. assistants (e.g., ChatBots), computer vision systems, unattended vehicle systems, and facial recognition systems.

Machine Learning (ML) refers to the components of A.I. systems that learn from data to perform tasks that require human intelligence. The ML components of an A.I. system include the data, model, and processes for training, testing, and validation (NIST, 2019c).

Blockchains are tamper-evident and tamper-resistant digital ledgers implemented in a distributed fashion (i.e., without a central repository) and usually without a central authority such as a bank, company or government (NIST, 2018). This technology became widely known in 2009 with the launch of the Bitcoin network, the first of many modern cryptocurrencies.

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of virtualized and configurable computing resources (e.g., networks, servers, storage, applications, and services). These cloud-based computing resources can be rapidly provisioned and released with minimal management effort or service provider interaction (NIST, 2011).

2. The Healthcare Ecosystem and Major Stakeholders

Healthcare is a complex system consisting of interdependencies among many stakeholders often with different interests that must be viewed as an ecosystem according to Woolf and Aron (2013-4). The primary stakeholder groups of a healthcare system include health professionals (physicians and nurses), health facilities (clinics, hospitals, urgent care), financing institutions or payers (insurance companies and other third-party payers), patients or consumers, and social services. Health professionals can come from various health sectors including dentistry, medicine, midwifery, nursing, psychology, physiotherapy, and many others (Dash, Shakyawar, Sharma & Kaushik, 2019).

Of the characteristics of big data (4 Vs) to healthcare, variety could be a defining feature because stakeholders of the healthcare ecosystem can generate a variety of data in different forms. Table 2 defines the main domains of data generated by healthcare providers, payers, patients, and social services.

Table 2: Major Types of Data by Major Healthcare Stakeholder

Healthcare Stakeholder	Major Forms of Data
Provider	Electronic Healthcare Record
Payer	Electronic Medical Record/Insurance Record
Patient/Consumer	Personal health record / user-generated health and health-related data
Social Service	HHS, CMS, Medicare, APCDs generate, aggregate, and curate large volumes of health care, social service, and performance data.

Healthcare providers information related to the medical history of a patient including medical diagnoses, medical imaging, pharmacy prescriptions, demographics, clinical narratives, and the results of various laboratory tests. This test information is increasingly stored, managed, and shared in a digital form known as Electronic Health Records (EHR). The adoption of EHRs was slow at the beginning of the 21st century, but it has grown substantially after 2009. HER provides several advantages to healthcare providers such as easy access to the entire medical history of a patient, faster reporting of quality indicators, better continuity of care, reduction of medical errors and greater efficiency related to electronic authorization and immediate insurance approval due to less paperwork (Dash et al., 2019). Like HER, an electronic medical record (EMR) stores the standard medical and clinical data gathered from patients.

The payers (insurance companies) commonly embrace big data more proactively than healthcare providers. Indeed, digitalization has revolutionized the insurance operation. Big data analytics has enormous influences on insurance value chains, including underwriting, rating, marketing, product design, claim management, and customer services.

Besides the traditional EMR and insurance record, insurance companies are developing capabilities to collect patient/consumer data at a more granular level. Over the past two decades, insurance companies have increasingly begun to deploy data from third-party data sources. Two new sources of data are relevant in the context of insurance (Geneva Association, 2018). The first source concerns data automatically generated and stored with people's online behavior, such as personal information shared via social media, individual online shopping behavior, and data generated by personalized search and browsing activities. Personal data captured as part of the online activities can reveal information about the habits and lifestyle, and complement data traditionally used by insurance companies. The collection of such data is highly concentrated with digital-first companies such as Alibaba, Alphabet (Google), Amazon, Facebook, and Microsoft. The second new source of information is captured by sensors built into various appliances and consumer goods, such as the wearable devices (i.e., Apple watch or fitness trackers). These, combined with traditional data such as claims record, medical record, or family member conditions, are increasingly used as inputs of automated predictive modeling and decision-making processes leveraging A.I. or ML.

At the patient level, the development of technologies allows individuals to build personal health records (PHR) in electronic format. More importantly, individuals can intentionally and unintentionally generate more data that can be used to infer their health condition. Notable examples include the data generated by fitness trackers and wearable devices and data created on the social network containing personal socio-behavior information, such as individual financial conditions, mental and physical status, diets, exercising records, and unique lifestyles.

Social services play an essential role in the healthcare ecosystem by extending and amplifying health care services, e.g., Medicare and Medicaid. Social services also affect the so-called social determinants of health (SDH). The development of BDA technologies enables the digitalization of social services in combination with the delivery of healthcare. The integration of Social and Health Care delivery services represents a significant challenge as well as an opportunity for improving the value of health care while reducing its cost.

3. Value Creation of Big Data Analytics for Healthcare Providers and Consumers

3.1 Healthcare Providers

Table 3 summarizes the significant opportunities and challenges of healthcare providers (see Raghupathi & Raghupathi, 2016; Kruse et al., 2016; Mehta, & Pandit, 2018).

Table 3: Opportunities and challenges of BDA to Healthcare Providers

Opportunities	Challenges
• Improve the quality of care • Early detection of diseases • Improve decision making • Cost reduction • Patient-centric care	• Data structure • Data security • Data standardization • Managerial issues • Regulatory compliance

There are significant opportunities for BDA use by healthcare providers. One is to improve the quality of care: the most certain benefit offered by BDA is to improve outcomes and improve the health-related quality of life because BDA provides the healthcare provider with the ability to predict outcomes better and help medication adherence at the patient's end.

A second is for early detection of diseases: BDA allows for the early detection of diseases, which helps in clinical objectives related to achieving improved treatments and higher patient outcomes. BDA can also help in the prevention and personalized disease management and monitoring through tracking healthy behaviors and assisting patients in monitoring their conditions.

BDA can also improve decision making: It enables the use of evidence-based medicine and helps healthcare providers to make more informed decisions because BDA can help to optimize the decision-making process by incorporated accurate and up-to-date information.

There can be significant cost reduction: savings can be realized through more cost-effective treatments and monitoring to improve medication adherence. Moreover, the cost of data-intensive tasks can be reduced due to the decrease in computing and communication expenses.

Improved patient-centric care can be realized: BDA plays an essential role in changing the health care sector from disease-centric care toward patient-centric care. It will allow patients to play an active part in their care by providing them timely and appropriate information and through active communication between patients and providers.

There are significant challenges of BDA when applied to healthcare providers. One concerns the nature of the data structure: most data in healthcare are unstructured such as natural language; as well, data such as HER are often fragmented and not well shared across different providers.

A second challenge is data security: there are considerable privacy concerns regarding the data given existing legislation and regulation. Moreover, because of the sensitivity of healthcare data, concerns arise when specific data sets are made available as an opensource platform or shared over the vendors on the value chains.

Data standardization is important because in the fragmented EHR platform, data are stored in formats that are hard to be compatible with various applications and technologies.

Managerial issues include lack of digital-first mindset and required analytical skills that may limit the adoption of BDA

Finally, regulatory compliance is a challenge: providers using a high volume of sensitive healthcare data will have to comply with existing regulations specific to healthcare. Moreover, providers will also need to pay attention to emerging rules for BDA and personal data.

The phrase *digital-first healthcare organization* means that organizations need to rethink how business gets done and their value positions. It also refers to the revolutions in the product and consumer experiences which informs the need to optimize existing business processes with new digital methods. Deloitte (2018) suggests that healthcare providers should implement a *use case-driven mentality* and focus on delivering value. Healthcare providers should adopt a design approach allowing rapid prototyping, testing, and piloting a variety of innovative business processes with customers and partners.

Wang et al. (2018) propose a BDA-enabled transformation model from the practice-based perspective. In this model, they argue that healthcare should develop and enhance capabilities based on the three main architectural components of BDA: data aggregation, data analysis, and data interpretation. Accordingly, healthcare providers shall develop three main BDA capabilities: traceability (related to data aggregation), analytical solutions (related to data analysis), and decision support capability (related to decision support capability). Using a case study approach that examines the relevant practices of healthcare providers, they reported that healthcare providers put a different level of priorities on those capability developments. The development of analytical solutions is given the highest priority, whereas the traceability is developed with the lowest priority.

3.2 Healthcare Payers

Table 4 summarizes the opportunities and challenges for healthcare payers (i.e., health insurance companies) drawn from Eling and Lehmann (2018) and McKinsey (2019). Traditionally, health insurance companies' business models are built on data such as historical medical claims data. However, the advances of new technologies in BDA, A.I., and IoT have fundamentally changed the role of data in the insurance business model and revolutionized the operations of insurance companies. For instance, the significant areas of insurance value chains, such as underwriting, product development, marketing, distribution, and claims management, are all affected by BDA (Geneva Association, 2018; Eling & Lehmann, 2018).

Table 4: Opportunities and challenges of BDA to Healthcare Payers

Opportunities	Challenges
• Tailored insurance product • Strengthen partnerships with providers • Fully automated insurance process with increased efficiencies • Better engagement of customers.	• Disruptions by technology start-ups and technology titans • Lack of skills • Regulatory compliance.

Regarding challenges, healthcare payers face both internal and external challenges related to BDA. Internal problems include the lack of skills for collecting, analyzing data, and making informed business decisions accordingly. Moreover, healthcare insurers also face more stringent regulatory requirements related to the enhanced use of personal data and the use of A.I. and ML in the decision-making process. The primary external challenge faced by health insurance companies is whether they will lose part or entire value chain to technology start-ups and technology titans. Data are essential for the operation of health insurance companies, and advantages of the technology include broad access to customers, respective data, and advanced data analytical tools. BDA might take over part or even the entire insurance value chain and issue their insurance product if the risk margin of being a producer were high enough (Eling & Lehmann, 2018). Indeed, technology-driven firms such as Tencent and Alibaba have already begun to start their joint ventures in property and life insurance business.

Similar to healthcare providers, healthcare payers need to respond to the influences of BDA and digitalization. Eling and Lehmann (2018) suggest that insurance companies will have to invest in IT and additional employee skills to integrate new technologies and innovative models within their organizations. Moreover, insurance companies also need to carefully evaluate the benefits and costs of running technological incubators and cooperating with technology-driven firms. McKinsey (2019) suggests that there shall be three core business capabilities in the digital health ecosystem: First are consumer components (e.g., personalized health services) provide digital access points for consumers with tailored experiences focused on individual needs. Second are provider components (e.g., integrated EHR) provide an integrated payer-provider data exchange platform and distributed data management among payers and providers. Third are internal components (automated digital core) provide technology foundations for the flexible data layer, sophisticated analytics, and automation.

3.3 Patients

Deloitte (2018) summarizes the benefits of the patient by adopting BDA as patient-centric healthcare that enables better health outcomes for consumers, provide simplified access to healthcare products and solutions, reduce costs, and increase transparency (Table 5).

Table 5: Opportunities and challenges of BDA to Patients

Opportunities	Challenges
• Patient-centric healthcare • Enhanced personalized medicine	• Data privacy • Potential Discriminations and insurability of health risk

More relevant to the context of the U.S. health care system are the recent advancements of value-based care models or programs. The Centers for Medicare & Medicaid Services (CMS) define value-based programs as those that reward health care providers with incentive payments based on the quality, rather than the quantity, of care provided to patients and that support a three-part aim of better care for individuals; better health for populations; and lower cost. [3] CMS's value-based programs include the End-Stage Renal Disease Quality Incentive Program, Hospital Value-Based Purchasing (VBP) Program, Physician Value-based Modifier (PVBM) program, and Hospital-Acquired Conditions Reduction Program. The goal of those programs is to link provider performance of quality measures to provider payment. For instance, the Hospital VBP program rewards acute care hospitals with incentive payments based on the quality of care provided in the inpatient hospital setting. Hospitals are assessed on measures of outcome such as mortality and complications, healthcare-associated infections, patient safety, patient experience, process, and efficiency and cost reduction. The program will adjust a part of hospitals' Medicare payments based on each measure of the hospitals' total performance.[4] These endeavors for promoting value-based care create massive demand for tracking, storing, sharing, analyzing, and validating the adopted quality measures. BDA will contribute significantly to these.

The potential challenges of BDA may include data privacy, issues related to discrimination, and the insurability of health risk to the patient. Concerns over data privacy related to the enhanced use of personal data by BDA are not new. Most Western countries provided protections regarding the collection and use of data through existing public policies and regulations. For instance, in the U.S., the Health Insurance Portability and Accountability of 1996 (HIPPA) governs healthcare data. But the adoption and practice of big data analytics impose new challenges because HIPPA does not regulate all the relevant health information. HIPPA does not govern health information that is generated by entities not covered by HIPPA (e.g., life insurance companies), or non-health information on which inferences about health are based, or user-generated health information. To address this, the European Union introduced the General Data Protection Regulation (GDPR) in 2018. GDPR strengthens the rights of individuals over their data by requiring "unambiguous consent" to pass personal data and introduce the right to be forgotten and data portability (Geneva Association, 2018). However, the lack of a unified personal data protection standard could cause an increase in regulatory compliance costs.

[3] https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Value-Based-Programs. Accessed on May 1, 2020.
[4] https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HVBP/Hospital-Value-Based-Purchasing. Accessed on May 1, 2020.

Moreover, the enhanced personal data use and automated decision-making process adopted by health insurance companies may contribute to fairness and discrimination problems for patients. Business models based on digital monitoring can either reward or penalize certain behaviors or lifestyle choices that are labeled "good" or "bad" by the insurance companies. Such models may be regarded as intrusive or "paternalistic" and as interring with an individual's autonomy in decision making (Geneva Association, 2018). Individuals may have to make choices such as to restrict their lifestyle options - eating less sugar and moving more often - because they cannot pay health insurance for high-risk behaviors. Furthermore, decision-making by A.I. or ML may not always be transparent and explainable. Sometimes it is difficult to determine which factor in the predictive model plays a determining role. Regulators and policymakers around the world have begun to respond to the regulatory and ethical challenges imposed by BDA.

One example is that Singapore introduced an A.I. governance framework in 2019 built on two main guiding principles: (1) decisions made by or with the help of A.I. shall be transparent, explainable, and fair; (2) A.I. systems shall be human-centric and safe. In the U.S., the National Association of Insurance Commissioners (NAIC) has tasked its Casualty Actuarial and Statistical Task Force (CASTF) to identify best practices for rating plans. These recommendations will serve as guidance to state insurance departments (and insurers) in their review of complex models that underly rating plans in personal property insurance, such as auto and homeowner insurance. A similar approach and practice can also be applied to health-related insurance in the future.

3.4 Social Services

Many health care services are delivered through various social services (e.g., Medicare in the U.S.). As presented in Table 6, BDA offers opportunities for a new form of social services characterized as "citizen-centered service delivery," which focuses on individual needs instead of specific services provided by siloed government divisions (e.g., homeless, unemployed, disabilities, and social needs). Modern information technologies can digitalize administrative activities in one consolidated information system. These systems will leverage data collection and decision-making components to provide a single source for citizen data, automated eligibility determination, and electronic document management. BDA can help to improve outcomes for citizens through cohesive delivery of social services while reducing the related costs at the same time.

Table 6: Opportunities and challenges of BDA to Social Services

Opportunities	Challenges
• Improved service delivery for citizens • Enhanced personalized medicine • Improvements in Public Health outcomes • Transition to value-based health care delivery • A more effective response to pandemics • Reduction in fraud	• Digitization of the workforce • Data integrity • Funding models • Government initiatives

As part of their continuous focus to transition to value-based care, the Department of Health and Human Services (HHS), the Centers for Medicaid (CMS), and the Department of Justice (DOJ), in partnership with the private sector, continue their efforts to eliminate wasteful spending, fraud, and abuse in Medicare and Medicare programs. All-Payer Claims Databases (APCDs) aggregate a broad range of health care data that is reported by the commercial sector along with Medicare and Medicaid providers, as described by Khan (2016-Part 2) these APCD data offer a wide array of information, including the utilization and cost related to the efficient delivery of quality health care. The state-level APCD data sets can be utilized to improve the recovery rates and overall effectiveness of the Health Care Fraud and Abuse Control Program, HCFAC (2013). BDA applied to APCDs can also provide a unique insight in managing and reducing the impact of global pandemics such as COVID-19.

While initially formed to coordinate and administer Medicare and Medicaid Programs as the Health Care Financing Administration under HHS, CMS's scope has substantially expanded since the creation of the agency more than 52 years ago. To reduce wasteful spending and fight fraud and abuse, CMS and the DOJ worked together to enact the Health Care Fraud and Abuse Control (HCFAC) Program, and to form the Healthcare Fraud Prevention Partnership (HFPP). This partnership enables CMS to shift away from the "pay and chase" model to a more proactive fraud prevention strategy, applying modern technology and BDA-based anti-fraud capabilities to focus on provider screening and enrollment.

A study by Berwick and Hackbarth (2012) identified six categories of wasteful health care spending: (1) failure of care delivery, (2) failure of care coordination, (3) overtreatment, (4) administrative complexity, (5) pricing failures, and (6) fraud and abuse. APCD data aggregated at the state level enables direct and indirect (derived via analytics) insight to improve outcomes for these. According to the Bass and Berry (2018-2) review of HCFAC in 2017, a large percentage of fraud and abuse cases typically fall under the False Claims Act as related to "upcoding" and "unbundling" violations that without BDA could have only be captured as overtreatment or pricing failures.

The combined wasteful and fraudulent spending for Medicare and Medicaid in 2011 can be estimated between $133 billion (about 67% of total) and $262 billion (about 65% of total). Based on DOJ HCFAC (2013) report for the years 2013 – 2017, the annual recovery was from $3.1 billion to $6.7 billion range, representing only a 2-3% recovery rate. BDA could present a substantial opportunity for further improvements. Leveraging and expanding the focus of investigations and counter-fraud solutions to include state-level APCD data sets can improve the recovery rates in several ways.

One is because the state-level APCDs aggregated health care data is self-reported by Medicare/Medicaid as well as by commercially administered programs, the comparison of cost, utilization, and quality between Medicare/Medicaid and commercial programs can be obtained. This comparative analysis will quickly show any possible outliers, differences, or anomalies that can be further studied by investigative units to determine the root causes.

Another is that the state-level APCD data are self-reported and represent state-wide health care market-specific rates, cost, and utilization statistics that make it possible to identify the median value for these attributes as well as potential outliers. Because APCD data are state- and market-specific, counter-fraud systems over time can "learn" and maintain state- and market-specific data points and interpret these data points within a national health care marketplace, access, and utilization of state-specific APCDs data sets vary from state to state. Many U.S. states enable free-of-charge access to APCD data for research and non-commercial purposes. In contrast, the cost of APCD data for commercial use is determined individually based on the requester's role and objectives. If the requester leverages APCD data to improve health care utilization and value by reducing waste and abuse, this will lead to establishing a win-win outcome and the relationship between the requester and the state-owner of the APCD data. Similar win-win engagements with multiple states and APCDs would also contribute to and accelerate APCD standardization activities.

Using state-level APCD data, counter-fraud solutions providers can substantially enhance the effectiveness of their systems as well as the operational performance of special investigation teams while improving APCD standardization, integration, and utilization at a national level.

The relationship between Public Health, Medical Care Systems, and Public Health outcomes are quite complicated (Woolf & Aron, 2013-4). BDA can be leveraged to obtain more comprehensive insight into function and interdependencies across these systems.

Several challenges exist for such a transformation of social and public health services.
First, it has been shown that providing proper training for the digitalization of workforce and operation models with decades of history can be very challenging. Second, many people concern about the data integrity issue related to a consolidated database that contains every bit of citizen information. Third, political tensions can arise due to the budget/cost of building, such as an integrated social services system and the question of the effectiveness of such a new system.

4. Conclusion

In this paper we reviewd the recent literature related to Big Data Analytics (BDA) and its applications in Healthcare. We described how the development of technologies and analytical tools have greatly reshaped the landscape of Healthcare. Looking into the future, we expect the Healthcare sector will continue to witness the creative disruptions brought by the advancements of BDA. As the Healthcare sector is best described as an ecosystem with many entangled stakeholders, it is important to keep a holistic and dynamic view when examining the fast-evolving technologies and business applications related to BDA. Moreover, researchers and practitioners must be vigilant and cautious about new risks (e.g., cyber risk and regulatory compliance issues) associated with BDA. We urge that prior to adopting new technologies and new analytic tools that organizations consider comprehensive cost and benefit analysis and scenario planning to address unintended outcomes.

References

Bass, F.M., Berry, F. & Sims, C. 2018. Health Care Fraud and Abuse Review 2017. Bass, Berry, and Sims Health Care Task Force, 2018(2). Retrieved from: http://www.bassberry.com.

Berwick, D. M., & Hackbarth, A. D. 2012. The Six Categories of Wasteful Health Care Spending. The *Journal of the American Medical Association*, 307 (14): 1515.

Dash, S., Shakyawar, S. K., Sharma, M. & Kaushik, S. 2019. Big Data in Healthcare: Management, Analysis and Future Prospects. *Journal of Big Data*, 6, 54. Retrieved from: https://doi.org/10.1186/s40537-019-0217-0.

Deloitte. 2018. The Digital Imperative: The Imperative for a Consumer-Centric, Digitally Enabled Health Ecosystem. Retrieved from: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-lshc-the-digital-imperative.pdf

Eling, M., & Lehmann, M. 2018. The Impact of Digitalization on the Insurance Value Chain and the Insurability of Risks. *Geneva Papers on Risk and Insurance – Issues and Practice*, 43: 359-396.

Galbraith, J. R. 2014. Organizational Design Challenges Resulting from Big Data. *Journal of Organization Design*, 3(1): 2-13.

Geneva Association. 2018. Big Data and Insurance: Implications for Innovation, Competition, and Privacy. Retrieved from: https://www.genevaassociation.org/research-topics/cyber-and-innovation-digitalization/big-data-and-insurance-implications-innovation

Grossman, R. L. &Siegel, K. P. 2014. Organizational Models for Big Data and Analytics. *Journal of Organization Design*, 3(1): 20-25.

Grover, V., Chiang, R. H. L., Liang, T. & Zhang, D. 2018. Creating Strategic Business Value from Big Data Analytics: A Research Framework. Journal of Management Information Systems, 35(2): 388-423. Retrieved from: https://doi.org/10.1080/07421222.2018.1451951.

Groves, P., Kayyali, B., Knott, D. & Kuiken, S.V. 2013. The 'Big Data' Revolution in U.S. Healthcare Accelerating Value and Innovation. McKinsay&Company. Retrieved from: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care

HCFAC. 2013. Annual Report for Fiscal Year 2013.The Department of Health and Human Services and The Department of Justice Health Care Fraud and Abuse Control Program, Retrieved from: https://www.oig.hhs.gov/publications/docs/hcfac/FY2013-hcfac.pdf

Kahn J.. 2016. Written CMS Testimony 2016(Part 2). Washington (D.C.). Retrieved from: https://www.ncvhs.hhs.gov/wp-content/uploads/2016/05/Panel-2-Jessica-Kahn-CMS-Written-20160June17.pdf

Kruse, C. S., Goswamy, R., Raval, Y. & S. Marawi, S. 2016. Challenges and Opportunities of Big Data in Health Care: A Systematic Review. JMIR Med Inform 4(4), e38. Retrieved from: http://medinform.jmir.org/2016/4/e38/.

McKinsey. 2019. Digital Health Ecosystems: A Payer Perspective. Retrieved from: https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/digital-health-ecosystems-a-payer-perspective

Mehta, N. & Pandit, A. 2018. Concurrence of Big Data Analytics and Healthcare: A Systematic Review. *International Journal of Medical Informatics*, 114: 57-65.

National Association of Insurance Commissioners. 2019. Regulatory Review of Predictive Models. Retrieved from: https://naic.org/documents/comte_c_catf_predictive_model_white_paper_190510_clean.docx

National Institute of Standards and Technology (NIST). 2016. Networks of 'Things. Retrieved from: http://dx.doi.org/10.6028/NIST.SP.800-183

NIST. 2011. The NIST Definition of Cloud Computing. Retrieved from: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

NIST. 2018. Blockchain Technology Overview. Retrieved from: https://doi.org/10.6028/NIST.IR.8202.

NIST. 2019a. NIST Big Data Interoperability Framework. NIST Special Publication 1500-1r2. Retrieved from: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r2.pdf

NIST. 2019b. U.S. Leadership in A.I.: A Plan for Federal Engagement in Developing Technical Standards and Related Tools. Retrieved from: https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.

NIST. 2019c. A Taxonomy and Terminology of Adversarial Machine Learning. Draft NISTIR 8269. Retrieved from: https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8269-draft.pdf.

Pasquale, F. & Ragone, T. A. 2017. Protecting Health Privacy in an Era of Big Data Processing and Cloud Computing. *Standford Technology Law Review*, 17: 595-654.

Pearlman, A. 2019. HIPAA is the Tip of the Iceberg When It Comes to Privacy and Your Medical Data. Retrieved from: https://blog.petrieflom.law.harvard.edu/2019/01/07/hipaa-is-the-tip-of-the-iceberg-when-it-comes-to-privacy-and-your-medical-data/

Price, W. N. & Cohen, G. 2019. Privacy in the Age of Medical Big Data. *Nature Medicine*, 2:37-43.

Raghupathi, W. & Raghupathi, V. 2014. Big Data Analytics in Healthcare: Promise and Potential. *Health Information Science and Systems*, 2:3. Retrieved from: http://www.hissjournal.com/content/2/1/3.

Shahbaz, M., Gao, C. & Hu, Y. 2019. Investigating the Adoption of Big Data Analytics in Healthcare: the Moderating Role of Resistance to Change. *Journal of Big Data*, 6, 6. Retrieved from: https://doi.org/10.1186/s40537-019-0217-y.

Wang, Y. & Hajli, N. 2017. Exploring the Path to Big Data Analytics Success in Healthcare. *Journal of Business Research*, 70: 287-299.

Wang, Y., Kung, L., Wang, Y. C. & Cegielski, C. G. 2018. An Integrated Big Data Analytics-Enabled Transformation Model: Application to Health Care. *Information & Management*, 55(1): 64-79.

National Research Council Institute of Medicine. 2013. U.S. Health in International Perspective: Shorter Lives, Poorer Health, Woolf S.H., & Aron L. (Eds). Washington, DC: National Academies Press. Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK154484/