

Wang Y, Kung K, Byrd TA.

[Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations.](#)

Technological Forecasting and Social Change 2016

DOI: <http://dx.doi.org/10.1016/j.techfore.2015.12.019>

Copyright:

© 2016. This manuscript version is made available under the [CC-BY-NC-ND 4.0 license](#)

DOI link to article:

<http://dx.doi.org/10.1016/j.techfore.2015.12.019>

Date deposited:

22/09/2016

Embargo release date:

26 August 2017



This work is licensed under a

[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence](#)

Big Data Analytics: Understanding Its Capabilities and Potential Benefits for Healthcare Organizations

Yichuan Wang*

Raymond J. Harbert College of Business, Auburn University
405 W. Magnolia Ave., Auburn AL 36849, USA

Phone: +1-318-278-4630

E-mail: yzw0037@auburn.edu

LeeAnn Kung

Rohrer College of Business, Rowan University
201 Mullica Hill Road, Glassboro, NJ08028

E-mail: kung@rowan.edu

Terry Anthony Byrd

Raymond J. Harbert College of Business, Auburn University
405 W. Magnolia Ave., Auburn AL 36849, USA

Phone: +1-334-844-6543

E-mail: byrdter@auburn.edu

* Corresponding author

Big Data Analytics: Understanding Its Capabilities and Potential Benefits for Healthcare Organizations

Abstract

To date, health care industry has not fully grasped the potential benefits to be gained from big data analytics. While the constantly growing body of academic research on big data analytics is mostly technology oriented, a better understanding of the strategic implications of big data is urgently needed. To address this lack, this study examines the historical development, architectural design and component functionalities of big data analytics. From content analysis of 26 big data implementation cases in healthcare, we were able to identify five big data analytics capabilities: analytical capability for patterns of care, unstructured data analytical capability, decision support capability, predictive capability, and traceability. We also mapped the benefits driven by big data analytics in terms of information technology (IT) infrastructure, operational, organizational, managerial and strategic areas. In addition, we recommend five strategies for healthcare organizations that are considering to adopt big data analytics technologies. Our findings will help healthcare organizations understand the big data analytics capabilities and potential benefits and support them seeking to formulate more effective data-driven analytics strategies.

Keywords: Big data analytics, big data analytics architecture, big data analytics capabilities, business value of information technology (IT), health care

1. Introduction

Information technology (IT)-related challenges such as inadequate integration of healthcare systems and poor healthcare information management are seriously hampering efforts to transform IT value to business value in the U.S. healthcare sector (Bodenheimer, 2005; Grantmakers In Health, 2012; Herrick, Gorman, & Goodman, 2010; The Kaiser Family Foundation, 2012). The high volume digital flood of information that is being generated at ever-higher velocities and varieties in healthcare adds complexity to the equation. The consequences are unnecessary increases in medical costs and time for both patients and healthcare service providers. Thus, healthcare organizations are seeking effective IT artifacts that will enable them to consolidate organizational resources to deliver a high quality patient experience, improve organizational performance, and maybe even create new, more effective data-driven business models (Agarwal, Gao, DesRoches, & Jha, 2010; Goh, Gao, & Agarwal, 2011; Ker, Wang, Hajli, Song, & Ker, 2014).

One promising breakthrough is the application of big data analytics. Big data analytics that is evolved from business intelligence and decision support systems enable healthcare organizations to analyze an immense volume, variety and velocity of data across a wide range of healthcare networks to support evidence-based decision making and action taking (Watson, 2014; Raghupathi & Raghupathi, 2014). Big data analytics encompasses the various analytical techniques such as descriptive analytics and mining/predictive analytics that are ideal for analyzing a large proportion of text-based health documents and other unstructured clinical data (e.g., physician's written notes and prescriptions and medical imaging) (Groves, Kayyali, Knott, & Kuiken, 2013). New database management systems such as MongoDB, MarkLogic and Apache Cassandra for data integration and retrieval, allow data being transferred between traditional and new operating systems. To store the huge volume and various formats of data, there are Apache HBase and NoSQL systems. These

big data analytics tools with sophisticated functionalities facilitate clinical information integration and provide fresh business insights to help healthcare organizations meet patients' needs and future market trends, and thus improve quality of care and financial performance (Jiang et al., 2014; Murdoch & Detsky, 2013; Wang, Kung, Ting, & Byrd, 2015).

A technological understanding of big data analytics has been studied well by computer scientists (see a systemic review of big data research from Wamba, Akter, Edwards, Chopin, & Gnanzou, 2015). Yet, healthcare organizations continues to struggle to gain the benefits from their investments on big data analytics and some of them are skeptical about its power, although they invest in big data analytics in hope for healthcare transformation (Murdoch & Detsky, 2013; Shah & Patak, 2014). Evidence shows that only 42 percent of healthcare organizations surveyed are adopting rigorous analytics approaches to support their decision-making process; only 16 percent of them have substantial experience using analytics across a broad range of functions (Cortada, Gordon, & Lenihan, 2012). This implies that healthcare practitioners still vaguely understand how big data analytics can create value for their organizations (Sharma, Mithas, & Kankanhalli, 2014). As such, there is an urgent need to understand the managerial, economic, and strategic impact of big data analytics and explore its potential benefits driven by big data analytics. This will enable healthcare practitioners to fully seize the power of big data analytics.

To this end, two main goals of this study are: first, to identify big data analytics capabilities; and second, to explore the potential benefits it may bring. By doing so, we hope to give healthcare organization a more current comprehensive understanding of big data analytics and how it helps to transform organizations. In this paper, we begin by providing the historical context and developing big data analytics architecture in healthcare, and then move on to conceptualizing big data analytics capabilities and potential benefits in healthcare. We conducted a content analysis of

26 big data implementation cases in health care which lead to the identification of five major big data analytics capabilities and potential benefits derived from its application. In concluding sections, we present several strategies for being successful with big data analytics in healthcare settings as well as the limitations of this study, and direction of future research.

2. Background

2.1 Big data analytics: past and present

The history of big data analytics is inextricably linked with that of data science. The term “big data” was used for the first time in 1997 by Michael Cox and David Ellsworth in a paper presented at an IEEE conference to explain the visualization of data and the challenges it posed for computer systems (Cox & Ellsworth, 1997). By the end of the 1990s, the rapid IT innovations and technology improvements had enabled generation of large amount of data but little useable information in comparison. Concepts of business intelligence (BI) created to emphasize the importance of collection, integration, analysis, and interpretation of business information and how this set of process can help businesses make more appropriate decisions and obtain a better understanding of market behaviors and trends.

The period of 2001 to 2008 was the evolutionary stage for big data development. Big data was first defined in terms of its volume, velocity, and variety (3Vs), after which it became possible to develop more sophisticated software to fulfill the needs of handling information explosion accordingly. Software and application developments like Extensible Markup Language (XML) Web services, database management systems, and Hadoop added analytics modules and functions to core modules that focused on enhancing usability for end users, and enabled users to process huge amounts of data across and within organizations collaboratively and in real-time. At the same

time, healthcare organizations were starting to digitize their medical records and aggregate clinical data in huge electronic databases. This development made the health data storable, usable, searchable, and actionable, and helped healthcare providers practice more effective medicine.

At the beginning of 2009, big data analytics entered the revolutionary stage (Bryant, Katz, & Lazowska, 2008). Not only had big-data computing become a breakthrough innovation for business intelligence, but researchers also were predicting that data management and its techniques were about to shift from structured data into unstructured data, and from a static terminal environment to a ubiquitous cloud-based environment. Big data analytics computing pioneer industries such as banks and e-commerce were beginning to have an impact on improving business processes and workforce effectiveness, reducing enterprise costs and attracting new customers. In regards to healthcare industry, as of 2011, stored health care data had reached 150 exabytes (1 EB = 10^{18} bytes) worldwide, mainly in the form of electronic health records (Institute for Health Technology Transformation, 2013). However, most of the potential value creation is still in its infancy, because predictive modeling and simulation techniques for analyzing healthcare data as a whole have not yet been adequately developed.

More recent trend of big data analytics technology has been towards the use of cloud in conjunction with data. Enterprises have increasingly adopted a “big data in the cloud” solution such as software-as-a-service (SaaS) that offers an attractive alternative with lower cost. According to the Gartner’s 2013 IT trend prediction, taking advantage of cloud computing services for big data analytics systems that support a real-time analytic capability and cost-effective storage will become a preferred IT solution by 2016. The main trend in the healthcare industry is a shift in data type from structure-based to semi-structured based (e.g., home monitoring, telehealth, sensor-based wireless devices) and unstructured data (e.g., transcribed notes, images, and video). The

increasing use of sensors and remote monitors are key factors supporting the rise of home healthcare services, meaning that the amount of data being generated from sensors will continue to grow significantly. This will in turn improve the quality of healthcare services through more accurate analysis and prediction.

2.2 Big data analytics architecture

To reach our goals of this study which are to describe the big data analytics capability profile and its potential benefits, it is necessary to understand its architecture, components and functionalities. The first action taken is to explore best practice of big data analytics architecture in healthcare. We invited four IT experts (two practitioners and two academics) to participate in a five-round evaluation process which included brainstorming and discussions. The resulted big data analytics architecture is rooted in the concept of data life cycle framework that starts with data capture, proceeds via data transformation, and culminates with data consumption. Figure 1 depicts the proposed best practice big data analytics architecture that is loosely comprised of five major architectural layers: (1) data, (2) data aggregation, (3) analytics, (4) information exploration, and (5) data governance. These logical layers make up the big data analytics components that perform specific functions, and will therefore enable healthcare managers to understand how to transform the healthcare data from various sources into meaningful clinical information through big data implementations.

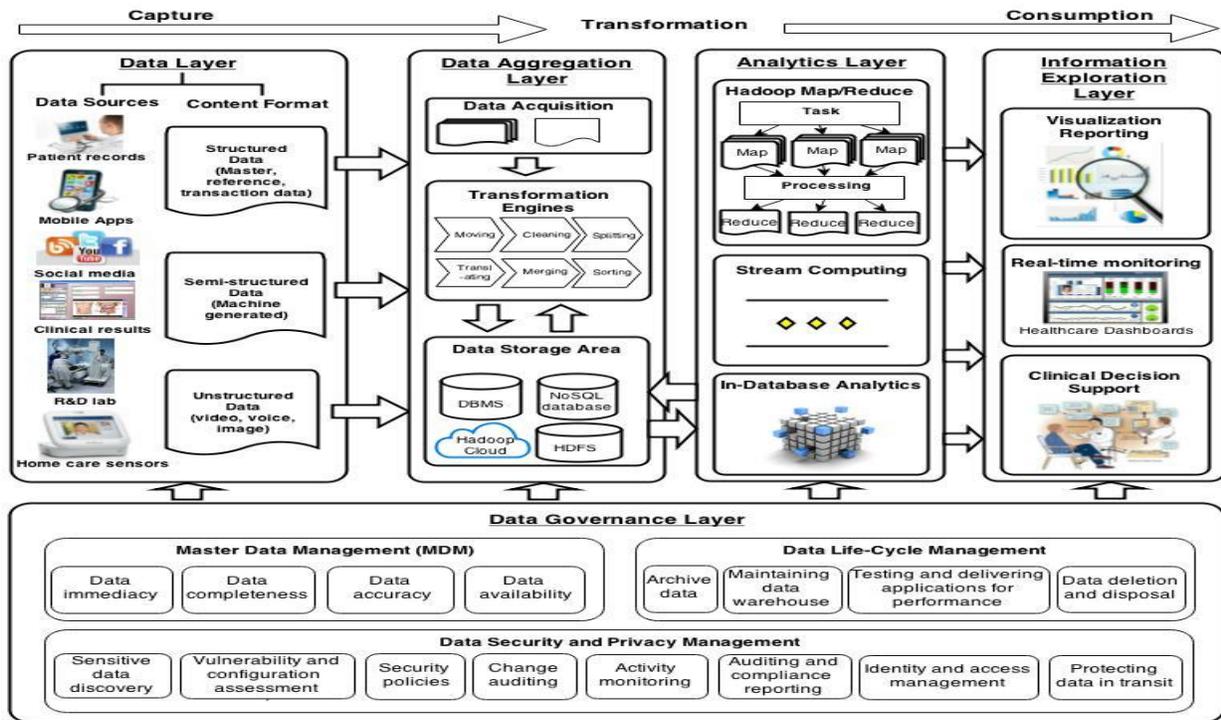


Figure 1. Big Data Analytics Architecture in Health Care

Data layer. This layer includes all the data sources necessary to provide the insights required to support daily operations and solve business problems. Data is divided into structured data such as traditional electronic healthcare records (EHRs), semi-structured data such as the logs of health monitoring devices, and unstructured data such as clinical images. These clinical data are collected from various internal or external locations, and will be stored immediately into appropriate databases, depending on the content format.

Data aggregation layer. This layer is responsible for handling data from the various data sources. In this layer, data will be intelligently digested by performing three steps: data acquisition, transformation, and storage. The primary goal of data acquisition is to read data provided from various communication channels, frequencies, sizes, and formats. This step is often a major obstacle in the early stages of implementing big data analytics, because these incoming data

characteristics might vary considerably. Here, the cost may well exceed the budget available for establishing new data warehouses, and extending their capacity to avoid workload bottlenecks. During the transformation step, the transformation engine must be capable of moving, cleaning, splitting, translating, merging, sorting, and validating data. For example, structured data such as that typically contained in an eclectic medical record might be extracted from healthcare information systems and subsequently converted into a specific standard data format, sorted by the specified criterion (e.g., patient name, location, or medical history), and then the record validated against data quality rules. Finally, the data are loaded into the target databases such as Hadoop distributed file systems (HDFS) or in a Hadoop cloud for further processing and analysis. The data storage principles are based on compliance regulations, data governance policies and access controls. Data storage methods can be implemented and completed in batch processes or in real time.

Analytics Layer. This layer is responsible for processing all kinds of data and performing appropriate analyses. In this layer, data analysis can be divided into three major components: Hadoop Map/Reduce, stream computing, and in-database analytics, depending on the type of data and the purpose of the analysis. *Mapreduce* is the most commonly used programming model in big data analytics which provides the ability to process large volumes of data in batch form cost-effectively, as well as allowing the analysis of both unstructured and structured data in a massively parallel processing (MPP) environment. *Stream computing* can support high performance stream data processing in near real time or real time. With a real time analysis, users can track data in motion, respond to unexpected events as they happen and quickly determine next-best actions. For example, in the case of healthcare fraud detection, stream computing is an important analytical tool that assists in predicting the likelihood of illegal transactions or deliberate misuse of customer

accounts. Transactions and accounts will be analyzed in real time and alarms generated immediately to prevent myriad frauds across healthcare sectors. *In-database analytics* refers to a data mining approach built on an analytic platform that allows data to be processed within the data warehouse. This component provides high-speed parallel processing, scalability, and optimization features geared toward big data analytics, and offers a secure environment for confidential enterprise information. However, the results provided from in-database analytics are neither current nor real time and it is therefore likely to generate reports with a static prediction. Typically, this analytic component in healthcare organizations is useful for supporting preventative healthcare practice and improving pharmaceutical management. The analytics layer also provides exceptional support for evidence based medical practices by analyzing EHRs, patterns of care, care experience, and individual patients' habits and medical histories.

Information exploration layer. This layer generates outputs such as various visualization reports, real-time information monitoring, and meaningful business insights derived from the analytics layer to users in the organization. Similar to traditional business intelligence platforms, reporting is a critical big data analytics feature that allows data to be visualized in a useful way to support users' daily operations and help managers to make faster, better decisions. However, the most important output for health care may well be its real-time monitoring of information such as alerts and proactive notifications, real time data navigation, and operational key performance indicators (KPIs). This information is analyzed from sources such as smart phones and personal medical devices and can be sent to interested users or made available in the form of dashboards in real time for monitoring patients' health and preventing accidental medical events.

Data governance layer. This layer is comprised of master data management (MDM), data life-cycle management, and data security and privacy management. This layer emphasizes the

"how-to" as in how to harness data in the organization. The first component of data governance, master data management, is regarded as the processes, governance, policies, standards, and tools for managing data. Data is properly standardized, removed, and incorporated in order to create the immediacy, completeness, accuracy, and availability of master data for supporting data analysis and decision making. The second component, data life-cycle management, is the process of managing business information throughout its lifecycle, from archiving data, through maintaining data warehouse, testing and delivering different application systems, to deleting and disposing of data. By managing data effectively over its lifetime, firms are better equipped to provide competitive offerings to meet market needs and support business goals with lower timeline overruns and cost. The third component, data security and privacy management, is the platform for providing enterprise-level data activities in terms of discovery, configuration assessment, monitoring, auditing, and protection (IBM, 2012). Due to the nature of complexity in data management, organizations have to face ethical, legal, and regulatory challenges with data governance (Phillips-Wren, Iyer, Kulkarni, & Ariyachandra, 2015). Particularly in healthcare industry, it is essential to implement rigorous data rules and control mechanisms for highly sensitive clinical data to prevent security breaches and protect patient privacy. By adopting suitable policies, standards, and compliance requirements to restrict users' permissions will ensure the new system satisfies healthcare regulations and creates a safe environment for the proper use of patient information.

2.3 Big Data Analytics Capability

Several definitions for big data analytics capability have been developed in the literature (see Table 1). In general, big data analytics capability refers to the ability to manage a huge volume of

disparate data to allow users to implement data analysis and reaction (Hurwitz, Nugent, Hapler, & Kaufman, 2013). Wixom, Yen, & Relich (2013) indicate that big data analytics capability for maximizing enterprise business value should encompass speed to insight which is the ability to transform raw data into usable information and pervasive use which is the ability to use business analytics across the enterprise. With a lens of analytics adoption, LaLalle, Lesser, Shockley, Hopkins, & Kruschwitz (2011) categorize big data analytics capability into three levels: aspirational, experienced, and transformed. The former two levels of analytics capabilities focus on using business analytics technologies to achieve cost reduction and operation optimization. The last level of capability is aimed to drive customer profitability and making targeted investments in niche analytics.

Moreover, with a view of adoption benefit, Simon (2013) defines big data analytics capability as the ability to gather enormous variety of data - structured, unstructured and semi-structured data - from current and former customers to gain useful knowledge to support better decision-making, to predict customer behavior via predictive analytics software, and to retain valuable customers by providing real-time offers. Based on the resource-based view, Cosic, Shanks, & Maynard (2012) define big data analytics capability as “the ability to utilize resources to perform a business analytics task, based on the interaction between IT assets and other firm resources (p. 4)”

In this study, we define big data analytics capability through an information lifecycle management (ILM) view. Storage Networking Industry Association (2009) describes ILM as “the policies, processes, practices, services and tools used to align the business value of information with the most appropriate and cost-effective infrastructure from the time when information is created through its final disposition (p. 2).” Generally, data regardless of its structure in a system has been followed this cycle, starting with collection, through repository and process, and ending

up with dissemination of data. The concept of ILM helps us to understand all the phases of information life cycle in business analytics architecture (Jagadish et al. 2014). Therefore, with a view of ILM, we define big data analytics capability in the context of health care as

-the ability to acquire, store, process and analyze large amount of health data in various forms, and deliver meaningful information to users that allows them to discover business values and insights in a timely fashion.

Table 1. The Definition of Big Data Analytics Capability from Prior Research

Sources	Viewpoints	Definitions
Cosic et al. (2012)	Resource based view	<ul style="list-style-type: none"> • The ability to utilize resources to perform a business analytics task
Hurwitz et al. (2013)	3V of big data	<ul style="list-style-type: none"> • The ability to manage a huge volume of disparate data to allow users to implement data analysis and reaction
LaLalle et al. (2011)	Analytics adoption	<ul style="list-style-type: none"> • Achieve cost reduction and operation optimization • Drive customer profitability and making targeted investments in niche analytics
Simon (2013)	Adoption benefit	<ul style="list-style-type: none"> • The ability to gather enormous variety of data from customers to gain business insights to optimize customer service
Trkman, McCormack, De Oliveira, & Ladeira (2010)	Business process	<ul style="list-style-type: none"> • Analytics in plan • Analytics in source • Analytics in make • Analytics in deliver
Wixom et al. (2013)	Business value	<ul style="list-style-type: none"> • Speed to insight • Pervasive use

2.4 Conceptualizing the potential benefit of big data analytics

To capture the potential benefits from big data analytics, a multidimensional benefit framework (See Table 2), including IT infrastructure benefits, operational benefits, organizational benefits, managerial benefits, and strategic benefits (Shang & Seddon, 2002) was used to classify

the statements related to the benefits from the collected 26 big data cases in health care. We choose Shang & Seddon's framework to classify the potential benefits of big data analytics for three reasons. First, our exploratory work is to provide a specific set of benefit sub-dimensions in the big analytics context. This framework will help us to identify the benefits of big data analytics into proper categories. Second, this framework is designed for managers to assess the benefits of their companies' enterprise systems. It has been refined by many studies related to ERP systems and specific information system (IS) architectures (Esteves, 2009; Gefen & Ragowsky, 2005; Mueller, Viering, Legner, & Riempp 2010). In this regard, this framework is suitable as a more generic and systemic model for categorizing the benefits of big data analytics system. Third, this framework also provides a clear guide for assessing and classifying benefits from enterprise systems. This guide also suggests the ways how to validate the IS benefit framework through implementation cases, which is helpful for our study.

Table 2. The Overview of Enterprise Systems' Multidimensional Benefit Framework

Benefit dimension	Description	Sub-dimensions
IT infrastructure benefits	Sharable and reusable IT resources that provide a foundation for present and future business applications	<ul style="list-style-type: none"> • Building business flexibility for current and future changes • IT cost reduction • Increased IT infrastructure capability
Operational benefits	The benefits obtained from the improvement of operational activities	<ul style="list-style-type: none"> • Cost reduction • Cycle time reduction • Productivity improvement • Quality improvement • Customer service improvement
Managerial benefits	The benefits obtained from business management activities which involve allocation and control of the firms' resources, monitoring of operations and supporting of business strategic decisions	<ul style="list-style-type: none"> • Better resource management • Improved decision making and planning • Performance improvement

Strategic benefits	The benefits obtained from strategic activities which involve long-range planning regarding high-level decisions	<ul style="list-style-type: none"> • Support for business growth • Support for business alliance • Building for business innovations • Building cost leadership • Generating product differentiation • Building external linkages
Organizational benefits	The benefits arise when the use of an enterprise system benefits an organization in terms of focus, cohesion, learning, and execution of its chosen strategies.	<ul style="list-style-type: none"> • Changing work patterns • Facilitating organizational learning • Empowerment • Building common vision

3. Research Methods

To reach our goals of this study, we used a qualitative approach, more specifically, a multiple cases content analysis to gain understanding and categorization of big data analytics capabilities and potential benefits derived from its application. The cases collection, approach and procedures for analyzing the cases are described in the following subsections.

3.1 Cases collection

Our cases were drawn from current and past big data projects material from multiple sources such as practical journals, print publications, case collections, and reports from companies, vendors, consultants or analysts. The absence of academic discussion in our case collection is due to the incipient nature of such in the field of healthcare. The following case selection criteria were applied: (1) the case presents an actual implementation of big data platforms or initiatives, (2) it clearly describes the software they introduce and benefits obtaining from the implementation. We excluded reports from one particular vendor due to their connection to one of our experts who were invited for the evaluation. We were able to collect 26 big data cases specifically related to the healthcare industries. Of these cases, 14 (53.8%) were collected from the materials released by

vendors or companies, 2 cases (7.7%) from journal databases, and 10 cases (38.4%) from print publications, including healthcare institute reports and case collections. Categorizing by region, 17 cases were collected from Northern America, 7 cases from Europe, and others from Asia-Pacific region. The cases we used are listed in Appendix A.

3.2 Research approach and process

We applied content analysis to gain insights from the cases collected. Content analysis is a method for extracting various themes and topics from text, and it can be understood as, “an empirically grounded method, exploratory in process, and predictive or inferential in intent.” Specifically, this study followed inductive content analysis, because the knowledge about big data implementation in health care is fragmented (Raghupathi, & Raghupathi, 2014). A three-phase research process for inductive content analysis (i.e., preparation, organizing, and reporting) suggested by Elo and Kyngäs (2008) was performed in order to ensure a better understanding of big data analytics capabilities and benefits in the healthcare context.

The preparation phase starts with selecting the “themes” (informative and persuasive nature of case material), which can be sentences, paragraphs, or a portion of a page (Elo & Kyngäs, 2008). For this study, themes from case materials were captured by a senior consultant who has over 15 years working experience with a multinational technology and consulting corporation headquartered in the United States, and currently is involved in several big data analytics projects. The senior consultant manually highlighted the textual contents that completely describe how a big data analytics solution and its functionalities create the big-data-enabled IT capabilities and potential benefits while reading through all 26 big data cases for a couple of times. Subsequently,

a total of 136 statements directly related to the IT capabilities and 179 statements related to the potential benefits were obtained and recorded in a Microsoft Excel spreadsheet.

The second phase is to organize the qualitative data emerged from phase one through open coding, creating categories and abstraction (Elo & Kyngäs, 2008). In the process of open coding, the 136 statements were analyzed by one of the authors, and then grouped into preliminary conceptual themes based on their similarities. The purpose is to reduce the number of categories by collapsing those that are similar into broader higher order generic categories (Burnard, 1991; Dey, 1993; Downe-Wamboldt, 1992). In order to increase the interrater reliability, the second author went through the same process independently. The two coders agreed on 84% of the categorization. Most discrepancies occurred between the two coders are on the categories of analytical capability. Disagreements were resolved after discussions and reassessments of the case to eventually arrive at a consensus. After consolidating the coding results, the two coders named each generic category of big data analytics capabilities using content-characteristic words.

4. Results

4.1 Capability Profile of Big data Analytics in Healthcare

Overall, the five generic categories of big data analytics capabilities we identified from 136 statements in our review of the cases are analytical capability for patterns of care (coded as part of 43 statements), unstructured data analytical capability (32), decision support capability (23), predictive capability (21), and traceability (17). These are described in turn below.

4.1.1 Analytical Capability for Patterns of Care

Analytical capability refers to the analytical techniques typically used in a big data analytics system to process data with an immense volume (from terabytes to exabytes), variety (from text to graph) and velocity (from batch to streaming) via unique data storage, management, analysis, and visualization technologies (Chen, Chiang, & Storey, 2012; Simon, 2013). Analytical capabilities in healthcare can be used to identify patterns of care and discover associations from massive healthcare records, thus providing a broader view for evidence-based clinical practice. Healthcare analytical systems provide solutions that fill a growing need and allow healthcare organizations to parallel process large data volumes, manipulate real-time, or near real time data, and capture all patients' visual data or medical records. In doing so, this analysis can identify previously unnoticed patterns in patients related to hospital readmissions and support a better balance between capacity and cost. Interestingly, analyzing patient preference patterns also helps hospitals to recognize the utility of participating in future clinical trials and identify new potential markets.

4.1.2 Unstructured data analytical capability

An analytical process in a big data analytics system starts by acquiring data from both inside and outside the healthcare sectors, storing it in distributed database systems, filtering it according to specific discovery criteria, and then analyzing it to integrate meaningful outcomes for the data warehouse, as shown in Figure 2. After unstructured data has been gathered across multiple healthcare units, it is stored in a Hadoop distributed file system and NoSQL database that maintain it until it is called up in response to users' requests. NoSQL databases support the storage of both unstructured and semi-structured data from multiple sources in multiple formats in real time. The core of the analytic process is the MapReduce algorithms implemented by Apache Hadoop.

MapReduce is a data analysis process that captures data from the database and processes it by executing “Map” and “Reduce” procedures, which break down large job objective into a set of discrete tasks, iteratively on computing nodes. After the data has been analyzed, the results will be stored in a data warehouse and made visually accessible for users to facilitate decision-making on appropriate actions.

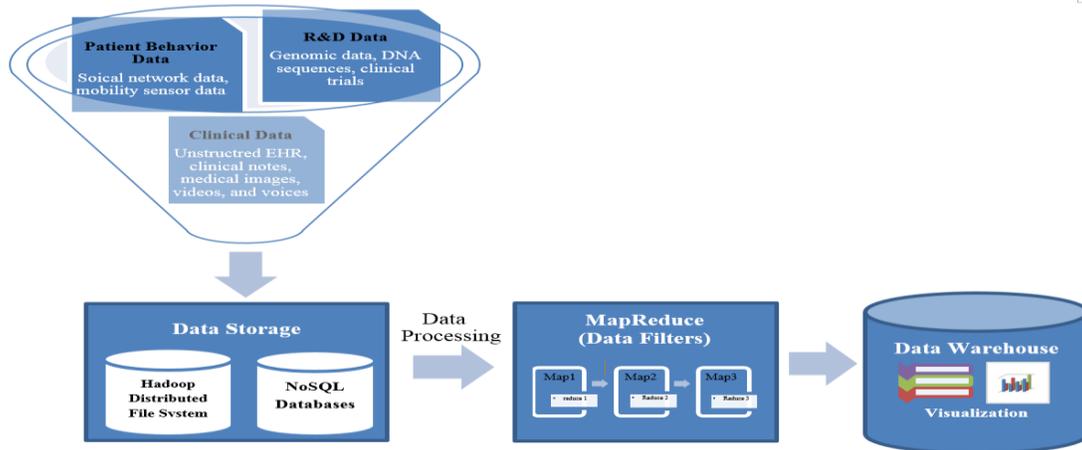


Figure 2. The Process of Analyzing Unstructured Data in Health Care Organizations

The main difference in analytical capability between big data analytics systems and traditional data management systems is that the former has a unique ability to analyze semi-structured or unstructured data. Unstructured and semi-structured data in healthcare refer to information that can neither be stored in a traditional relational database nor fit into predefined data models. Some examples are XML-based EHRs, clinical images, medical transcripts, and lab results. Most importantly, the ability to analyze unstructured data plays a pivotal role in the success of big data analytics in healthcare settings since 80% of health data is unstructured. According to a 2011 investigation by the TDWI research (Russom, 2011), the benefits of analyzing unstructured data capability are illustrated by the successful implementation of targeted marketing, providing

revenue-generating insights and building customer segmentation. One of our cases, Leeds Teaching Hospitals in the UK analyze approximately one million unstructured case files per month, and have identified 30 distinct scenarios where there is room for improvement in either costs or operating procedures by taking advantage of natural language processing (NLP). This enables Leeds to improve efficiency and control costs through identifying costly healthcare services such as unnecessary extra diagnostic tests and treatments.

4.1.3 Decision support capability

Decision support capability emphasizes the ability to produce reports about daily healthcare services to aid managers' decisions and actions. In general, this capability yields sharable information and knowledge such as historical reporting, executive summaries, drill-down queries, statistical analyses, and time series comparisons. Such information can be utilized to provide a comprehensive view to support the implementation of evidence-based medicine, to detect advanced warnings for disease surveillance, and to develop personalized patient care. Some information is deployed in real time (e.g., medical devices' dashboard metrics) while other information (e.g., daily reports) will be presented in summary form.

The reports generated by the big data analytics systems are distinct from transitional IT systems, showing that it is often helpful to assess past and current operation environment across all organizational levels. The reports are created with a systemic and comprehensive perspective and the results evaluated in the proper context to enable managers to recognize feasible opportunities for improvement, particularly regarding long-term strategic decisions. From our case analysis, we found that Premier Healthcare Alliance collects data from different departmental systems and sends it to a central data warehouse. After near-real-time data processing, the reports

generated are then used to help users recognize emerging healthcare issues such as patient safety and appropriate medication use.

4.1.4 Predictive capability

Predictive capability is the ability to build and assess a model aimed at generating accurate predictions of new observations, where new can be interpreted temporally and or cross-sectionally (Shmueli & Koppius, 2011). Wessler (2013) defines predictive capability as the process of using a set of sophisticated statistical tools to develop models and estimations of what the environment will do in the future. By definition, predictive capability emphasizes the prediction of future trends and exploration of new insights through extraction of information from large data sets. To create predictive capability, organizations have to rely on a predictive analytics platform that incorporate data warehouses, predictive analytics algorithms (e.g., regression analysis, machine learning, and neural networks), and reporting dashboards that provide optimal decisions to users. This platform makes it possible to cross reference current and historical data to generate context-aware recommendations that enable managers to make predictions about future events and trends.

In healthcare, predictive analytics has been widely utilized to reduce the degree of uncertainty such as mitigating preventable readmissions, enabling managers to make better decisions faster and hence supporting preventive care (Bardhan, Oh, Zheng, & Kirksey, 2014; Simon, 2013). From our case analysis, we found that Texas Health Harris Methodist Hospital Alliance analyzes data from medical sensors to predict patients' movements and monitor patients' actions throughout their hospital stay. In doing so, Texas Health Harris Methodist Hospital Alliance is able to leverage reports, alerting, key performance indicators (KPIs), and interactive visualizations created by

predictive analytics to provide needed services more efficiently, optimize existing operations, and improve the prevention of medical risk.

Moreover, predictive analytics allows healthcare organizations to assess their current service situations to help them disentangle the complex structure of clinical costs, identify best clinical practices, and gain a broad understanding of future healthcare trends based on an in-depth knowledge of patients' lifestyles, habits, disease management and surveillance (Groves et al., 2013). For instance, I+Plus, an advanced analytical solution with three-level analysis (i.e., claims, aggregated, and admission) used in an Australian healthcare organization, provides claim-based intelligence to facilitate customers claim governance, balance cost and quality, and evaluate payment models (Srinivasan, & Arunasalam, 2013). Specifically, through these predictive analytical patterns managers can review a summary of cost and profit related to each healthcare service, identify any claim anomalies based on comparisons between current and historical indicators, and thus make proactive (not reactive) decisions by utilizing productive models.

4.1.5 Traceability

Traceability is the ability to track output data from all the system's IT components throughout the organization's service units. Healthcare-related data such as activity and cost data, clinical data, pharmaceutical R&D data, patient behavior and sentiment data are commonly collected in real time or near real time from payers, healthcare services, pharmaceutical companies, consumers and stakeholders outside healthcare (Groves et al., 2013). Traditional methods for harnessing these data are insufficient when faced with the volumes experienced in this context, which results in unnecessary redundancy in data transformation and movement, and a high rate of inconsistent data. Using big data analytics algorithms, on the other hand, enables authorized users to gain access to

large national or local data pools and capture patient records simultaneously from different healthcare systems or devices. This not only reduces conflicts between different healthcare sectors, but also decreases the difficulties in linking the data to healthcare workflow for process optimization.

The primary goal of traceability is to make data consistent, visible and easily accessible for analysis. Traceability in healthcare facilitates monitoring the relation between patients' needs and possible solutions through tracking all the datasets provided by the various healthcare services or devices. For example, the use of remote patient monitoring and sensing technologies has become more widespread for personalized care and home care in U.S. hospitals. Big data analytics, with its traceability, can track information that is created by the devices in real time, such as the use of Telehealth Response Watch in home care services. This makes it possible to gather location, event and physiological information, including time stamps, from each patient wearing the device. This information is immediately deposited into appropriate databases (e.g., NoSQL and the Hadoop distributed file system), for review by medical staff when needed with excellent suitability and scalability. Similarly, incorporating information from radio frequency identification devices (RFID) into big data analytics systems enables hospitals to take prompt action to improve medical supply utilization rates and reduce delays in patient flow. From our case analysis, we found that Brigham and Women's Hospital (BWH) provides a typical example of the use of in-depth traceability in large longitudinal healthcare databases to identify drug risk. By integrating big-data algorithms into the legacy IT systems, medical staff can automatically monitor drug safety by tracking warning signals triggered by alarm systems.

In the next subsection, we will describe the results of our second research objective, which are the benefits healthcare organizations could drive from big data analytics.

4.2 Potential Benefits of Big Data Analytics

Our results from content analysis reveal that the big data analytics derived benefits can be classified into five categories: IT infrastructure benefits, operational benefits, organizational benefits, managerial benefits, and strategic benefits, as summarized in Table 3. The two most compelling benefits of big data analytics are IT infrastructure (coded as part of 79 statements) and Operational benefits (73). The results also show that *reduce system redundancy* (19), *avoid unnecessary IT costs* (17), and *transfer data quickly among healthcare IT systems* (17) are the elements most mentioned in the category of IT infrastructure benefit; *improve the quality and accuracy of clinical decisions* (21), *process a large number of health records in seconds* (16), and *reduce the time of patient travel* (15) are the elements with high frequency in the category of operational benefits. This implies that big data analytics has a twofold potential as it implements in an organization. It not only improves IT effectiveness and efficiency, but also supports the optimization of clinical operations. In addition, our results also indicate that big data analytics is still at an early stage of development in healthcare due to the limited benefits of big data analytics at the organizational, managerial, and strategic levels.

Table 3. Breaking down the Potential Benefits Driven by Big Data Analytics in Health Care

Potential benefits of big data analytics	Elements	Frequency	
IT infrastructure benefits	Reduce system redundancy	19	79
	Avoid unnecessary IT costs	17	
	Transfer data quickly among healthcare IT systems	17	
	Better use of healthcare systems	13	
	Process standardization among various healthcare IT systems	9	
	Reduce IT maintenance costs regarding data storage	4	

Operational benefits	Improve the quality and accuracy of clinical decisions	21	73
	Process a large number of health records in seconds	16	
	Reduce the time of patient travel	15	
	Immediate access to clinical data to analyze	8	
	Shorten the time of diagnostic test	8	
	Reductions in surgery-related hospitalizations	3	
	Explore inconceivable new research avenues	2	
Organizational benefits	Detect interoperability problems much more quickly than traditional manual methods	8	13
	Improve cross-functional communication and collaboration among administrative staffs, researchers, clinicians and IT staffs	3	
	Enable to share data with other institutions and add new services, content sources and research partners	2	
Managerial benefits	Gain insights quickly about changing healthcare trends in the market	5	9
	Provide members of the board and heads of department with sound decision-support information on the daily clinical setting	2	
	Optimization of business growth-related decisions	2	
Strategic benefits	Provide a big picture view of treatment delivery for meeting future need	3	5
	Create high competitive healthcare services	2	
Total		179	

5. The Strategies for Success with Big Data Analytics

To create a data-driven organization, practitioners have to identify the strategic and business value of big data analytics, rather than merely concentrating on a technological understanding of its implementation (Wang, Kung, Wang, & Cegielski, 2014). However, evidence from a survey of 400 companies around the world shows that 77% of companies surveyed do not have clear strategies for using big data analytics effectively (Wegener & Sinha, 2013). These companies failed to describe how big data analytics will shape their business performance and transform their business models. Especially for healthcare industries, healthcare transformation through implementing big data analytics is still in the very early stages. Attention is sorely needed for research to formulate appropriate strategies that will enable healthcare organizations to move

forward to leverage big data analytics most efficiently and effectively. Thus, we recommend the following five strategies for being successful with big data analytics in healthcare settings.

5.1 Implementing (big) data governance

Data governance is an extension of IT governance that focuses on leveraging enterprise-wide data resources to create business value. Indeed, big data analytics is a double-edged sword for IT investment, potentially incurring huge financial burden for healthcare organizations with poor governance. On the other hand, with appropriate data governance, big data analytics has the potential to equip organizations to harness the mountains of heterogeneous data, information, and knowledge from a complex array of internal applications (e.g., inpatient and ambulatory EHRs) and healthcare networks' applications (e.g., laboratory and pharmacy information systems). Success in data governance requires a series of organizational changes in business processes since all the data has to be well understood, trusted, accessible, and secure in a data-driven setting. Thus, several issues should be taken into consideration when developing data governance for a healthcare organization.

The first step is to formulate the missions of data governance, with clearly focused goals, execution procedures, governance metrics, and performance measures. In other words, a strong data governance protocol should be defined to provide clear guidelines for data availability, criticality, authenticity, sharing, and retention that enable healthcare organizations to harness data effectively from the time it is acquired, stored, analyzed, and finally used. This allows healthcare organizations to ensure the appropriate use of big data and build sustainable competitive advantages. Second, healthcare organizations should review the data they gather within all their units and realize their value. Once the value of these data has been defined, managers can make

decisions on which datasets to be incorporated in their big data analytics framework, thereby minimizing cost and complexity. Finally, information integration is the key to success in big data analytics implementation, because the challenges involved in integrating information across systems and data sources within the enterprise remain problematic in many instances. In particular, most healthcare organizations encounter difficulties in integrating data from legacy systems into big data analytics frameworks. Managers need to develop robust data governance before introducing big data analytics in their organization.

To create a strong data governance environment, The University of Kansas Hospital has established a data governance committee for managing the availability, usability, integrity, and security of the organization's data. This committee has three different groups with specific responsibilities. The data governance executive group is responsible of overseeing vision and strategy for improvement data quality, while the data advisory group establishes procedures and execution plans to address data quality issues, work priorities and the creation of working groups. The data governance support group is composed of technology, process improvement and clinical experts that provide support to the former two groups. With respect to the best practices of data governance, this committee provides users a secure commitment from senior leaders, implements data sharing processes and technologies that users could rely on for quality data pulled from disparate sources and systems, and identifies a data gap and a disruption in reporting key organizational metrics. With the strong data governance in big data analytics platforms, The University of Kansas Hospital has achieved more than 70 standardized enterprise data definition approvals in the first year and created a multi-year business intelligence/data governance roadmap.

5.2 Developing an information sharing culture

A prerequisite for implementing big data analytics successfully is that the target healthcare organizations foster information sharing culture. This is critical for reducing any resistance to new information management systems from physicians and nurses. Without an information sharing culture, data collection and delivery will be limited, with consequent adverse impacts on the effectiveness of the big data analytical and predictive capabilities. To address this issue, healthcare organizations should engage data providers from the earliest stage of the big data transition process and develop policies that encourage and reward them for collecting data and meeting standards for data delivery. This will significantly improve the quality of data and the accuracy of analysis and prediction.

5.3 Training key personnel to use big data analytics

The key to utilize the outputs from big data analytics effectively is to equip managers and employees with relevant professional competencies, such as critical thinking and the skills of making an appropriate interpretation of the results. Because incorrect interpretation of the reports generated could lead to serious errors of judgment and questionable decisions. Therefore, it is important that healthcare organizations provide analytical training courses in areas such as basic statistics, data mining and business intelligence to those employees who will play a critical support role in the new information-rich work environment. According to a recent survey by the American Management Association (2013), mentoring, cross-functional team-based training and self-study are beneficial training approaches to help employees develop the big data analytical skills they will need. Alternatively, healthcare organizations can adjust their job selection criteria to recruit prospective employees who already have the necessary analytical skills.

5.4 Incorporating cloud computing into the organization's big data analytics

Most hospitals are small and medium sized enterprises (SMEs), and often struggle with cost and data storage issues. Due to the rapid changes of technology, big data, and the general increase in data-intensive operations, healthcare organizations are facing some challenges: storage, analysis, and bottom line. The needs to store different formats of data and access to them for decision making have pushed healthcare organizations seeking better solutions other than traditional storage servers and processes. A typical model for the storage of big data is clustered network-attached storage (NAS), which is a costly distributed file system for SMEs. A usage-based charging model such as cloud computing services is an attractive alternative. Cloud computing is a network-based infrastructure capable of storing large scale of data in virtualized spaces and performing complex computing near real time. The combination of lower cost and powerful and timely processing and analyzing make cloud computing an ideal option for healthcare SMEs to fully take advantage of big data analytics.

However, storing healthcare data in a public cloud raises two major concerns: security and patient privacy (Sahoo et al., 2014). Although the public cloud is a significant cost savings option, it also presents higher security risk and may lead to the loss of control of patient privacy since the access to data is managed by a third party vendor. A private cloud, on the other hand, provides a more secure environment and keeps the critical data in-house, but increases the budget for big data analytics projects. Healthcare managers must strike a balance between the cost-effectiveness of the different cloud choices and patient information protection when adopting big data analytics.

5.5 Generating new business ideas from big data analytics

New idea generation is not only necessary for organizational innovation, but also can lead to changes in business operations that will increase productivity and build competitive advantages. This could be achieved through the use of powerful big data predictive analytics tools. These tools can provide detailed reporting and identify market trends that allow companies to accelerate new business ideas and generate creative thinking. In addition to using big data analytics to answer known questions, managers should encourage users to leverage outputs such as reports, alerting, KPIs, and interactive visualizations, to discover new ideas and market opportunities, and assess the feasibility of ideas (Kwon, Kwak, Kim, 2015).

6. Limitation, Future Research and Conclusion

Through analyzing big data cases, our research has provided a better understanding how healthcare organizations can leverage big data analytics as a means of transforming IT to gain business value. However, like any other study, ours has limitations. The primary limitation of this study is the data source. One challenge in the health care industry is that its IT adoption usually lags behind other industries, which is one of the main reasons that cases are hard to find. Although efforts were made to find cases from different sources, the majority of the cases identified for this study came from vendors. There is therefore a potential bias, as vendors usually only publicize their “success” stories. Further and better discovery could be done through collecting and analyzing primary data. Given the growing number of healthcare organizations adopting big data analytics, the sample frame for collecting primary data becomes larger. Examining the impact of big data analytics capabilities on healthcare organization performance with quantitative analysis method based on primary data could shed different lights.

In addition to requiring empirical analysis of big data analytics enabled transformation, our study also expose the needs for more scientific and quantitative studies, focusing on some of the business analytics capability elements we identified. This especially applies to analytical and decision support capabilities, which are cited frequently in the big data cases. With a growing amount of diverse and unstructured data, there is an urgent need for advanced analytic techniques, such as deep machine learning algorithm that allows computers to detect items of interest in large quantities of unstructured data, and to deduce relationships without needing specific models or programming instructions. We thus expect future scientific studies to take developing efficient unstructured data analytical algorithms and applications as primary technological developments.

Finally, the foundation to generate any IT business value is the link among the three core dimensions: process, IT, and people (Melville, Kraemer, & Gurbaxani, 2004). However, this study merely focuses on the IT angle, ignoring the people side of this capability as the cases barely highlight the importance of analytical personnel. Indeed, analytical personnel who have an analytic mindset play a critical role in helping drive business value from big data analytics (Davenport, Harris, & Morison, 2010). We thus expect that future research should take analytical personnel into consideration in the big data analytics framework.

In conclusion, the cases demonstrate that big data analytics could be an effective IT artifact to potentially create IT capabilities and business benefits. Through analyzing these cases, we sought to understand better how healthcare organizations can leverage big data analytics as a means to create business value for health care. We also identified five strategies that healthcare organizations could use to implement their big data analytics initiatives.

References

- Agarwal, R., Gao, G., DesRoches, C., & Jha, A. K. (2010). Research commentary-The digital transformation of healthcare: current status and the road ahead. *Information Systems Research*, 21(4), 796-809.
- American Management Association (AMA). (2013). *Conquering Big Data: Building Analytical Skills in Your Organization*, American Management Association Press.
- Bardhan, I., Oh, J. H., Zheng, Z., & Kirksey, K. (2014). Predictive Analytics for Readmission of Patients with Congestive Heart Failure. *Information Systems Research*, 26(1), 19-39.
- Bodenheimer, T. (2005). High and rising health care costs. Part 1: seeking an explanation. *Annals of Internal Medicine*, 142(10), 847-854.

- Bryant, R.E., Katz, R.H., and Lazowska, E.D. Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society Computing, In Computing Research Initiatives for the 21st Century, *Computing Research Association*, 2008. Available at http://www.cra.org/ccc/files/docs/init/Big_Data.pdf.
- Burnard, P. (1991). A method of analysing interview transcripts in qualitative research. *Nurse Education Today*, 11(6), 461-466.
- Chen, H., Chiang, R.H.L. & Storey, V.C. (2012). Business intelligence and analytics: from big data to dig impact, *MIS Quarterly*, 36(4), 1165-1188.
- Cortada, J.W. Gordon, D. and Lenihan, B. (2012), *The Value of Analytics in Healthcare: From Insights to Outcomes*. Somers, NY: IBM Global Business Services.
- Cosic, R., Shanks, G., Maynard, S. (2012). Towards a business analytics capability maturity model. In *Proceeding of 23rd Australasian Conference on Information Systems*, Geelon, Australia.
- Cox, M., and Ellsworth, D. Application-controlled demand paging for out-of-core visualization, In Proceedings of the 8th IEEE Conference on Visualization, *IEEE Computer Society Press*, 1997, Los Alamitos, CA.
- Davenport, T. H., Harris, J. S., & Morison, R. (2010). *Analytics at Work*, Cambridge, MA: Harvard Business School Press.
- Dey, I. (1993). *Qualitative Data Analysis. A User-Friendly Guide for Social Scientists*. Routledge, London.
- Downe-Wamboldt, B. (1992). Content analysis: method, applications, and issues. *Health Care for Women International*, 13(3), 313-321.
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107-115.

- Esteves, J. (2009). A benefits realisation road-map framework for ERP usage in small and medium-sized enterprises. *Journal of Enterprise Information Management*, 22(1/2), 25-35.
- Gartner. (2013). Gartner Says Cloud Computing Will Become the Bulk of New IT Spend by 2016. Retrieved from <http://www.gartner.com/newsroom/id/2613015>.
- Gefen, D., & Ragowsky, A. (2005). A multi-level approach to measuring the benefits of an ERP system in manufacturing firms. *Information Systems Management*, 22(1), 18-25.
- Goh, J. M., Gao, G., & Agarwal, R. (2011). Evolving work routines: Adaptive routinization of information technology in healthcare. *Information Systems Research*, 22(3), 565-585.
- Grantmakers In Health. (2012). *Transforming Health Care Delivery: Why It Matters and What Will it Take*, Washington, DC: Grantmakers In Health.
- Groves, P., Kayyali, B., Knott, D., & Kuiken, S.V. (2013). *The "Big Data" Revolution in Healthcare: Accelerating Value and Innovation*, McKinsey & Company.
- Herrick, D.M., Gorman, L., & Goodman, J.C. (2010). *Health Information Technology: Benefits and Problems*, Dallas, Texas: National Center for Policy Analysis.
- Hurwitz, J., Nugent, A., Hapler, F., & Kaufman, M., (2013). *Big Data for Dummies*, Hoboken, New Jersey: John Wiley & Sons.
- IBM. (2012). *Understanding Holistic Database Security: 8 Steps to Successfully Securing Enterprise Data Sources*, IBM Corporation.
- Institute for Health Technology Transformation. (2013). *Transforming Health Care through Big Data: Strategies for leveraging big data in the health care industry*, New York: Institute for Health Technology Transformation.

- Jagadish, H.V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J.M., Ramakrishnan, R., & Shahabi, C. (2014). Big Data and its Technical Challenges. *Communications of the ACM* 57(7), 86-94.
- Jiang, P., Winkley, J., Zhao, C., Munnoch, R., Min, G., & Yang, L. T. (2014). An intelligent information forwarder for healthcare big data systems with distributed wearable sensors. *Systems Journal, IEEE*, PP(99), 1-9.
- Ker, J. I., Wang, Y., Hajli, M. N., Song, J., & Ker, C. W. (2014). Deploying lean in healthcare: Evaluating information technology effectiveness in US hospital pharmacies. *International Journal of Information Management*, 34(4), 556-560.
- Kwon, T. H., Kwak, J. H., & Kim, K. A study on the establishment of policies for the activation of a big data industry and prioritization of policies: Lessons from Korea. *Technological Forecasting and Social Change*, 96, 2015, pp. 144-152
- LaLalle, S., Lesser, E., Shockley, R., Hopkins, M.S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 52(2), 21-31.
- Melville, N., Kraemer, K., & Gurbaxani, V. (2004). Review: Information technology and organizational performance: An integrative model of IT business value. *MIS Quarterly*, 28(2), 283-322.
- Mueller, B., Viering, G., Legner, C., & Riempp, G. (2010). Understanding the economic potential of service oriented architecture. *Journal of Management Information Systems* 26(4), 145-180.
- Murdoch, T.B. & Detsky, A.S. (2013). The inevitable application of big data to health care. *The Journal of the American Medical Association*, 309(13), 1351-1352.

- Phillips-Wren, G., Iyer, L. S., Kulkarni, U., & Ariyachandra, T. (2015). Business Analytics in the Context of Big Data: A Roadmap for Research. *Communications of the Association for Information Systems*, 37(1), 448-472.
- Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1), 3.
- Russom, P. Big data analytics, Renton, WA: TDWI research, 2011.
- Sahoo, S. S., Jayapandian, C., Garg, G., Kaffashi, F., Chung, S., Bozorgi, A., ... & Zhang, G. Q. (2014). Heart beats in the cloud: distributed analysis of electrophysiological 'Big Data' using cloud computing for epilepsy clinical research. *Journal of the American Medical Informatics Association*, 21(2), 263-271.
- Shah, N.D., & Pathak, J. (2014). Why Health Care May Finally Be Ready for Big Data. Retrieved January 12 2015 from Harvard Business Review: <https://hbr.org/2014/12/why-health-care-may-finally-be-ready-for-big-data>.
- Shang, S., & Seddon, P. B. (2002). Assessing and managing the benefits of enterprise systems: the business manager's perspective. *Information Systems Journal*, 12(4), 271-299.
- Sharma, R. Mithas, S. and Kankanhalli, A. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, 23(4), 433-441.
- Shmueli, G., & Koppius, O. R. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 35(3), 553-572
- Simon, P., *Too Big to Ignore: The Business Case for Big Data*, Hoboken, New Jersey: John Wiley & Sons, 2013.

- Srinivasan, U., & Arunasalam, B. (2013). Leveraging big data analytics to reduce healthcare costs, *IT Professional*, 15(6), 21-28.
- Storage Networking Industry Association (2009). *The information lifecycle management maturity model*. Storage Networking Industry Association Press.
- The Kaiser Family Foundation. (2012). *Health Care Costs: A Primer, Key Information on Health Care Costs and Their Impact*. The Henry J. Kaiser Family Foundation.
- Trkman, P., McCormack, K., De Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, 49(3), 318-327.
- Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234-246.
- Wang, Y., Kung, L., Ting, C., & Byrd, T. A. (2015). Beyond a technical perspective: understanding big data capabilities in health care. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on* (pp. 3044-3053). IEEE.
- Wang, Y., Kung, L., Wang, Y.C., & Cegielski, C. (2014). Developing IT-enabled transformation model: the case of big data in healthcare. In *Proceedings of 35th International Conference on Information Systems (ICIS)*, Auckland, New Zealand.
- Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, 34(1), 1247-1268.
- Wegener, R., & Sinha, V. (2013). *The value of Big Data: How analytics differentiates winners*. Bain and Company.
- Wessler, M. (2013). *Big Data Analytics for Dummies*, Hoboken, NJ: John Wiley & Sons, Inc.

Wixom, B., Yen, B., & Relich, M. (2013). Maximizing value from business analytics. *MIS Quarterly Executive*, 12(2), 111-123.

Appendix A. Case List

Case	Case name	Country	Sources	
1	Wissenschaftliches Institut der AOK (WIdO)	Germany	Released by vendors or companies	IBM
2	Brigham and Women's Hospital	United States		IBM
3	The Norwegian Knowledge Centre for the Health Services (NOKC)	Norway		IBM
4	Memorial Healthcare System	United States		IBM
5	University of Ontario Institute of Technology	Canada		IBM
6	Premier healthcare alliance	United States		IBM
7	Bangkok Hospital	Thailand		IBM
8	Rizzoli Orthopedic Institute	Italy		IBM
9	Universitätsklinikum Erlangen	Germany		IBM

10	Fondazione IRCCS Istituto Nazionale dei Tumori (INT)	Italy		IBM
11	Fraunhofer FOKUS	Germany		IBM
12	Leeds Teaching Hospitals	UK		Intel/ Microsoft
13	Beth Israel Deaconess Medical Center	United States		Microsoft
14	Atlantic Health System	United States		EMC ² /Intel
15	Private health institution in Australia	Australia		IT Professional
16	University Hospitals Case Medical Center	United States	Practical journals	Journal of the American Medical Informatics Association
17	Texas Health Harris Methodist Hospital	United States		Medcitynews /Modern Healthcare.com
18	Mount Sinai Medical Center	United States		MIT Technology Review/ Science Translational Medicine
19	Indiana University Health	United States	Print publications	Health Catalyst
20	Mission Health	United States		
21	MultiCare Health System	United States		
22	North Memorial Health Care	United States		
23	OSF HealthCare	United States		
24	Partners HealthCare	United States		
25	The University of Kansas Hospital	United States		
26	Texas Children's Hospital	United States		