Spring 3-22-2017

# Big data: Harnessing the beast!!

Rebecca Hovemeyer
*Marshall University*

Debra Stinson
*Marshall University*

Beteseb Gebremariam
*Marshall University*

Alberto Coustasse Dr. PH, MD, MBA
*Marshall University*, coustassehen@marshall.edu



Recommended Citation

Hovemeyer, Rebecca, M.S., Stinson, Debra, M.S., Gebremariam, Beteseb, M.S. & Alberto Coustasse, Dr. PH. (2017), "Big data: Harnessing the beast!!," in Business & Health Administration Proceedings, Avinandan Mukherjee, Editor, pp.117-125.

# BIG DATA: HARNESSING THE BEAST!!

Rebecca Hovemeyer, MS
Health Care Administration Program
Marshall University Graduate College
100 Angus E. Peyton Drive
South Charleston, WV 25303

Debra Stinson, MS
Health Care Administration Program
Marshall University Graduate College
100 Angus E. Peyton Drive
South Charleston, WV 25303

Beteseb Gebremariam, MS
Health Care Administration Program
Marshall University Graduate College
100 Angus E. Peyton Drive
South Charleston, WV 25303

Alberto Coustasse, Dr. PH, MD, MBA MPH – Contact Author
Professor
Health Care Administration Program
Marshall University Graduate College
100 Angus E. Peyton Drive
South Charleston, WV 25303
(304) 746-1968
(304) 746-2063 FAX
coustassehen@marshall.edu

# BIG DATA: HARNESSING THE BEAST!!

***Rebecca Hovemeyer, Marshall University***
***Debra Stinson, Marshall University***
***Beteseb Gebremariam, Marshall University***
***Alberto Coustasse, Marshall University***

***ABSTRACT***

*The healthcare industry's growing use of health information technology has contributed to the enormous accumulation of health care data, leading to active use of the term big data. Although there has been large amounts and varieties of complex data captured during patient care, this data has remained vastly underutilized. The purpose of this study was to assess the variety of benefits and barriers of obtaining meaningful information from big data in healthcare.*

*The methodology utilized was a qualitative literature review that referenced 17 sources published between 2005 and 2016.Findings suggest that applied big data analytics within the healthcare arena can enable the identification of specific patient groups and pre-disease stage patients, help identify the most effective treatment methods, and assist in developing personalized treatment. Big Data can also help identify potential health hazards, disease patterns, and contribute to disease epidemiology tracking leading to the mitigation of public health hazards. Obtaining meaningful information from big data in healthcare can lead to improved healthcare clinical practices, a reduction of overall healthcare cost, and applied epidemiology applications. However, there are several barriers to big data use in healthcare including big data complexity, security and privacy concerns, and poor data quality. Health care providers need to invest in the ability to integrate enormous amounts of data in order to derive meaningful information and fully realize the potential benefits of big data.*

***Key words:*** *Big Data, Healthcare Data, Health Information Management, Data Analysis*

## INTRODUCTION

The health care system of the United States (U.S.) has a number of challenges; the most commonly stated are the ever-increasing cost of care, low quality performance, and inaccessibility (McCarthy & Hart, 2011). Various attempts have been made to moderate these challenges, including the introduction and use of Information Technology (IT) in health care (Naylor, Kudlow, Li, & Yuen, 2011). The Health Information Technology for Economic and Clinical Health (HITECH) Act initiative has provided incentives to health care providers who employed meaningful use of Electronic Health Records (EHR) since 2009; EHR applications have been a significant step, and have resulted in the creation of patient-care related data in an electronic format (CDC, 2012). The collective move of the healthcare industry to the use of health information technology has contributed to a huge accumulation of health care data and has brought to use a terminology that has become familiar to the industry in the last few years: big data (Ward & Adam, 2013).

### Significance of the Problem

Big data definitions have changed over time dependent on the industry in which it has been applied; in health care, big data has been defined as a gathering of data elements whose size, speed, type, and/or complexity require individuals to seek, develop, and implement new hardware and software mechanisms to successful store, analyze, and visualize the data (Belle, et al., 2015). Demchenko, Zhao, Grosso, Wibisono, & Laat (2012) described the primary characteristics of health care big data as five V's: Volume, Velocity, Variety, Veracity, and Value, where volume refers to the huge amount of health-related data being created and gathered constantly. In 2011, the U.S. healthcare system reached 150 exabytes, and was soon predicted to reach the zettabyte (1021 gigabytes) scale and, not long after, the yottabyte (1024 gigabytes) (Raghupathi & Raghupathi, 2014). Velocity refers to the continuous flow of new data amassing at an astonishing rate, variety refers to the data complexity level, veracity includes questions of trust and uncertainty regarding the data itself and the resulting analysis of that data, and value indicates the quality of the data compared to the intended or desired results (Herland, Khoshgoftaar, & Wald, 2014).

The complexity of health care big data has been attributed to the fact that data created during patient care is highly varied and includes sources of data in numerous multimedia formats and designs including structured, unstructured and semi-structured design (Ward & Adam, 2013). Big data also comes in two distinct forms: data at rest - datasets that are analyzed after they are collected, and big data in motion - datasets processed and analyzed immediately as they are developing (Yan, 2013). The various sources of health care data include clinical data, patient data in Electronic Patient Records (EPRs), machine generated/sensor data, social media posts, emergency care data, news feeds, and articles in medical journals (Raghupathi & Raghupathi, 2014).

Although the application of IT has enabled the capture of huge amounts and varieties of patient information, this data has remained vastly underutilized, wasting the opportunity to gain insights from big data analysis which might assist to save lives, improve care delivery, expand healthcare access, align payment with performance, and help curb growing healthcare costs (Belle, et al., 2015). Additionally, despite 2014 health care spending in the US accounting for over 17% of the GDP, substantially high compared to other countries with modern health care systems who spend 12% or less, the US healthcare system clinical outcomes are only comparable or worse (McClellan & Rivlin, 2014).

The purpose of this study was to assess and enumerate the variety of benefits and barriers to obtaining meaningful information from big data in relation to healthcare.

## METHODOLOGY

The primary hypothesis of this study is that meaningful information obtained from big data in healthcare will increase positive health outcomes in the U.S. Secondly, this study will research how obtaining meaningful information will decrease the cost of healthcare in the U.S.

The methodology for this study was a literature review. The approach for this research study followed a series of systematic search steps and the research framework utilized by Yao, Chao-Hsien, and Li (2010). Figure 1 depicts the major steps of Big Data utilization and obtaining meaningful information for use by the healthcare delivery system and the barriers to impeding its implementation, as well as the benefits of obtaining meaningful information. The application of this conceptual framework with this study was considered suitable; it identifies major benefits of utilizing big data analytics in health care, and then highlights barriers to obtaining that meaningful information. (Figure 1).

Utilizing Big Data Analytics in the US Health Care Delivery System

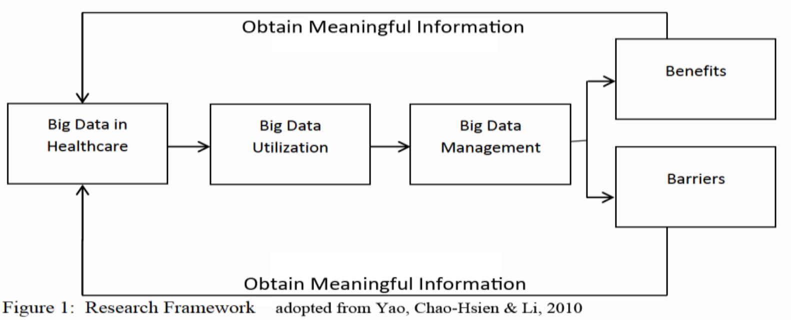


Figure 1: Research Framework adopted from Yao, Chao-Hsien & Li, 2010

**Search Strategy**

The key phrases used were: Big Data or Big Data Management, or Health Information Management, or Data Warehouse, and Healthcare Data or Data Analytics, or Data Analysis, or healthcare, or benefits, or barriers, or outcomes as inclusion criteria to explore scholarly databases for articles and references. The following databases were used for this research: PubMed, EBSCO host, Academic Search Premier, Google Scholar, and Google. Information from various websites was included, such as, Center for Disease Control and Prevention (CDC), Mckinsey & Company, Harvard Business Review, and the Institute for Health Technology Transformation ($iHT^2$).

Literature was selected by DS, BG, and RH for review on the basis of relevance to the study, and barriers and benefits to using Big Data Management to obtain meaningful information. References utilized were limited to those written in English and published between 2005 and 2015 in order to keep current. This led to the inclusion of 17 articles. References were analyzed and selected if they satisfied the inclusion criteria by containing accurate knowledge regarding Big Data analytics in healthcare, with special consideration on the benefits and barriers to obtaining meaningful information; this literature was further validated by AC who acted as a second reader verifying that all references met the inclusion criteria.

## RESULTS

**Benefits to Obtaining Meaningful Information from Big Data**

A review of the literature found a myriad of benefits that can be attained from meaningful information derived from Big Data (Table 1). Benefits ranged from direct applications in healthcare practice through to public health and research applications, and have been loosely organized into the following categories: improved healthcare practices, reduction of overall healthcare cost, and epidemiology applications (Table 1).

Table 1: Benefits to Obtaining Meaningful Information from Big Data in Healthcare

| **Benefits** | **Details (Citation)** |
|---|---|
| Improved Health Care Practices | • Identifying broad cohorts and segmenting those cohorts into targeted risk populations for targeted treatment (Bates, et. al., 2014)<br>• Identify the pre-disease state before the occurrence or serious deterioration of diseases for applied intervention (Li & Chen, 2014)<br>• Integrate clinical, medical and public health electronic systems to potentially enable providers, payers, and other stakeholders to better coordinate care for the entire community. (Nash, 2014)<br>• Make better use of prior trial data, finding opportunities to improve clinical trials and traditional treatment protocols. (Groves et al., 2013)<br>• Use big data generated genomics with prior treatment to offer personalized care deliver to improve patient outcomes. (Jee & Kim, 2013) |
| Reduce Overall Healthcare Cost | • Identification of fraud/ questionable patterns in clinical laboratory claims (Goozner, 2014)<br>• Reduce costs by basing treatment on sound clinical evidence, reduce waste of time and money on ineffective practices (Sarasohn-Khan, 2014)<br>• Using big data predictive analytics to predict treatment costs, enabling potential realized treatment cost savings. (Maguire & Dhar, 2013) |
| Epidemiology Uses | • Use geocoding to detect epidemics or identify 'hot spots' (of diseases, high costs, etc.) thereby enabling timely intervention. (Bates, et. al., 2014)<br>• Investigate occurrences of health issues, analyzing origins and contributing factors to health hazards and identify health trends (Ola & Sedig, 2014). |

Within the direct treatment of patients vast benefits can be gained; using knowledge gained from big data, physicians and hospitals can better identify gaps in the care of specific patient groups, enabling them to anticipate readmissions within segment populations so that these populations can be handled in such a way as to reduce preventable readmissions. (Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014) This patient group targeting can not only assist to pinpointing what works best for specific patient groups, but can lead to reasons why these particular patients would be more likely to be readmitted, enabling the potential for more personalized treatment specific these patients' needs (Goozner, 2014). The sharing of research data and results through big data warehousing of research information can offer physician researchers the opportunity to make better use of prior trial data, enabling them to expand on these clinical trials to potentially improve traditional treatment conventions (Groves, Kayyali, Knott, VanKuiken, 2013). Physicians can also integrate knowledge gained from big data generated genomics with prior patient treatment information to potentially offer personalized care delivery to improve outcomes for patients with cancer and other diseases (Jee & Kim, 2013)(Table 1).

Preventative medical approaches often have more potential to reduce costs to the global healthcare community than treatment does; knowledge derived from big data can lead to the use of genomic and sensor-input data markers identifying the pre-disease state of a patient prior to severe disease deterioration so patient healthcare needs can be anticipated before hospitalization (Li & Chen, 2014)(Table 1). This potential for use in preventative care can mean huge benefits for both patient and provider from a cost standpoint; it has been stated that if the US healthcare system uses big data effectively, the healthcare sector could create more than $300 billion in value every year with two-thirds of this value in the form of reducing US healthcare expenditure (Groves, et al., 2013). Healthcare cost savings can also be realized by selecting the most effective and efficient treatment available; a study by Dhar and Maguire (2013) used big data predictive analytics to predict diabetic treatment costs associated with the various different treatment patterns, which can enable potential cost savings through provider treatment selection. (Maguire & Dhar, 2013).

Public Health also stands to benefit by the use of healthcare Big Data analytics in combining epidemiologic and geocoding information; big data knowledge from these disparate systems can be used to discover disease patterns, eliciting epidemic intelligence for early identification of potential health hazards assisting public health officials and healthcare providers to anticipate and prepare for necessary intervention (Bates, et al., 2014). Further addition of geographic and environmental information into big data can assist in pinpointing the origins and factors contributing to epidemics and diseases, enabling early identification and reduction of public health hazards, subsequently reducing the burden on healthcare facilities and providers. (Ola & Sedig, 2014)

**Barriers to Obtaining Meaningful Information from Big Data**

Despite the benefits of utilizing big data analytics in healthcare, there are several barriers. One major barrier has been the complexity and high volume of data involved; the healthcare system's ability to produce health care data far outpaces its analytical resources (Martin-Sanchez & Versppoor, 2014). Healthcare big data is inherently complex due to integrating data from diverse hospital systems each with it's own unique ways of structuring data, different levels of discrete versus non-discrete data, and a myriad of ways to code the data (Li & Chen, 2014). This, along with other factors such as provider's fear to release patient data, has created information silos in both hospitals and their respective departments (Jee & Kim, 2013). (Table 2)

Another factor adding to the complexity has been the security and privacy of healthcare data; patient data is sensitive, requiring consent if being used for research (Cohen, Amarasingham, Shah, Bin, & Lo, 2014). In cases where the data can be de-identified, there is concern that when merging so many resources, linking techniques could be used to re-identify the data (Sarasohn-Khan, 2014). Once access to patient data is obtained, there is a risk of non-equitable representation of populations leading to unfair targeting of groups of people for illness – a sensitive issue since Tuskegee (Sarasohn-Khan, 2014). This data security and privacy concern has also been a limiting factor in sharing healthcare data; many other industries use cloud computing to share the computational and storage burden of the massive data utilized in big data analytics, however, this has not been possible with health data due to unclear residency and ownership of patient data (Li & Chen, 2014). (Table 2)

Even once data has been obtained, poor data quality is a huge factor; data analysts must develop algorithms to validate and clean data, as analysis output is only as good as the input (Ola & Sedig, 2014). There are numerous sources of inconsistently stored non-discrete data; this inconsistency makes developing algorithms to handle this data tedious and extremely timely consuming if not next to impossible (Chute, et al., 2014). Even after data scrubbing, there may still be erroneous and incorrectly entered data that cannot be caught by algorithms, as it will appear correct to a computer algorithm (Modern Healthcare, 2014).

In order to correctly develop algorithms and understand results, the analyst must have a good background in clinical and data analytics; unfortunately, there is a lack of competent data scientists in healthcare (Groves, et al., 2013). This dearth of proficient data scientists in healthcare makes the cost for acquiring them extremely high (Jee & Kim, 2013). In one 2015 report, the median base salary for individual data scientists was $120,000, compared to $95,000 for other predictive analytics professionals; similarly, the median base salary for managerial data scientists was $170,000, while other managerial predictive analytics professionals had a median base salary of $145,000 (Burtch, 2015). Not only has there been a lack of expertise, but also a lack of tools supporting large complex big data sets, and there is need to develop new statistical or computational methods (Ola & Sedig, 2014). Lack of data analytics tools capable of supporting big data has spurred entrepreneurship in big data analytics healthcare tools, which is leading to a lack of standards between multiple vendors and lack of consistency in deriving meaningful consistent knowledge (Baro, Degoul, Beuscart, & Chazard, 2015). (Table 2)

Similar to what happened in introducing EHR systems, there has been provider apprehension due to fear of using a system that provides more accurate and better information than any one physician could derive; meaning it could change the physician's role within healthcare (Cohen, et al., 2014). Healthcare providers also have a reluctance to share data for fear of losing a competitive advantage (Goozner, et al., 2014). Additionally, due to poor quality data, results are never 100% reliable and no direct causality can be achieved (only association), there is a lack of trust, unclear return on investment, and unclear results from big data investment (Baro, et al., 2015). (See Table 2)

Table 2: Barriers to Obtaining Meaningful Information through Big Data Analytics

| **Barriers** | **Details (Citation)** |
|---|---|
| Interoperability, Complex Data, & High Volume Data | • Integrating data from heterogeneous resources (multiple EHR systems, different database integration and structures) (Li & Chen, 2014)<br>• Ability to Produce Health Care data outpaces analytical resources; Large Batches of Data (Martin-Sanchez & Verspoor, 2014)<br>• Information Silos by entity, department, or guarded by patient security regulations (Jee & Kim, 2013) |
| Lack of Expertise in Analytics in Healthcare | • Lack of Analytical Expertise in Healthcare (Groves, et al, 2013)<br>• Existing Tools fall short in supportive cognitive activities --large complex data sets; need new statistical/computational methods (Ola & Sedig, 2014).<br>• Cost for competent data scientists is too high (Jee & Kim, 2013). |
| Poor Data Quality, Lack of ability to validate/clean data with algorithms | • Lack of good data; any system is only as good as what is put in (Modern Health, 2014)<br>• Different Database Structures – data stored inconsistently; Numerous sources of non-discrete data. (Chute, et al, 2014). |
| Security and Privacy | • Patient Data is Sensitive; Consent through informing or signed consent regulations. (Cohen, et al., 2014).<br>• Risk of Non-Equitable Representation (Sarasohn-Khan, 2014)<br>• De-Identifying Data can be undone with linkage techniques in big data sets from multiple sources. (Sarasohn-Khan, 2014)<br>• Can't Use Cloud Computing - a common Big Data tool. (Li & Chen, 2014) |
| Accuracy, Lack of Standards | • Big data can never be 100% accurate; Deriving meaningful consistent knowledge -- One data scientists may not find the same results as another and it would be actionable. (Baro, et al, 2015)<br>• Spur in Entrepreneurship in Big Data Analytics Tools; no standards between result interpretation (Goozner, 2014) |
| Provider/Healthcare Professional Support | • Reluctance to Share Data; Fear of Competitive Advantage (Goozner, 2014)<br>• Provider Reluctance to Utilize; Fear of using a system that provides more accurate and better information than any physician could -- Physicians used to relying on their individual education. (Cohen, et al, 2014). |

| | |
|---|---|
| | • Lack of trust in Results: No direct causality can be achieved, only association; Unclear return on investment (Baro, et al, 2015). |

## DISCUSSION

The intention of this paper has been to compile and evaluate the benefits and barriers to obtaining meaningful information from big data in relation to healthcare.

The concepts and issues found in the studies included both the benefits and barriers to obtaining meaningful information from big data; these concepts were fairly consistent across studies. However, the overall definition of big data occasionally varied, as some research approached big data as being merely the large amount of gathered patient health care data, and some researchers viewed big data as requiring the inclusion of a more varied array of sources including but not limited to social media, news feeds, existing research articles, epidemiology and geocoding information. The authors of the articles provided some indirect insight as to the reasons for this difference through the intention of their study; authors with a more healthcare-data related viewpoint were studying topics directly related to the provision of healthcare or the cost thereof, and authors with a larger viewpoint were studying topics which required use of additional datasets outside of the direct patient-related dataset.

The findings suggest that big data analytics within the healthcare arena can lead to improved healthcare practices, a reduction of overall healthcare cost, and epidemiology applications which can be of benefit to healthcare. In improved healthcare practices, physicians and hospitals can target specific patient groups, reducing preventable readmissions and personalizing treatment. With knowledge gained from integrating various big datasets, physicians can offer more personalized care delivery for cancer and other chronic diseases. Physicians can also use information from big data research databases to improve traditional treatment conventions, enabling a reduction of costs. The findings suggest big data analytics knowledge can be used in preventative care to identify pre-disease stage patients as well as determine the most effective and efficient treatment methods, reducing the number of patients and unnecessary treatments which leads to a reduction in healthcare cost. Lastly, the findings suggest big data analysis can be used in applied epidemiology to discover disease patterns, quickly identify potential health hazards, and pin-point origins and factors contributing to disease epidemics, leading to the prevention or reduction of public health hazards, facilitating both less burden on the healthcare arena and better healthcare when epidemics do happen.

The findings suggest that there are multiple barriers including the complexity of big data, security and privacy of health care data, quality of the data, lack of data scientists, and fear of participation from health care providers. In terms of big data complexity, the results showed multiple factors such as high volume, unique data structuring, discrete versus non-discrete data, and silos of data. The sensitivity of patient data requires in depth attention to security policies and procedures. The findings also suggest that in order to be usable, the data must be extensively cleaned and validated. In order to validate, clean, analyze, and obtain meaningful information from big data, the health care system needs competent data scientists and tools to support large complex data sets; however, results showed both proficient big data healthcare scientists and tools to be in short supply. Additionally, the findings suggest an apprehension to big data analytics use in health care, which can be attributed to physician's fear of replacement, fear of sharing data, unclear return on investment, or the unclear nature of results.

### Study Limitation

This research study has limitations. The quality, number of databases accessed, and the search strategy utilized may have limited the literature review. Big Data is a relatively new topic and therefore the quantity and quality of publications on the topic is limited. Additionally, publication and research bias cannot be ruled out due to research evaluation to determine articles study relevancy.

### Practical Implications

Overall, there is a positive view of big data analytics as a way to reduce health care costs and improve health care outcomes. Hospitals and health care providers will need to invest in the future of health care through investing in the ability to integrate enormous amounts of data and then derive meaningful information. This initial investment in big data will require obtaining the necessary hardware and software with the capability to manage these quantities of data and subsequent time and investment in planning and collaboration of multiple hospitals and health care providers to share data and remove it from silos. It will be necessary to train individuals capable of developing the tools needed to clean, integrate, analyze, and comprehend the data. Lastly, physicians will need to review their role as individuals seeking improved patient outcomes, expanding their mindset to incorporate interpreting results from knowledge generated from big data analytics regarding the whole patient pool instead only the results of labs from few patients. Furthermore, hospital administration will need to view investment in big data analytics as a way to reduce costs through the mass prevention of disease, unnecessary readmissions, and reduction of inefficient treatments. Due to the nature of healthcare data, healthcare providers also need to keep in mind their legal obligations to ensure the privacy of their patients, and make sure necessary security safeguards are in place in the implementation of big data technology.

## CONCLUSION

Obtaining meaningful information from Big Data in healthcare will have a positive impact on healthcare outcomes in the U.S and has the potential to eventually reduce the cost of healthcare though the unlocking of proactive knowledge. However, obtaining that meaningful information is fraught with challenges such as obtaining and managing usable healthcare data, lack of analytical expertise, security and privacy concerns innate to healthcare, and physician reluctance similar to the reluctance of EHR implementation.

## REFERENCES


Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big Data In Health Care: Using Analytics To Identify And Manage High-Risk And High-Cost Patients. *Health Affairs*, *33*(7), 1123-1131. doi:10.1377/hlthaff.2014.0041

Baro, E., Degoul, S., Beuscart, R., & Chazard, E. (2015). Toward a Literature-Driven Definition of Big Data in Healthcare. *Biomed Research International*, *2015*, 639021. doi:10.1155/2015/639021

Belle, A., Thiagarajan, R., Soroushmehr, S. R., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big Data Analytics in Healthcare. *Biomed Research International*, *2015*, 370194. doi:10.1155/2015/370194

Burtch, L. (2015). *Salaries of Data Scientists*. Retrieved November 1, 2015 from http://www.burtchworks.com/files/2015/05/DS-2015_DS-vs-PAP.pdf

Center for Disease Control and Prevention (CDC). (2012). *Meaningful Use: Introduction* Retrieved September 25, 2015 from: http://www.cdc.gov/ehrmeaningfuluse/introduction.html

Chute, C. G., Ullman-Cullere, M., Wood, G. M., Lin, S. M., He, M., & Pathak, J. (2013). Some experiences and opportunities for big data in translational research. Genetics in Medicine, 15(10), 802-809.

Cohen, I. G., Amarasingham, R., Shah, A., Bin, X., & Lo, B. (2014). The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care. *Health Affairs*, *33*(7), 1139-1147. doi:10.1377/hlthaff.2014.0048

Demchenko, Y., Zhao, Z., Grosso, P., Wibisono, A., & De Laat, C. (2012). Addressing big data challenges for scientific data infrastructure. In *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on* (pp. 614-617). IEEE.

Groves P, Kayyali B, Knott D, & Van Kuiken S. (2013). The 'big data' Revolution in Healthcare, *McKinsey & Company*, Retrieved September 25, 2015 from: http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care

Goozner, M. (2014). Better rules needed to boost use of big data. *Modern Healthcare*. 44(28) p. 38

Herland, M., Khoshgoftaar, T., & Wald, R. (2014). A review of data mining using big data in health informatics. *Journal Of Big Data*, 1(2) doi:10.1186/2196-1115-1-2

Jee K, & Kim GH. (2013). Potentiality of Big Data in The Medical Sector: Focus on How to Reshape the Healthcare System. *Healthcare Informatics Research*. 19(2):79-85.

Li, Y., & Chen, L. (2014). Big Biological Data: Challenges and Opportunities. *Genomics, proteomics & bioinformatics*, *12*(5), 187-189.

Maguire, J., & Dhar, V. (2013). Comparative effectiveness for oral anti-diabetic treatments among newly diagnosed type 2 diabetics: data-driven predictive analytics in healthcare. *Health Systems*, *2*(2), 73-92.

Martin-Sanchez, F., & Verspoor, K. (2014). Big data in medicine is driving big changes. *Yearbook of medical informatics*, *9*(1), 14.

McCarthy, C. J., & Hart, S. (2011). Designing groups to meet evolving challenges in health care settings. *The Journal for Specialists in Group Work*, *36*(4), 352-367.

McClellan, M., & Rivlin, A. (2014). Improving health while reducing cost growth: what is Possible? *Engelberg Center for Health Care Reform at Brookings.* Retrieved November 1, 2015 from http://www.brookings.edu/~/media/events/2014/04/11%20health%20care%20spending/improving_health_reducing_cost_growth_mcclellan_rivlin.pdf

Modern Healthcare. (2014). Applying Data and Analytics to the Provision of Better Healthcare. (2014). *Modern Healthcare*, *44*(35), 33.

Nash, D. B. (2014). Harnessing the Power of Big Data in Healthcare. *American Health & Drug Benefits, 7*(2), 69–70.

Naylor, K., T., Kudlow. P., Li., F., and Yuen, K. (2011) Improving healthcare with information technology. *University of Western Ontario Medical Journal* 80:1 Retrieved September 27, 2015 from: http://www.uwomj.com/wp-content/uploads/2011/08/Naylor.pdf

Ola, O., & Sedig, K. (2014). The challenge of big data in public health: an opportunity for visual analytics. *Online journal of public health informatics*, *5*(3), 223.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, *2*(1), 3.

Sarasohn-Khan, J. (2014, July). *Big Data Come to Health Care...With Big Challenges*. Retrieved from http://healthpopuli.com/2014/07/09/big-data-come-to-health-care-with-big-challenges-health-affairs-july-2014/

Ward, J. S., & Barker, A. (2013). Undefined by data: a survey of big data definitions. Retrieved December 6, 2015 from http://arxiv.org/abs/1309.5821

Yan, J. (2013). Big Data, Bigger Opportunities. Retrieved September 27, 2015 from http://www.meritalk.com/pdfs/bdx/bdx-whitepaper-090413.pdf

Yao, W., C. Chao-Hsien, and Z. Li. "The Use of RFID in Healthcare: Benefits and Barriers." Proceedings of the 2010 IEEE International Conference on RFID-Technology and Applications (RFID-TA) (2010): 128–34