https://doi.org/10.48009/3_iis_2016_1-10

Issues in Information Systems
Volume 17, Issue III, pp. 1-10, 2016

IMPACTS OF HEALTHCARE BIG DATA: A FRAMEWORK WITH LEGAL AND ETHICAL INSIGHTS

Mark Fox, Indiana University South Bend, mfox1@iusb.edu
Ganesh Vaidyanathan, Indiana University South Bend, gvaidyan@iusb.edu

ABSTRACT

Healthcare organizations have amassed massive amounts of data in the past years. In addition, with the prolific use of social media and the Internet of Things, big data analysis has captured the attention of many researchers. Big data may be used for the common good, for example, to predict an outbreak of a new virus. However, healthcare data is unique as it profiles the privacy of patients and is protected by federal and state laws. Two main issues arise: first, how big data analysis can be implemented and minimizing the risk of leaking patient data and, second, how big data analysis can be implemented while being protected regarding patient privacy. To understand these issues, this study establishes a framework encompassing five dimensions of healthcare big data that includes legal and ethical factors, patient privacy risk factors, security threat and data breach factors, preventive practice factors, and data application factors.

Keywords: Big Data, Legal, Ethical, Healthcare, Framework

INTRODUCTION

Big data is generating a lot of excitement in every industry, including healthcare. Big data describes the large volume of data, including both structured and unstructured data. Volume, velocity, variety, variability, and veracity characterize big data (Hitzler & Janowicz, 2013). Organizations collect huge volumes of data from business transactions, social media and other pertinent sources. Velocity is the speed of data creation which may be generated, which could be as quickly as real-time (or nearly in real-time) from batch to streaming data. Variety of data includes structured, unstructured, numeric, transactional, patient records, text document, email, video, audio, social media tweets, likes, and patient profiles. Trending as well as daily, seasonal, and event triggered peaks in data defines the variability in big data. Apart from the required volume, velocity, variety, and variability in big data applications, the data must have veracity—quality and accuracy—to produce credible results.

In healthcare, big data is ultimately being driven by the concept of personalized medicine that will improve patient care (Costa, 2013). As Issa et al (2015, p. 293) observe: "We are in the era of the '-omics', wherein an individual's genome, transcriptome, proteome and metabolome can be scrutinized to the finest resolution to paint a personalized biochemical fingerprint that enables tailored treatments, prognoses, risk factors, etc." Big data provides the digitized information that can be used for personalized medicine.

Big data can be analyzed for insights and patterns that may lead to better decisions and strategic initiatives. In the medical context, big data has the potential to reveal hidden insights to improve patient care (Adamson, 2016). There are a number of initiatives to use big data for common good. HealthConnect, a Kaiser Permanente implementation, ensures data exchange across all medical facilities and promote the use of electronic health records. AstraZeneca and HealthCore have joined in alliance to determine the most effective and economical treatments for some chronic illnesses and common diseases based on their combined data (Kayyali, Knott, and Kuiken, 2013).

HIPAA compliance is critical as the privacy and security of patient data is of utmost importance. Security goals include authentication, availability, integrity, and confidentiality. Big data implementation is based on integrating data from various sources. The management of integrating data and, in particular, security of data during and after such integration poses major challenges. There are some options for healthcare organizations to ensure the security of big data. They can use commercial software rather than open source software which may have questionable

security (Anderson, 2016). Some commercial software, such as Cloudera, has implemented a Payment Card Industry (PCI) compliant Hadoop environment supporting authentication, authorization, data protection, and auditing (Anderson, 2016).

However, data privacy is a concern when opening up access to a large, diverse group of users (Anderson, 2016). There is no problem if a healthcare institution grants access to a few data scientists and often research uses of such data will also be governed by IRB protocols. Big data in healthcare can be used to predict epidemics, improve quality of life, and avoid preventable deaths. While it is possible to predict an outbreak of a new virus using big data analysis and use the *a priori* knowledge from the analysis to prepare before the outbreak, it is a predicament as data from healthcare organizations cannot be made available to researchers due to HIPAA compliance. Key questions in this regard are: Should the data be made available? Who owns the data? Who should own data? To answer these questions, this paper focuses on the following issues:

- How can big data analysis be implemented while minimizing the risk of leaking patient data?
- How can big data analysis be implemented while protected within laws regarding patient privacy?

To understand those issues, this study establishes a framework encompassing five dimensions of healthcare big data that includes legal and ethical factors, patient privacy risk factors, security threat and data breach factors, preventive practice factors, and data application factors.

The paper is structured as follows. Next, we address privacy in general and the legal issues with respect to privacy of big data. In the following section, we include an analysis of ethics with respect to big data. Using the knowledge gained from the legal issues of privacy in big data, we will construct a framework of privacy with respect to healthcare big data in the next section. The final section includes conclusion and future research prospects.

PRIVACY IN BIG DATA ANALYTICS

In healthcare, Electronic Medical Records (EMR) can generate large volumes of data, but neither the volume nor velocity of such data is enough for big data analytics. Adamson (2016) points out that only a small fraction of the tables in an EMR database are relevant to healthcare data analytics and the majority of the data could be considered recreational. Smart phone apps and, more recently, wearable sensory devices such as Fitbit have been used by consumers to track their health progress. This data is being uploaded into servers to be compiled and potentially used for big data analysis. This personal data can be used along with EMRs to predict the state of the health of the general public and even allow specific health problems to be spotted before they occur (Marr, 2015). Amidst this backdrop, a number of problems loom regarding security and privacy. Cyber thieves routinely target medical records. The Federal Bureau of Investigation (FBI) warned healthcare providers to guard their data against cyber-attacks after one of the largest U.S. hospital operators, Community Health Systems Inc., reported that Chinese hackers have stolen the personal information of 4.5 million patients (Humer and Finkle, 2014). In 2014, hackers stole names and addresses relating to 80 million patients from Anthem, one of the large US health insurance companies. Fortunately, details of patient illnesses and treatments were not exposed. Can the potential leak of patient data be minimized and protect the privacy of patient data?

Alliances between medical institutions and data professionals have formed to implement big data analytics. For example, Pittsburgh Health Data Alliance plans to draw data from multiple sources including medical records, insurance records, wearable sensors, personal data, and social media data. The alliance is working towards a goal to offer tailored healthcare packages to patients (Marr, 2015). Can such activities be sustained while protecting patient privacy?

Various categories of organizations such as health plans, health care clearinghouses, business associates, and health care providers are covered by HIPAA. HIPAA recognizes that health care providers and health care plans also rely on other organizations that use or disclose information on their behalf. Accordingly, business associates include:

(i) A Health Information Organization, E-prescribing Gateway, or other person that provides data transmission services with respect to protected health information to a covered entity and that requires access on a routine basis to such protected health information.

(ii) A person that offers a personal health record to one or more individuals on behalf of a covered entity.

(iii) A subcontractor that creates, receives, maintains, or transmits protected health information on behalf of the business associate.

Hence, business associates are organizations that perform data analysis and data aggregation activities, such as those that would be necessary for health care data analytics.

The Privacy Rule permits the disclosure of information from, say, health care providers to business associates, but only does so "if the providers or plans obtain satisfactory assurances that the business associate will use the information only for the purposes for which it was engaged by the covered entity, will safeguard the information from misuse, and will help the covered entity comply with some of the covered entity's duties under the Privacy Rule." (HHS, 2016a).

HIPPAA is relevant to data analytics in terms of two key provisions. First, and generally, it outlines the HIPAA Privacy Rule (which relates to Protected Health Information or PHI). Secondly, and more specifically, it outlines the HIPAA Security Rule, which deals with electronic Protected Health Information. The Privacy Rule is intended to safeguard the privacy and use of "protected health information" (PHI). To understand what PHI is we first need to understand what Individually identifiable health information (IIHI) is. This is defined as health information that:

"(1) Is created or received by a health care provider, health plan, employer, or health care clearinghouse; and
(2) Relates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual." (HHS, 2016a).

This information that could identify or could be reasonably be used as the basis for identifying an individual. Such information not only includes healthcare information, but also includes demographic and financial information that could be used to identify an individual. Turning now to PHI. This is a subset of IIHI that includes data that is:

"(i) Transmitted by electronic media;
(ii) Maintained in electronic media; or
(iii) Transmitted or maintained in any other form or medium." (HHS, 2016a)

However, several exceptions apply, notably that PHI *excludes*:

"(i) Education records covered by the Family Educational Rights and Privacy Act, as amended, 20 U.S.C. 1232g;
(ii) Records described at 20 U.S.C. 1232g (a) (4) (B) (iv); and
(iii) Employment records held by a covered entity in its role as employer." (HHS, 2016a)

The question for big data analysts is how to get data in a form that does not constitute IIHI. The answer to this is found in the process of de-identification, i.e., creating health information "that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information." (HHS, 2016a, Section 164.514(a) of the HIPAA Privacy Rule)

A key point that Malik (2013) makes is that personally identifiable data needs to be masked before it is transmitted. Malik also talks about key security principles that need to be adhered to: "separation of duties, separation of concern, principle of least privilege, and defense in depth …" (page 9). There are two general means for de-identification that HIPAA allows, namely "expert determination" and "safe harbor". The expert determination method involves:

"A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:
(i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
(ii) Documents the methods and results of the analysis that justify such determination;" (HHS, 2016b)

The second means of de-identification, *safe harbor*, allows for the removal of eighteen "identifiers of the individual

or of relatives, employers, or household members of the individual" (HHS, 2016b). These identifiers include name, telephone numbers, SSNs. Further, the entity covered by HIPPA would "not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information." (HHS, 2016b). Safe harbor de-identification does provide a minor risk of re-identification (estimated at 0.04%) (Sweeney, 2007).

Bearing in mind that the Security Rule relates to electronic protected health information, related legislation (the Health Information Technology for Economic and Clinical Health, "HITECH" Act) specifies particular sections of the Security Rule that business associates must comply with. Business associates protect ePHI through appropriate administrative, physical and technical safeguards. HITECH is based on the premise that there will be increased sharing of ePHI and therefore more stringent data security and accountability for privacy breaches are needed.

Another key provision of HITECH involves notifications of breaches (HHS, 2016c). Breach notifications are requires business associates or their associates to provide notifications of any breach they notice or reasonably should have noticed "without unreasonable notice and in no case later than 60 days after discovery of a breach" (HHS, 2016c). In the case where contact information is available, written notification is requited to occur via first class mail. Notice to the media within given states are also required if 500 or more residents of those states were, or could reasonably be considered, to have been affected by the breach.

Under HITECH, business associates are also directly liable for violations of HIPAA rules. Areas where business associates can be liable include failure to comply with the Security Rule, impermissible use and disclosure, and failure to provide breach notifications. Patil and Seshadri (2014) observe that HIPAA is only a starting point in ensuring data security and patient privacy: "At the outset, patient information is stored in data centers with varying levels of security. Moreover, most healthcare data centers have HIPAA certification, but that certification does not guarantee patient record safety. The reason being, HIPAA is more focused on ensuring security policies and procedures than on implementing them. Furthermore, the inflow of large data sets from diverse sources places an extra burden on storage, processing and communication." (page 763).

ETHICS IN BIG DATA ANALYTICS

Big data involves trade offs between patient privacy and gaining optimal precision from the available data. Perfect de-identification of health care data, even using HIPAA de-identification practices) is not possible (Barth-Jones & Janisse, 2014).

A focus on big data has also raised concerns that patient-care-providers could be impaired due to the focus on data collection itself (Medical Ethics Advisor, 2013). The potential for commercial use of the data, for example, to market medicines directly to consumers has also been raised as a concern (Medical Ethics Advisor, 2013). In the realm of public health concerns have been raised about global justice (Vayena et al., 2015). For example, ethical concerns arise if access to big data is limited to those who can afford it, or if the concerns of those analyzing big data focus mainly on those who can afford certain sorts of treatments.

FRAMEWORK

The number of threats to the security and privacy of patient information have expanded over the years (Ponemon, 2015). Data breaches alone have been estimated to cost the healthcare industry about $6 billion. The Ponemon study reported that malware attacks caused security breaches in 78% of healthcare organizations. Moreover, a majority of healthcare organizations failed to perform risk assessments for security incidents. The study also found that healthcare organizations are struggling to comply with federal and state privacy and security regulations.

Privacy and ethics are intertwined in healthcare data analysis. Even before big data became vogue, researchers in the general area of healthcare data analysis have questioned awareness and appropriateness of such data and the lack of

respect and patient consent to retrieve data for research (Robling et al. 2004). They also expressed fear of unauthorized data access and highlighted public concerns when accessing medical records without patient consent. Chen et al (2012) suggested gleaning of clinical data and knowledge for a deeper understanding of patient disease patterns. Patil and Seshadri (2014) called for collection, linkage, and analysis of multidimensional data in real-time in patient care with a goal to understand disease control and prevention of adverse health events. However, retrieval of such data may pose many ethical and legal issues. Security treats continue when data is stored in cloud. Storage of big data in cloud is prevalent but users who outsource data are concerned about privacy. As patient and personal information are continued to be collected, this concern has become serious. Information on individuals gives rise to concerns on profiling, stealing, and loss of control (Tene & Polonetsky, 2012). Moreover, the volume and diversity of big data cannot be protected by traditional security solutions due to the sheer complexity of distributed software solutions with varying data sources and formats (Patil & Seshadri, 2014).

Security breaches induce disclosure of personal information and have an impact on patient privacy. Medical identity thefts have doubled in the past five years to 2.3 million in 2014. However, most healthcare organizations have no protection services for patients whose medical information have been breached (Ponemon, 2015). Security breaches To thwart security breaches and to comply with regulations, organizations have established preventive practices. Chen et al (2015) indicate that preventive processes such as Security Education, Training, and Awareness (SETA) programs and security monitoring has impacted security culture in organizations. From a legal and ethics point of view, patient consent, transparency, and global justice may be seen as important factors. Global justice of big data affordability without limitations to people is an ethical concern (Vayena et al., 2015). From the above mentioned literature, a number of impact factors on big data analysis can be derived. The current literature highlights certain factors that are important to big data analysis. Those factors have been extracted and captured to form the dimensions. Figure 1 summarizes those factors as dimensions of impacts of healthcare big data analytics.

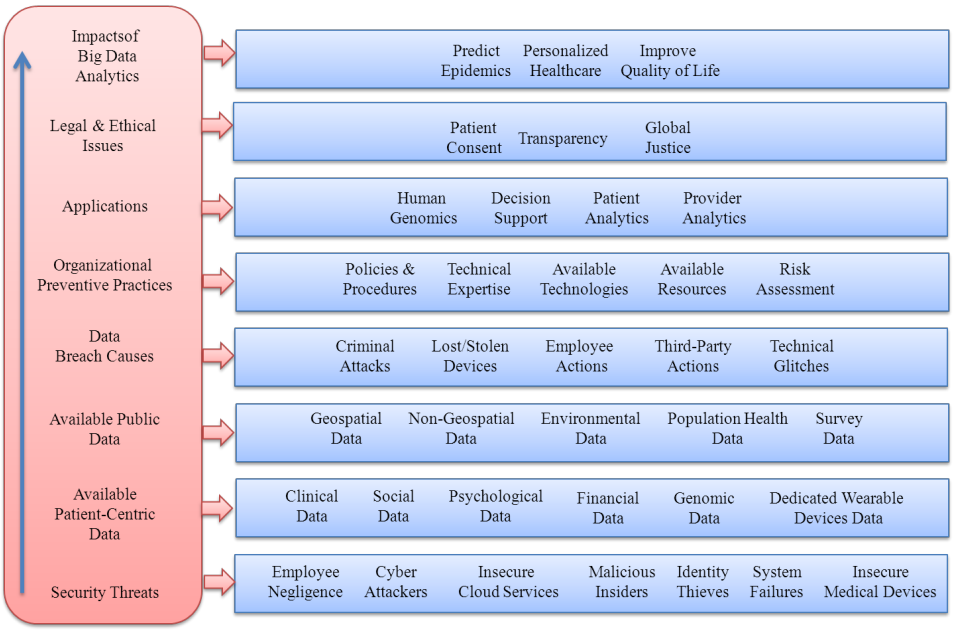

Figure 1. Impact dimensions of Healthcare Big Data Analytics

In Figure 1, security threats are targeted towards patient-centric data containing clinical, financial, and genomic data among others. Publically available data can be hacked and modified to produce false impacts. Data breaches such as criminal attacks and employee actions can be identified by organizations and using preventive practices, they can try to minimize compromises to privacy. Data from disparate sources can be analyzed to provide false predictions. For example, patients looking for side effects online and not making progress with their Fitbit device may be predicted for some type of illness but only clinical data can actually confirm that illness (Groves et al., 2013). Applications to predict and confirm have to be developed and used for proper medical intervention. Ethical choices need to be examined as well as legal outcomes need to be considered right from extraction of big data through implementation of its outcomes. These five dimensions can impact big data analytics to contribute big benefits for common good.

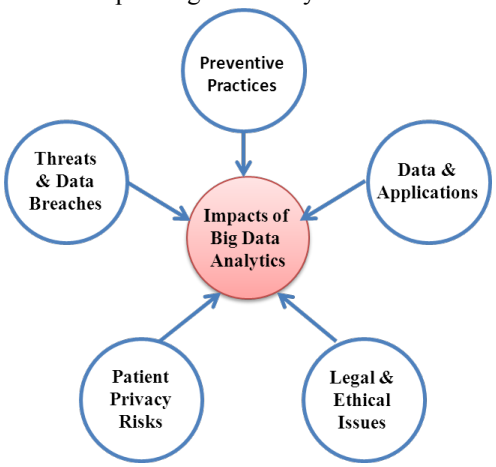

Figure 2. Impacts of Healthcare Big Data Analytics Framework

Figure 2 illustrates the framework of project culture is also formulated using the impact dimensions of healthcare big data analytics. In this section, each one of the five project culture dimensions are defined and described.

Security Threats

The potential for unauthorized access of data by third parties is a concern that arises from big data's use of the Internet, cloud computing and of data pooling (Feldman, 2012). The benefits of healthcare data analysis will not be complete if security challenges are not considered (Kotz et al. 2015).

Today's sophisticated medical devices like infusion pumps and vital-sign monitors are networked. Such network capable medical devices are vulnerable to security threats and require security protection mechanisms and software against threats. Organizations need to use audit logs for all their health systems to monitor inappropriate behavior (Gunter et al., 2011). The ability to detect anomalies as well as automated analysis of audit logs is required in such medical systems.

The perceptions of healthcare organizations on security threats are mainly concerned about lost or stolen devices, phishing, and malware attacks. 96% of respondents agreed that security threats are due to lost or stolen devices and about 83% of respondents reported phishing and malware attacks being the culprits. The respondents also said that they are more worried about employee negligence than cyber attackers and use of public cloud services (Ponemon, 2015).

Data Breaches

Security breaches lead to disclosure of personal information and violate HIPPA, and therefore are a profound impact on patient privacy (Kotz et al. 2015; Patil & Seshadri, 2014). About 38% of the causes of data breaches were lost paper files, 27% related to misplaced portable memory devices and only 11% were due to hackers (Jaeger, 2013).

Malicious insiders are a big threat because they have the knowledge, resources, and access to systems and devices in organizations (Vance et al., 2013). The root causes of data breaches in healthcare organizations are criminal attacks, lost or stolen devices, unintentional employee actions, third-party snafus, technical system glitches, malicious insider, and intentional non-malicious employee actions (Ponemon, 2015). A survey revealed that 45% of respondents said that breaches in healthcare organizations were due to criminal attacks and about 43% of respondents dis that the breaches were due to lost or stolen computing devices. The same survey claimed that 69% of the respondents believed that audit assessment exposed data breaches and 44% of the respondents said that employees themselves detected data breaches (Ponemon, 2015).

Data

Diverseness in data is a key factor to big data. The integration of large amounts of pertinent clinical, financial, genomic, social, and environmental data is crucial for big data analysis to understand population health for disease control and predictive analysis (Pail & Seshadri, 2014). For example, researchers at the Johns Hopkins School of Medicine used data from Google Flu Trends to predict surges in a flu-related emergency room. Others have used Twitter updates to track the spread of cholera in Haiti after an earthquake (McAfee et al 2012). Big data can effectively use unstructured data such as email messages, news updates, images, social network comments, sensor outputs, cell phone signals, etc. Chen et al (2012) view that collection of data from genomics and sequence datasets, electronic health records, and health and patient social media will help healthcare analytics. Miller (2012) concludes that two main sources of health big data are genomics-driven big data and payer–provider big data that includes insurance records, pharmacy prescription, patient feedback and responses. The data that has been targeted in healthcare organizations are medical files, billing and insurance records, payment detail, prescription details, and monthly statements.

Applications

Chen et al (2012) have suggested that applications such as human genomics, healthcare decision support, and patient community analysis are promising in healthcare data analytics. A new development in healthcare applications, the Medical Body Area Networks (MBANs), enables continuous monitoring of patient's condition by sensing and transmitting measurements such as heart rate, electrocardiogram (ECG), body temperature, respiratory rate, chest sounds, and blood pressure (Wang et al., 2015). MBANs will allow real-time and historical monitoring of millions of patients' health over the Internet and poses a potential risk to privacy intrusion.

Legal and Ethical Issues

Ethical concerns regarding big data include the potential misrepresenting either the quality of data or its limitations (Morgan, 2015) or using data in harmful or unnecessary purposes. For example, there is a risk that big data could predict irrelevant information and service providers can use that information to provide unnecessary services. (Richards & King, 2014).

Transparency is another key ethical issue in big data, particularly as this applies to health care. As Richards and King (2014, p. 419) note: "transparency fosters trusts by being able to hold others accountable." With regards to health care information, transparency means letting patients know what their personalized health information may be used for, how it could be used, and what protections are in place for the storage and sharing of this information. It also means informing patients of breaches or potential breaches and doing so in a timely manner.

Organizational Preventive Practices

Organizations rely on both policies and procedures to achieve federal and state compliances in order to secure patient sensitive information. In a survey conducted by Ponemon (2014), 55% of organizations reported that they have policies and procedures to prevent or quickly detect unauthorized patient data access, loss, or theft. Organizational risk assessment practices include manual, automated, and ad-hoc processes as well as incident report management techniques. In the same study, 49% of the respondents concluded that technical expertise to prevent or

detect unauthorized patient data access and knowledge about data breach notification laws are essential attributes to protect patient data privacy. Organizations have to allocate sufficient resources to prevent or detect unauthorized patient data access (Ponemon, 2014).

Impacts

Improved healthcare quality, improved long-term care, patient empowerment are some of the benefits of big data analytics in healthcare (Chen et al 2012). Other benefits include personalized medicine and individual analytics applied for patient profile, performance based pricing for personnel, disease patterns analysis, and improvement of public health in general.

CONCLUSION ND FUTURE RESEARCH

The Ponemon survey (2015) concludes that policies and procedures established by healthcare organizations continue on a trend to be effective in privacy and security of healthcare data. According to the same survey, only a few health organizations offer credit monitoring and identity monitoring of patient records. The survey shows that the percentage of criminal-based security incidents including malware attacks are on the rise. The study has prompted us to understand the vital factors that contribute to the success of big data analysis. We have responded by establishing a framework that depicts some of the important dimensions that lead to the success of healthcare big data analytics. The framework established in this paper addresses some of the important factors that need to be understood clearly so that healthcare organizations can minimize the risk of leaking patient data and protect patient privacy.

In light of the intimate nature of health care data "we must respect and protect it with the highest security possible … One of our biggest barriers to adoption is trust so we have security audits and make sure we exceed all of the current compliance and legislation." (Feldman, 2012, p.39). Tene & Polonetsky (2012) have prompted researchers to develop models where impacts of big data analysis balance individual privacy rights.

The framework established in this paper can be used to evaluate security programs in healthcare organizations and to improve such systems in order to promote the privacy rights of patients. An empirical investigation of the framework may be conducted to check to test its validity. Impacts of those dimensions on each other can be empirically validated and researched. The research community needs to address the practical challenges posed in this paper to achieve the high level of security that is warranted for successful big data applications.

REFERENCES

Adamson, D. (2016). Big Data in Healthcare Made Simple: Where It Stands Today and Where It's Going. *Health Catalyst.* Available: https://www.healthcatalyst.com/big-data-in-healthcare-made-simple.

Barth-Jones, D.C. & Janisse, J. (2014). Challenges associated with data-sharing: HIPAA de-identification. *The National Academies of Sciences, Engineering, and* Medicine. Available: https://www.nationalacademies.org/hmd/~/media/Files/Activity%20Files/Environment/EnvironmentalHealthRT/2014-03/Daniel-Barth-Jones_March2014.pdf

Chen, Y. Ramamurthy, K., & Wen, K. (2015). Impacts of Comprehensive Information Security Programs on Information Security Culture. *Journal of Computer Information Systems,* 55 (3), 11-19.

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly, 36*(4), 1165-1188.

Costa F.F., (2014). Big data in biomedicine. *Drug Discovery Today, 19*(4), 433-440.

Feldman, B., Martin, E.M., & Skotness, T. (2012). Big data in healthcare: hype and hope. *Dr. Bonnie, 360.* Available: https://www.ghdonline.org/uploads/big-data-in-healthcare_B_Kaplan_2012.pdf

Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The big data revolution in healthcare: accelerating value and innovation. *Mckinsey Quarterly,* 2. Available: https://digitalstrategy.nl/wp-content/uploads/E2-2013.04-The-big-data-revolution-in-US-health-care-Accelerating-value-and-innovation.pdf

Gunter, C.A., Liebovitz, D.M., and Malin, B. (2011). Experience based access management: A life-cycle framework for identity and access management systems. *IEEE Security & Privacy,* 9(5), 48-55.

Habte, M. L. (2014). Federal and state privacy laws: strategies for analysis of big data in healthcare. *Healthcare Informatics.* Available: http://www.healthcare-informatics.com/article/federal-and-state-privacy-laws-strategies-analysis-big-data-healthcare

HHS (2016a). Business Associates 45 CFR 164.502(e), 164.504(e), 164.532(d) and (e). *Health Information Privacy.* Available: http://www.hhs.gov/hipaa/for-professionals/privacy/guidance/business-associates/

HHS (2016b), Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Available: http://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

HHS (2016c). Breach Notification Rule. Available: http://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html

Hitzler, P., & Janowicz, K. (2013). Linked Data, Big Data, and the 4th Paradigm. *Semantic Web, 4*(3), 233-235.

Hirsch, R., & Deixler, H. (2014). HIPAA business associates and health-care big data: big promise, little guidance. *Bloomberg BNA.* Available: http://www.bna.com/hipaa-business-associates-and-health-care-big-data-big-promise-little-guidance/

Humer, C., and Finkle, J. (2014). Your medical record is worth more to hackers than your credit card. *Reuters.* Available: http://www.reuters.com/article/us-cybersecurity-hospitals-idUSKCN0HJ21I20140924

Issa, N.T., Byers, S.W., & Dakshanamurthy, S. (2014). Big data: the next frontier for innovation in therapeutics and healthcare. *Expert Review Clinical Pharmacology, 7*(3), 293-8.

Jaeger, J. (2013). Human error, not hackers, cause most data breaches. *Compliance Week, 10*(110), 56–57.

Kayyali, B., Knott, D., and Kuiken, S.V. (2013). The big-data revolution in US health care: accelerating value and innovation. *McKinsey & Company.* Available: http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care.

Kotz, D., Fu, K., Gunter, C., & Rubin, A. (2015). Privacy and Security for Mobile and Cloud Frontiers in Healthcare. *Communications of the ACM, 58*(8). 21-23.

Landi, W, & Rao, R.B. (2003). Secure de-identification and re-identification. *AMIA Annual Symposium Proceedings.* November 8-12, Washington D.C.

Malik, P. (2013). Governing big data: principles and practices. *IBM Journal of Research & Development, 57*, 3-4.

Marr, B. (2015). How big data is changing healthcare. *Forbes.* Available: http://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/#6c6d5cd432d9

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data. The management revolution. *Harvard Business Review, 90*(10), 61-67

Medical Ethics Advisor (2013). Big data in health care raises some ethical concerns. *Medical Ethics Advisor,* September.

Miller, K. (2012). Big Data Analytics in Biomedical Research. *Biomedical Computation Review.* Available: http://biomedicalcomputationreview.org/content/big-data-analytics-biomedical-research

Morgan, L. (2015). Big Data Ethics: 8 Key Facts to Ponder. Available: http://www.informationweek.com/big-data/big-data-analytics/big-data-ethics-8-key-facts-to-ponder/d/d-id/1322143?image_number=6

Patil, H.K. & Seshadri, R. (2014). Big data security and privacy issues in healthcare, *Proceedings of 2014 IEEE International Congress on Big Data,* Washington D.C., October 27-30, 762-765.

Ponemon (2015). Fifth annual benchmark study on privacy & security of healthcare data. *Ponemon Institute Research Report.* Available: https://www2.idexpertscorp.com/fifth-annual-ponemon-study-on-privacy-security-incidents-of-healthcare-data.

Richards, N. M., & King, J. H. (2014). Big data ethics. Wake Forest Law Review, 49, 393-432.

Robling, M.R., Hood, K., Houston, H., Pill, R., Fay, J., & Evans, H.M. (2004). Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study. *Journal of Medical Ethics, 30,* 104-109.

Sweeney, L. (2007). Testimony before National Committee on Vital and Health Statistics. Available: http://www.ncvhs.hhs.gov/transcripts-minutes/transcript-of-the-august-23-2007-ncvhs-ad-hoc-workgroup-for-secondary-uses-of-health-data-hearing

Tene, O. & Polonetsky, J. (2012). Privacy in the age of big data: a time for big decisions. *Stanford Law Review Online,* 63. Available: http://www.stanfordlawreview.org/online/privacy-paradox/big-data

Terry, N. (2012). Protecting Patient Privacy in the Age of Big Data. *University of Missouri-Kansas City Law Review, 81*(2), 1-31.

Vayena, E, Salathé, M., Madoff, L..C, & Brownstein, J.S. (2015) Ethical Challenges of Big Data in Public Health. Bourne PE, ed. *PLoS Computational Biology*, *11*(2), 1-7.

Vance, A., Lowry, P.B., & Eggett, D. (2013). Using accountability to reduce access policy violations in information systems. *Journal of Management Information Systems, 29*(4), 263–290.

Wang, L., Ranjan, R., Kolodziej, J., Zomaya, A. Y., & Alem, L. (2015). Software Tools and Techniques for Big Data Computing in Healthcare Clouds. *Future Generation Computer Systems, 43*, 38-39.