

# Big Data Analytics Role in Reducing Healthcare Costs in the United States

Judy Phillips  
Indiana University  
Bloomington, Indiana  
judkphil@iu.edu

## ABSTRACT

In the United States more money is spent on health care than in any other industrialized country in the world. Yet, health care access is often problematic and health care quality indicators are lower or mediocre as compared to other countries with similar economic status. Insights offered by Big Data Analytics can find solutions that will significantly lower costs and improve delivery of health care in the United States. These solutions have to potential can save billions of dollars in health care costs and to improve the quality of care for millions of Americans.

## KEYWORDS

I523, H12332, health care costs, predictive analytics, electronic health records, big data

### ACM Reference format:

Judy Phillips. 2017. Big Data Analytics Role in Reducing Healthcare Costs in the United States. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17)*, 11 pages.  
<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

## 1 INTRODUCTION

Health care spending in the United States greatly exceeds the spending of other industrialized countries. Americans spend 3 trillion dollars annually on health care. Health expenditures currently account for 17.6 percent of the Gross National Product (GDP) and are expected to increase at an average rate of 5.8 percent through 2025. Health care spending has exceeded growth of the Gross National Product (GDP) in 42 of the previous 50 years [2]. Health spending threatens the nations fiscal health [29]. Despite the excessive spending, the United States ranks among the worst on measures of health care quality, health access equity, and quality of life [22]. Policy makers do not know how to respond.

Big data analytics has the potential to help manage address some of the cost issues while simultaneously improving patient health outcomes. Big Data ability gives us the ability to combine and analyze data from a wide variety of sources in ways that have never before been possible. This new information is providing new and valuable insights into ways to provide more effective and efficient

patient care. The associations, patterns, and trends in big data may hold the key to reducing expenditures, improving care, and saving lives [29]. The information is being used to achieve more accurate and timely diagnoses, better match treatment plans to patient needs, and predict and identify at risk patients and populations [22]. Mobile applications are being used to monitor patient care in real time. Big data can reduce health care waste, improve coordination of care, expose fraud and abuse, and to speed up the research and development pipeline.

The cost savings estimates are substantial. McKinsey and Company estimates that Big data analytics has the potential to reduce health care costs in the United State by 12 to 17 percent. This equates to a savings of between 348 to 493 billion dollars annually [6].

Some of the tools and methodologies that big data uses to introduce efficiencies into the American health care system include: Outcome based reimbursement methodologies, electronic health records, medical device monitoring, predictive analytics, evidence based medicine, genomic analysis, and claim prepayment fraud analysis. Big data technologies is adding value and improving efficiency in almost every area of health care including, clinical decision support, administration, pharmaceutical research and development, and population health management.

## 2 COMPARISON TO OTHER COUNTRIES

According to the Organization for Economic Cooperation and Development (OECD), the United States spends 2.5 times per person than the average of OECD related industrialized nations. In 2016, the United States spent 9822 dollars per person annually on health care. In comparison, the average amount spent per person among all OECD nations was 4033 dollars. The next highest spender was Switzerland at 7919 dollars per person [28]. The average spending as a percentage of Gross National Product (GDP) among OECD nations was 9 percent. Switzerland was again the next highest spender at 12 percent of their Gross National Product (GDP) being spent on health care. According to a McKinsey and Company analysis, the United States spends 600 billion dollars more annually than the estimated benchmark amount as calculated based upon the countries size and wealth as compared to other OECD related nations [18].

The United States lags in many standard indicators of health quality. According to a Commonwealth Fund study of 11 developed countries in 2013, the United States ranked fifth in quality and worst in infant mortality. The United States also ranked last in the prevention of deaths from treatable conditions such as strokes, diabetes, high blood pressure and treatable cancers. The average life expectancy in the United States is 76.3 years. The average life expectancy among all OECD countries is 77.9 years. The incidence of obstetric trauma is 9.6 per 100000 births in the United States

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Conference'17, July 2017, Washington, DC, USA

© 2017 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

compared to 5.7 incidents per 100000 in other countries. The statistics for preventable hospital admissions also compare poorly in comparison to other nations. In the United States the hospital admission rate for asthma and COPD was 262 per 100000 in comparison to the average of 236 per 100000. Thirty eight percent of the population in the United States is obese. The average obesity rate in other countries at nineteen percent. The United States has fewer physicians and hospitals. In the United States, there are 2.6 practicing physicians and 2.8 hospital beds per 1000 population. This compares to an average of 3.4 physicians and 4.7 hospital beds on the average in the other countries [28].

The United States has material problems with health care access. Most other OEDC countries have achieved almost universal insurance coverages. On the average, 98 percent of persons in OEDC countries have health insurance. In the United States only 90 percent have health insurance. In addition, cost sharing requirements often make access additionally prohibitive. In 2016, 22.3 percent of the persons in the United States had skipped a medical consultation due to cost concerns. In comparison, the average percentage of individuals who had skipped medical visits due to cost in OEDC nations was 10.5 percent. In the United States 11.6 percent of the population had skipped taking a prescribed medication due to cost in 2016. This compares to an average of 7.1 percent of the population in other OEDC countries who reported foregoing foregone a prescribed medication due to cost [28].

### 3 HEALTH COST DRIVERS

Why is health care so much more expensive in the United States than it is anywhere else in the world? Some of the contributing factors include: the basic health care economic payment structure, inefficient and wasteful use of resources, medical errors, lack of transparency within the system, unnecessary administrative costs, and fraud and abuse.

#### 3.1 Health Care Payment Structure

Many of the cost issues can be contributed to the complex, uncoordinated, multi-payer payment structure. Private insurance companies, Medicaid, and Medicare are the primary payers. An individuals eligibility by payer is dependent upon factors such as employment status, income level, age, and whether or not they are disabled. Most citizens obtain private insurance through their employment. Individuals who are 65 years of age or older or disabled are eligible for Medicare. These individuals may also purchase private Medicare Supplement insurance on their own to pay expenses that Medicare does not cover. Low income individuals may be eligible for Medicaid. If an individual is not eligible for any of these programs, he can purchase individual health insurance from a private insurance company on his own. However, individual health insurance is expensive. According data from E-care, in 2016, the average monthly premium for an individual was 393 dollars per month. The average cost for family coverage was and 1021 dollars per month [39]. In addition, individual insurance policies often include fairly high cost sharing features. Even though subsidies are available through the Affordable Care Act to offset some of these costs, many people choose to forego insurance entirely due to the prohibitive expense.

The system is inefficient and flawed because the basic economic concepts such as supply and demand and competition do not work in this sector. This is because none of the players are incentivized to manage or reduce costs [3]. Consumers do not manage medical utilization because it is being paid for by a third party, the insurance company. Insurance coverage thus insulates patients from the true costs of medical care [3]. Providers are not incentivized to provide efficient, cost effective care. Most providers on paid via a traditional fee for service methodology. That is, providers are paid for each service that they provide. Traditional fee for service provider payment methodologies that reward health caregivers for quantity instead of quality often result in overutilization of unnecessary tests and treatment procedures. The structure is such that it encourages the production of inefficient and low value services [3]. Insurance companies pass the cost of services on to the consumers in the form of higher premiums year after year. The cost inflation cycle goes on and on.

Administrative waste is another result of the complexity of the United States multi payer payment structure. Each payer has their own rules and standards. Benefit and coverage options can vary dramatically among individuals even within the same insurance company. According the OEDC 2008 estimates, the United States spends 7.3 percent of health care expenses on administrative activities. This is more than any other country. Comparatively, Germany spends 5.6 percent, Canada spends 2.6 percent and France spends 1.9 percent [28]. Administrative activities include transaction related activities such as billing and claims payment, and regulatory compliance such as those required to comply with government and nongovernment accreditation and regulation including licensing requirements.

#### 3.2 Clinical and Operational Waste

McKinsey and Company estimates that clinical waste amounts to 273 dollars annually [29]. According to the Congressional budget office, 30 percent of United States spending is wasteful or not necessary [8]. There are two types of waste: operational, and clinical [3].

Operation waste results from duplication of services or inefficient production processes. An example would be a duplicate medical service because of lost medical records or the same service already being provided by another caregiver [3].

Clinical waste is created by the creation of low value outputs or care that is not optimally managed. One type of clinical waste is the spending on goods and services that provide marginal or no health benefit over less costly alternatives. Some clinical waste is the result of the uncertainty in the science of medicine. An example would be when a patient is misdiagnosed or when the treatment protocol is uncertain [3]. Other types of clinical waste may be symptoms of a flawed fee for service payment structure. These may include such things as over screening, excessive office visits, or the use of branded instead of generic drugs. Another example is when a newer or more modern treatment is marketed and sold even when it does not provide a better outcome as compared to the traditional treatment. An example was a 2 million dollar prostate cancer machine that was being marketed in 2014. It made the price of the procedure significantly more, but it did nothing to improve

the health outcome [8]. Other examples of types of treatment that are the result of clinical waste include avoidable emergency room use, unnecessary hospital admissions, and excessive antibiotic use [3].

### 3.3 Medical Errors

Medical errors cost the United States system between 17 and 29 billion dollars annually [3]. This amount could be as much as 1 trillion dollars a year if lost productivity is taken into account [27]. This compares to an estimate of 750 million in Canada [3]. The Institute of Medicine estimates that preventable medical errors claim between 44000 and 98000 lives in hospitals each year [3].

### 3.4 Fraud and Abuse

The National Healthcare Anti-Fraud Association estimates losses due to health care fraud at 80 billion dollars annually. Other industry sources estimate fraudulent related losses to be around 200 billion. This accounts for approximately 2 to 3 percent of total health care spending. Research indicates that only 5 percent of these losses are ever recovered [10].

## 4 BIG DATA

Big data refers to electronic data sets that are so large and complex that they cannot be managed with traditional hardware and software. A report delivered to the United States Congress in August 2012 defines big data as large volumes of high velocity, complex, variable data that require progressive techniques and technologies to capture, storage, distribution, manage, and analyze the information. Big data characteristics include variety, velocity, and veracity, and volume [29]. Health care data is big data because it involves the processing of overwhelmingly large complex data sets, from a wide variety of sources and a very rapid speed [29]. In addition, the data is extremely difficult to sort, organize, and decipher [11]. Recent advances in Big Data technology gives us the ability to capture, share and store healthcare data at an unprecedented pace.

### 4.1 Volume

The health care industry has always generated large amounts of data. Data is needed for record keeping, compliance and regulatory reporting and patient care. Historically, this data has been stored in hard copy format. Now, more and more data is being created and stored digitally. In 2011, there were estimated to be 150 Exabytes of health related data. The amount of health related big data is growing rapidly. It is expected to soon reach the zettabyte scale and then soon after that, into the yottabytes [29].

### 4.2 Velocity

Traditionally, health care data has been static: for example, paper files, x-ray films, and prescriptions [29]. Ironically, in many medical situations, the speed of the response can mean the difference between life and death. Increasingly, more and more of the data is being collected in real time and at a rapid pace. For example, medical monitoring devices information collect data continuously, and can support immediate response [29].

### 4.3 Variety

There is an enormous variety of data being collected. The data is in multimedia including images, video, text, numerical, multimedia, paper, and electronic records. Formats include structured, unstructured, and semi-structured. Sources of data include patients, physicians, hospitals, laboratories, research companies, insurance companies, and government agencies. Data comes from web and social media such as Facebook, twitter, health plan websites and smart phone applications. Machine to machine data comes from patient sensors. Biometric data is available such as fingerprints, genetics, hand writing information and imaging reports [29]. Physicians generate electronic medical records, physician notes, and medical correspondence. Pharmaceutical companies maintain research and development information in medical databases. The United States government houses databases concerning clinical drug trials. Data is collected by the United States Centers Disease Control and Prevention [6].

### 4.4 Veracity

The characteristic Veracity addresses whether the information is credible and error free. Veracity is extremely important in health care because life or death decisions on being based upon the information provided. There is a particular concern because interpretations of unstructured data such as physician notes could be incorrect or imprecise. Big data architecture, platforms, methodologies and tools are designed to take into account the uncertainties of big data analytics [29].

### 4.5 Unstructured Big Data

Unstructured data now makes up about 80 percent of the health care information that is available and is growing exponentially. Sources of unstructured data include: medical devices, physician and nurses notes, and medical correspondence. Being able to access to this information is an invaluable resource for improving patient care and increasing efficiency [22]. Big data technology gives us the ability to capitalize and make use of the valuable clinical information that is unstructured [15].

Traditional databases have well defined structures. The data exists in a table and column format, tables have well defined schemas, and each piece of data is stored within its own well defined space. Big data is not like that at all. Data is extracted from the source systems in its raw format. Massive amounts of this data are stored in a somewhat chaotic fashion in a distributed file system. For example, the Hadoop Distributed File System (HDFS) stores data in directories of files in a hierarchical form. The convention is to store files in 64 Megabyte files in the data nodes using a high degree of compression [15].

Big data is raw data. Big data is not cleansed or transformed in any way. No business rules are applied. The approach is to transform and apply business rules or bind the data semantically as late in the process as possible. In other words, the approach is to bind as close to the application layer as possible [15].

Big unstructured data is less expensive than traditional databases. Most traditional relational databases require propriety software that is associated with expensive licensing and maintenance agreements. Relational databases also need significant specialized resources for

design, administration, and maintenance. Because of its unstructured format and open source concept, big unstructured data is much less expensive to own and operate. Big data needs little design work and is easy to maintain. A Hadoop cluster is built using inexpensive commodity hardware and runs on traditional disk drives using a directed attached (DAS) configuration instead of an expensive storage area (SAN). The practice of storage redundancy makes the configuration more tolerable to hardware failures. Hadoop clusters are designed so that they are able to rebuild failed nodes easily [15].

Big unstructured data is more difficult to use. Traditional relational database users are able to access the data using a simple structured query language (SQL) that uses a sophisticated query engine that has been optimized to extract the data. Unstructured data is much more difficult to query. A sophisticated data user, such as a data scientist may be needed to manipulate the data. However tools are being developed to solve this problem. One tool is SparkSQL. This tool leverages conventional SQL for querying and works by converting SQL queries into MapReduce jobs. Another example is Microsoft Polybase which can join data from Hadoop and traditional databases and return a single result set [15].

To summarize, advances in Big Data technology, including data management of unstructured datasets and cloud computing are facilitating the development of platforms for more effectively capturing, storing, and manipulating large data sets sourced from multiple sources [29].

#### 4.6 Big Data Trends for Healthcare

The costs for storing and parallel processing are decreasing [22]. Previously, we had to choose what data to capture and store because storage costs were so high. Now we can capture and store everything [17]. The use of the Internet of Things is growing. Internet connected technology is everywhere and has become a common and accepted part of our culture. For example, wearable fitness devices are continuously generating health information and sending it to the cloud.

Another trend is the establishment of standards and incentives in the industry that encourage the digitization and sharing of health care data. The Health Insurance Portability and Accountability Act (HIPAA) establishes national standards for electronic healthcare transactions for the submission of claims. Claims are the documents that health providers submit to insurance companies to get paid. Such standards encourage the widespread use of Electronic Document exchange. These standards have made it possible to effectively and easily share and exchange medical information between providers and insurance companies [22]. Medicare and Medicaid have set up Electronic Medical Record (EHR) incentive programs to encourage professionals and hospitals to adopt and demonstrate meaningful use of EHRs. The Affordable Health Care Act (ACA) encourages the shift from fee for service to value based payment structures by financing initiatives to test new payment models [33].

### 5 VALUE BASED REIMBURSEMENT

One of the most important strategies that we can take to reduce health care in the United States is to change the way that we reimburse providers from the traditional fee for service methodology to

outcome based reimbursement. McKinsey and Company estimates that this strategy alone could reduce health care spending in the United States by 1 trillion dollars over the next decade [23]. This will also mitigate medical inflation because it will automatically promote preventative care and discourage the use of low value expensive technologies. Other benefits include: improved care coordination and the reduction of redundant care. All of this results in better health outcomes, and enhanced patient satisfaction.

With the fee for service payment structure providers are paid a fee for each and every service that they perform. This tends to encourage overutilization instead of the efficient use of medical resources. The United States tends to perform more and more expensive diagnostic services and treatment services than any other country in the world. The United States is well known for over testing and over treatment [26]. Hospitals are rewarded for preventable readmissions. Physicians are rewarded as much for a failed medical procedure as they are for a successful one. It is up to each individual physician to determine what tests and treatment services to order. From a clinical perspective, many of these tests are not medically necessary. This is a wasteful use of resources.

The goal of value based reimbursement structures are to align payment incentives with the administration of efficient, high quality medical care. Basing provider reimbursement on performance and patient outcomes encourages providers to work towards optimizing patient health instead of just providing more health care services. Caregivers are also incentivized to be more innovative and to search for ways to improve health care delivery [5].

Many payers, including private health insurance companies, Medicare, and Medicaid are starting to base reimbursement on value based incentives. The Affordable Health Care Act includes provisions to encourage the development and adoption of more effective care delivery models. Some payers are also starting to reward pharmaceutical companies by basing reimbursements on drug effectiveness [18]. Systems that have been adopted to date include: patient centered medical homes, episode based payments, global payments, shared savings programs, value based contracting, and population models, including accountable care organizations.

In the patient centered home model, the primary care physician coordinates the patients care and is rewarded for improving quality and reducing costs for individual patients. Another value based system is a population model that rewards providers for improving the health of the entire population [20]. An example of this type of program is an Accountable Care Organization (ACO). In Accountable Care Organizations, groups of doctors, hospitals, and other providers work together to provide coordinated care for patients. In Medicare supported Accountable Care Organizations, providers share in Medicare savings when they deliver high quality care and manage costs wisely [7].

Big Data Analytics can play an integral role in the development and testing of new payment model methodologies. The development and adoption of such models are still in the infancy stage. Big Data Analytics has the potential to provide information that will result in innovative payment structure and reward insights. Big data can also play a role developing clinical best practices and in identifying reasons for unjustified clinical variability in current practices.

Big Data will help to support the implementation of models that have already been adopted. Value based health care depends upon quality data collection and precise data analytics [20]. First, the data must be collected and analyzed in order to define what defines quality care. Big Data is collected and analyzed in order to establish clinical guidelines that promote a more rational use of specific diagnostic tests and treatment protocols. Second, this information must be made available to health care givers in a format that they can use for day to day clinical decision making. This is often in the form of a cloud based integration platform [20]. Next, data must be collected on an ongoing basis to provide feedback indicating whether the providers are meeting the defined standards and if not, what can be done to improve performance. In addition, the same data can benefit future patients when data analytics are taken beyond the initial reporting and are used to develop care protocols for entire patient populations [20].

One example is in which big data is being used to track and modify provider behavior is at Memorial Care, a six hospital system in Fountain Valley, California. Memorial Care uses physician performance analytics to analyze performance of hospital doctors and outpatient providers. So far, such tracking has resulted in the reduction 280 dollars per hospital stay for the average adult patient. This equates to a 13.8 million annual dollar savings for the Fountain Valley Hospital system [9].

## 6 ELECTRONIC HEALTH RECORDS

An Electronic Medical Record (EMR) is a digitized version of a patients medical chart. Whereas, an electronic medical record (EMR) typically includes information from one health provider, an electronic health record (EHR) includes information from multiple providers and documents all of the available information about the patient. The objective is to provide in one place, an electronic record of a patients health. This enables the sharing of information between providers. An electronic health record (EHR) contains medical history, diagnosis, medications, immunizations dates, allergy information, radiology images, and test results [36]. These records are made available to providers in real time. Electronic health record (EHR) systems often include electronic prescription subscribing systems. Also, they can include and be integrated with evidence based tools that help providers make immediate decisions about patients care. For example, an Electronic Health record system can also automatically check for problems such as medication conflicts and notify clinicians with alerts [13].

Electronic Health Records (EHRs) improve patient health care in so many ways. Physicians have better organized, more accessible, and more complete information about the patient. A clinicians ability to make an accurate diagnosis is improved. Easily accessible patient information reduces medical errors and unnecessary tests. There is a reduction in the incidence of duplicate tests. Coordination of care is improved because every caregiver is made aware of simultaneous care that is being provided by other caregivers. It easier to communicate critical clinical information to all applicable providers in a timely fashion. Because information is made available to providers in real time, there is a drastic reduction in the probability of errors caused by such things as allergic reactions or drug interactions, especially in emergency situations. Because

electronic subscribing allows physicians to communicate directly with the pharmacies, prescriptions are no longer lost or misread [13]. Preventative care improves because it is easier to track and manage when patients are due for vaccinations and screenings. It becomes possible to track prescriptions to determine if a patient has been following doctors orders [34]. Productivity is increased, overlap care is reduced, and coordination of care is enhanced [5]. In general, electronic health records (EHRs) improve quality of care enhance patient safety, and contribute to better outcomes [13].

Electronic Health records (EHRs) have significantly improved the ability to treat chronically ill patients. In the past, providers had to limit the decisions to the amount of information that was available to them at the time. The planning of care of a chronically diseased patient that had many symptoms was often mismanaged or delayed. Electronic health records (EHRs) enable the physicians to facilitate personalized treatment for these patients in a way that has never before been possible [5]. Providers have a comprehensive record of historical treatments, diagnostic data, medical history, and meticulous medical information all in one place [5]. The result is more efficient and effective treatment for chronically ill patients. There is a reduction in the number of potential side effects and an increase the patients quality of life all at a much reduced cost. [5].

Electronic health records (EHRs) also save money by reducing administrative costs. They reduce transcription costs and eliminate chart storage and access costs.

Between 2001 and 2014 Electronic Health record (EHR) usage in physician offices rose from 20 percent to 82 percent. According to Health Information Technology for Economic and Clinical Health (HITECH) research, electronic health records are being used in 94 percent of hospitals in the United States [34]. This amount of data that is being collected by large health systems and treatment centers around the country is massive [31].

## 7 PREDICTIVE ANALYTICS

### 7.1 Definition

Predictive analytics is the process of learning from historical data in order to make predictions about the future. The objective of predictive health analytics is to provide insights that enable personalized medical care for each individual patient [30]. Traditionally, physicians have always used predictive analytics, as they have always provided health care based upon what they know about the medical history of each individual patient. Predictive Health analytics seeks to supplement that knowledge with software tools that enable physicians to make more informed choices about the patients treatment based upon data from population cohorts [31]. Patients are directed to specific treatment plans based upon their specific conditions as compared to other patients in a similar cohort. This additional knowledge has the potential to provide physician with the information they need to provide a more effective treatment plans [31]. This becomes especially important for patients with complex medical histories who are suffering from multiple conditions [34]. Predictive analytics can also improve the accuracy of diagnosing patient conditions, better match treatments with outcomes, and better predict the specific patients at risk for disease [34].

Predictive analytics takes advantage of disparate data sources including: clinical, claims, research, sensors, social media, and genomic analysis.

Predictive analytics has the potential to materially reduce health care costs and improve patient care. Insights provided can in clinical decision support, prevent hospital readmission preventions, aid in adverse incidence avoidance, and help chronic disease management. In addition, predictive analytics can identify treatments and programs that do not deliver demonstrable benefits or that cost too much [29]. Some predictive models reduce readmissions by identifying environmental of lifestyle factors that increase risk or trigger adverse events so that treatment plans can be adjusted according. [29].

## 7.2 Patient Profile Analytics

Patient Profile Analytics is a specific type of predictive analysis in which patient profiles are developed to identify individuals who may be at risk for developing a disease and who could benefit from proactive management, such as lifestyle modifications. For example, patient profile analytics can be used to identify patients who may be at risk for developing diabetes.

## 7.3 Risk Stratification

One area in which predicting patients at risk can yield the greatest results is in identifying the patients who are at the greatest risk for the most adverse outcomes or costliest diseases [29]. Risk stratification is a methodology that can be used to identify and track the sickest and potentially costliest patients. The tool ranks or stratifies patients by potential risk and flags high risk cases for additional management. A risk stratification predictive tool takes into account risk factors such as missed doctors appointments in addition the symptoms. The tool enables doctors to intervene earlier to avoid hospital admissions and costly treatment [9].

## 7.4 Predictive Analytic Examples

Hundreds of thousands of dollars are spent on cancer care. Big data can be used to develop individualized, personalized cancer care programs. There is a web based application, which was sponsored by the National Cancer Institute that uses data from the Prostate, Lung, Colorectal, and Ovarian Cancer Screening trial together with patient risk factor and demographic data to help develop patient specific treatment regimens [6].

Congestive heart failure accounts for more medical spending than any other diagnosis. The earlier this condition is diagnosed, the easier it is to treat and to avoid dangerous and expensive complications. However, early manifestation is difficult to recognize and can easily be missed by physicians [22]. Machine learning algorithms have the ability to take into account many more factors than doctors alone. Predictive modeling and machine learning using large sample sizes can identify nuances and patterns that were previously impossible to see. As a result, machine learning models in the form of predictive analytics substantially improved clinicians ability to accurately diagnose persons with congestive heart failure [34].

Optum labs has developed a database with the electronic health records of over 30 million patients. They use the database to develop predictive analytic tools, the objective of which is to help doctors make Big data informed decisions that will improve patients treatment [22].

Parkland Hospital in Dallas, Texas uses predictive modeling to identify high risk patients in the coronary care unit and to predict likely outcomes when the patients are sent home. To date, Parkland has reduced readmissions for Medicare patients with heart failure by 31 percent. This equates to a 500000 dollar annual savings for this one hospital [9].

# 8 INTERNET CONNECTED MEDICAL DEVICES

. Internet connected medical devices are becoming more affordable and are being used more and more commonly. Gartner, the analysis firm, estimates that there will be more than 25 billion connected health devices by the year 2020 [15]. These devices collect data in real time and send information into the cloud. Devices include blood pressure monitors, pulse oximeters, glucose monitors, and electronic scales [15]. Some of these devices are being used as preventive care devices. Other devices are being used by health care providers to aid in the monitoring of patient conditions. Big Data is required because the process involves the capture and analysis of large volumes of fast moving data from in hospital and in home devices in real time.

## 8.1 Preventative Care

Millions of people are using mobile technology help live healthier lifestyles. Smart phone applications together with wearable devices such as Fitbit, Jawbone, and Samsung Gear Fit are designed to track the wearers exercise and activity levels [12]. Measures that are typically tracked include: the number of steps taken, number of calories burned, and number of stairs climbed. The objective is to encourage the users to take a more active role in their own health and wellbeing by being more physically active. Such devices can provide individuals with the information that they need to make more informed decisions, better manage their health, and to more easily track and adopt healthier behaviors [3]. In the future, it is conceivable that it will be routine to share this information with personal physicians and that it will be incorporated into regular health care management.

An individuals data can be uploaded from the device to the cloud where it is aggregated with information from other users [15]. In an initiative between Apple and IBM, a big data platform is being developed that will allow iPhone and Apple Watch users to share their data with IBMs Watson Health cloud health care analytics service. The information will use the combination of real time activity information in combination with biometric data to discover new medical insights [12].

## 8.2 Medical Monitoring

Remote monitoring enable medical professional to monitor a patient remotely using various technological devices. The devices can be worn by patients with health conditions at home and in medical facilities to stream data continuously to provide real time

remote patient monitoring. The devices can improve care by giving patients the ability to self-manage their conditions. Processing of real time events can be supplemented with machine learning algorithms to help provide physicians with information they need to make lifesaving interventions [22]. Patient care tends to be more proactive as patient vital signs can be monitored constantly [22]. Medical alerts can be sent to care providers such that they immediately aware of changes in a patients condition and can respond accordingly. Devices are often used for adverse risk prediction. Remote monitoring is typically used to monitor conditions such as heart disease, diabetes mellitus, and asthma. One example of the use of personal devices in patient care is pediatricians monitoring asthmatics to identify environmental triggers for attacks [6].

Real time systems analysis improves patient care while simultaneously reducing health care costs [5]. The devices are especially advantageous to individuals who reside in remote areas. Other advantages include: a reduced incidence of severe events, improved in patient safety, and high patient satisfaction levels.

## 9 PUBLIC HEALTH

Data science is being used in cities throughout the United States to predict and impede potential public health issues before they even start. For example, the Chicago Department of Public Health is modeling a program to target lead exposure in children. Information is collected from multiple sources such as, home inspection records, assessor values, health records, and census data. Predictive analytic algorithms then determine which houses have the highest potential risk. This information is then being incorporated into Electronic health records (EHRs) to automatically alert physicians to possible lead exposure risk concerning their pediatric and pregnant patients. Chicago has similar programs in place for food protection and tobacco control [14].

In San Diego, California the public health department routinely gathers big data health related information and publishes it on a user friendly web site. Information is gathered from sources such as marketing companies, mobile apps and demographic data. The data includes everything from vegetable consumption to diabetes occurrences. In one initiative, Live Well, the information was able to reduce the obesity rates at a local elementary school by 5 percent. A project that is currently in progress is the study and analysis of areas that have high rates of Alzheimers [19].

## 10 TRANSPARENCY

In the United States, health care price information is rarely made available to the health care consumers when they receive the care. Patients usually become aware of the costs when they receive the bill. The price of health procedures can vary radically by provider. Prices can even vary by payer for the same provider. In one study, it was estimated that consumers paid 10 to 17 percent less when they were given access to comparative price data. According a paper that was published by the American Economic Journal Economic Policy, if patients had access to price data and were willing to shop around, they could be pay significantly less for everything from routine screenings to knee surgery [2]. This tended to work best for consumers who had to pay for at least some portion of their own care.

Online pricing is a potential Big Data solution. Health related price web sites provide approximate prices for health services and procedures in fairly transparent formats. Online resources are now being made available by insurers, government agencies, internet companies and medical care providers. National insurers such as Anthem, United Health group, Humana, Aetna, and Cigna offer pricing tools to their customers. Some states, including New Hampshire, Maine, Oregon, and Massachusetts publish health pricing websites. The internet company Healthcarebluebook.com publishes information for all consumers in the United States [35].

The trend towards pay for performance reimbursement agreements will also help the cost transparency issue. This is because these pricing structures encourage health care providers to share information [5].

## 11 EVIDENCE BASED MEDICINE

Evidence based medicine (EBM) is an approach to medical practice that emphasizes the use of evidence from well designed and well conducted research to optimize decision making [37]. Evidence based medicine is an approach that supplements a clinicians knowledge, which may be limited by knowledge gaps or bias, with the formal and explicit information such as scientific literature or best practice methodology. Evidence based medicine eliminates guesswork for health care providers. Instead of having to rely only on their own personal judgement, providers can base treatment and protocols on credible scientific data [5].

Big Data analytics supports the research and development of evidence based best practice treatment protocols. Structured and unstructured data from a variety of sources is combined and big data algorithms are applied. Sources may include electronic medical records, financial and operational data, clinical data, and genomic data [29]. The aggregating individual data sets into big data sets enable analysis for conditions that typically have small populations. An example is the study of individuals with gluten allergies [18].

## 12 DRUG COSTS

It is a well known fact that drugs in the United States are priced higher than they are in other countries. There are many complicated contributing factors. One factor is lack of price regulation. Another factor is the economic structure of the health care system. Because the system includes multiple payers, there is no one payer with the power to effectively negotiate with the pharmaceutical companies as there are in other economies. Therefore, drug companies typically set drug prices at whatever the market will bear. Newly developed drugs usually have higher price tags. Big Data analytics cannot fix all of the problems with the drug market, but there are some areas in which it may have an impact: medication therapy management capabilities, drug comparison technology, and pharmaceutical research and development process improvements [4].

### 12.1 Medication Therapy Management

Big data analytics can play a significant role in improving the Medication Therapy Management process. Adverse drug events cost billions of dollars and result in thousands of patient deaths. Physicians and pharmacist are often overwhelmed to the point of not

having the time to implement appropriate drug therapies. Drug therapies are becoming more difficult to manage as more patients are taking multiple medications. Big Data cloud analytics are helping clinicians better co manage drug therapies, and to identify drug interactions, adverse side effects, and additive toxicities in real time. The results include a reduction in the number of patient deaths, emergency room visits, hospital admissions, and hospital readmissions [9].

## 12.2 Comparison of Competitor Drugs

In the research, there tends to be a lot of information about individual drugs. However, there is not much information about how drugs perform in comparison to their competitors. There needs to be more drug comparative information so that physicians are better informed about the true benefits of prescribing a more costly medication as compared to a less expensive or generic drug [4]. Big data technology can play a role in making such comparisons easier to accomplish.

## 12.3 Pharmaceutical Research and Development

Big Data can help to streamline the Pharmaceutical Research and development process. As a result, important drugs can be delivered to the market more quickly and the cost of drug development will be reduced.

Big data can enhance the process of identifying appropriate patients to enroll in the clinical trials. First, multiple sources are now available from which to select patients. For example, social media can be incorporated into the selection process and used in addition to physician information. Secondly, the participate selection criteria can include more inclusive factors, such as genetic information. This will enable better targeting of potential trial subjects which will result in more pertinent information, while at the same time shorting trail times and reducing expenses [24].

Trial can be monitored and tracked in real time. Real time trail monitoring can decrease the number of safety and operational issues. The result is the avoidance of potentially costly issues such as adverse events or unnecessary delays [24].

Electronically captured data can improve communication. Information can be shared easily between functions and external parties. All interested individuals can have access to the data at the same time including all departments, external partners, physicians, and contract research organizations (CROs). This will replace the issue of having rigid departmental data silos that hinder interaction [24].

Genomic and proteomic data can be used to speed drug development by providing the capability to better target treatments based upon genetic indicators [17].

## 13 ADMINISTRATIVE COSTS

According to the Institute of Medicine (IOM), the United States spends 361 billion annually on health care administration. This is more than twice our total spending on heart disease and three times our spending on cancer. Also according to the IOM, fully half of these expenditures are unnecessary [9].

One way that providers can save money is to digitize billing processes such as benefit verification, denial management, and

claims submission. A benefit verification that is done electronically costs 49 cents per patient. Comparatively, the same process done manually costs 8 dollars. It is estimated that providers could save 9.4 dollars annually by transitioning to electronic processing [21].

One example in which digitized processes are being used to streamline billing processes effectively is at the Phoenix Childrens Hospital in Arizona. They use a tool that automatically converts the clinical notes in the electronic health record (EHR) system to billable diagnostic codes [21].

## 14 FRAUD AND ABUSE

Common types of fraud and abuse include: billing for services that are not rendered, billing for more expensive procedures than were actually delivered, and the performance of unnecessary services.

In the past, the process of identifying misrepresented claims was tedious and time consuming. Big Data analytics makes it possible to easily identify and tag such claims. According to an article by RevCycle Intelligence, when there is repeated misrepresentation of some key fact or event, patterns are created in the data that can be detected by comparing the information to legitimate claims [10]. Anthem Health Insurance, one of the nations biggest insurance payers, uses big data and machine learning algorithms to tag suspicious claims as the claims are being processed. Tagged claims are then sent to clinical coding experts for review. The objective is to identify and address fraudulent claims before they are actually paid [10].

The Center for Medicare and Medicaid Services used predictive data analytics to identify and recover 210.7 million [22] in health care fraud in 2015. They did this by assigning risk scores to claims and providers via algorithms. This enabled the identification of abnormal billing patterns in claim submissions [10].

United Healthcare realized a 2200 percent return on their investment in a Hadoop Big Data platform that was used to identify and tag inaccurate claims using a systemic and repeatable methodology [22].

Other uses of Big Data analytics in fighting fraud and abuse include: identifying links between providers to access whether an identified unethical activity is being practiced by related providers, identification of a hospitals overutilization of services in a short time period, recognizing patients who are receiving health care services from different hospitals in different locations at the same time, and detecting prescriptions that are filled for the same patient in multiple locations at the same time. Big Data analytics can also utilize machine learning algorithms combined with historical information to detect trends in anomalies and suspicious data patterns.

## 15 GENOMICS ANALYTICS

Big data is playing a major role in the field of genomics and precision medicine. These technologies are helping clinicians choose the best treatment plan for individuals based upon their genetic makeup. Combining data from electronic health records (EHRs), clinical trials, and genetic testing gives researchers information to develop more effective treatments for complex diseases such as cancer and diabetes [25], and HIV. Genetic testing that has been made possible



by the mapping of the human genome will cut costs and improve survival rates [1].

One area in which genomics can have a dramatic impact is in pharmaceuticals management. In the United States, 300 million dollars are spent annually on pharmaceuticals. Studies indicate that between 20 to 75 percent of patients are not responsive to prescribed drug therapies. This can often be contributed to incorrect dosing or drug mismatches. However, 50 percent of the time it is because of a molecular mismatch between the patient and the drug. According to Alan Mertz, president of the American Clinical Laboratory Association, an estimated 30 to 110 billion can be saved by using genetic test to select a drug that is a precise match for the genetics of the patient. By using each patient's unique genomic profile, therapy can become more targeted and the instances of inappropriate care will be reduced [1].

For breast cancer patients, genetic testing can identify which 30 percent of women have an overabundance of the HER2 protein. Regular chemotherapy will not help these women, but a drug called Herceptin does. Having this information not only provides doctors with the information they need to prescribe the correct medication, it enables thousands of women to avoid needless harsh, expensive chemotherapy treatment. As a result, genetic testing has been shown to reduce the risk of death by 33 percent and the risk of recurrence by 52 percent for breast cancer patients. The resulting savings are estimated to be 24 thousand dollars per patient [1].

Genetic tests can help physicians select the appropriate drug for patients with metastatic colon cancer. According to one estimate, 700 million dollars could be saved annually by obtaining this information before administering treatment [1].

According to a 2006 Brookings/AEI estimate, using genetic tests to determine the appropriate dose of the blood thinner, warfarin, could save the United States 1.1 billion dollars annually. According to a study in June 2010 by the Journal of American College of Cardiology, this test could reduce hospital admissions that are caused by inaccurate dosages by 31 percent [1].

Genomic technology is also good for the United States economy. According to Battelle, a global research organization, human genome sequencing projects generated 796 billion in economic output, 244 billion in personal income and 3.8 million job-years of employment in the United States [1].

The process of gene sequencing continues to become more efficient and cost effective. It is expected to become a regular part of medical care in the near future [15].

## 16 TELEMEDICINE

Telemedicine is receiving medical treatment and advice remotely, on a computer over the internet with a physician [12]. Telemedicine has been in the market for 40 years, but with the availability of internet connected technology such as smartphones, wireless devices, and video conferences, it is becoming commonplace. It is primarily used for initial diagnosis, remote patient monitoring, and medical education. However, it is also being used for more complicated care such as telesurgery. Telesurgery is a technique in which doctors perform surgery via robots with the assistance of high speed real time data delivery technology [34].

Telemedicine is especially beneficial to patients who live in rural communities who may have to travel long distances to see a doctor or specialist. Telemedicine also gives doctors who are located in multiple locations the ability to discuss and share information. Telemedicine facilitates medical education by giving caregivers the ability to observe and be trained by subject experts no matter where their location.

Telemedicine has the potential to significantly reduce costs by reducing the number of outpatient and hospital visits [38].

## 17 USE CASES

Valence Health has built a data lake that they use as their primary data repository using a MapR Converged Data Platform. The system includes 3000 inbound data feeds and contains 45 different types of data including: lab test results, patient vitals, prescriptions, immunizations, pharmacy benefits, claims information from doctors and hospitals. The system reports dramatically better system performance than legacy system technology. For example, previously, it took 22 hours to process 20 million laboratory records. Now the processing time for the same number of records is 20 minutes. In addition, the new system requires less hardware [22].

The National Institute of Health developed a data lake which combines data sets from separate institutions. Now that all of the data is housed in the same location, analysis is more efficient and can be more easily shared [22].

United Healthcare uses Hadoop to maintain a platform with tools that they use to analyze information generated from claims, prescriptions, provider contracts, plan subscriber, and review information [22].

Novartis, a global healthcare company, uses Hadoop and Apache Spark to build a workflow system that aids in the integration, processing, and analysis of Next Generation Sequencing research as it relates to Genomic Analytics [22].

## 18 CHALLENGES

One of the most compelling challenges is clinicians' willingness and ability to change behavior based upon the information provided by the data. Studies have shown that it takes more than a decade of compelling clinical evidence before a new finding becomes common clinical practice. Therefore, we need to do a better job of working with clinicians on finding ways to use the data to provide higher quality care [17].

In health care, the privacy, security, and confidentiality of the patient is paramount [15]. Big data technology has inconsistent security technology. The Health Insurance Portability and Accountability Act (HIPAA) is a federal law that was passed in 1996 that sets a national standards to protect the confidentiality of medical records and personal health information. The HIPAA law is applicable to any component of the information can be used to identify a person. The protections apply to both electronic and non-electronic forms of information [32]. HIPAA regulations make it a federal offense to breach patient security. It is important to work with vendors who understand the importance of security [15]. Liason Technologies is one company that provides solutions to the healthcare and life sciences industry that has experience meeting the HIPAA security requirements [22].

Health care data has inconsistent formatting and definitional issues [17]. There is proliferation of data formats and data representations. There are inconsistent variable definitions. A value may have different meanings for different groups. For example, a cohort definition for an asthmatic patient often differs from one group of clinicians to another [16]. Big data has the challenge of bringing all of this information together.

Another issue is lack of technical experts. The manipulation and extraction of data from often unstructured data sets require special knowledge. There have been some recent changes in tooling that will make it easier for individualized with less specialized skills to manipulate the data. For example, Big data is starting to use include SQL as a tools for querying and data manipulation. Examples are Microsoft Polybase, Impala, and SQL Hadoop [15].

## 19 CONCLUSION

Big data analytics has huge potential to save the United States billions of dollars in health care costs while drastically improving health outcomes. Vast amounts of information is being captured, stored and combined in ways that offer insights have never before been possible. Innovative Big data tools are reducing medical waste, decreasing medical errors, fighting fraud, and keeping people healthier. Value based reimbursement solutions have the potential to revolutionize the health delivery system in the United States by motivating providers to find ways to deliver the best possible medical care with the most economical use of resources. The development of most of these tools is only in the preliminary stage. Therefore, we are only beginning to realize some of the potential benefits. Big data really does have the potential to bend the cost curve. Big data in health care is here to stay.

## ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski and the teaching assistants in the Data Science department at Indiana University for their support and suggestions to write this paper.

## REFERENCES

- [1] American Clinical Library Association. 2011. Genetic Testing Can Help the United States Cut Costs and Improve Care. Web page as article. (July 2011). <https://www.prnewswire.com/news-releases/genetic-testing-can-help-the-us-cut-costs-and-improve-health-care-126105103.html>
- [2] American Economic Association. 2017. Would Price Transparency Lower Healthcare Costs. Web page as article. (Feb. 2017). <https://www.aeaweb.org/research/health-care-price-transparency>
- [3] Effros Rachel M Palar Kartika Keeler Emmett B Bentleu, Tanya. 2018. Waste in the US Health System - A conceptual framework. *The Milbank Quarterly* 86 (Dec. 2018), 629–659. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2690367/>
- [4] Business Insider. 2016. Why the Price of Prescription drugs in the US is Out of Control. Web page as article. (Aug. 2016). <http://www.businessinsider.com/why-the-us-pays-more-for-prescription-drugs-2016-8>
- [5] Christian Ofori Boateng. 2016. Top 3 Ways Big Data Helps Decrease the Cost of Health Care. Web page as Article. (Nov. 2016). <https://go.christiansteven.com/top-3-ways-big-data-helps-decrease-the-cost-of-health-care>
- [6] CIO. 2015. How Big Data can save 400 billion in healthcare costs. Web page as Article. (Oct. 2015). <https://www.cio.com/article/2993986/big-data/how-big-data-can-help-save-400-billion-in-healthcare-costs.html>
- [7] CMS Centers for Medicare and Medicaid Services. 2017. Accountable Care Organizations. Web page. (Nov. 2017). <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO/>
- [8] Consumer Reports. 2014. Why is Healthcare so Expensive. Web page. (Sept. 2014). <https://www.consumerreports.org/cro/magazine/2014/11/it-is-time-to-get-mad-about-the-outrageous-cost-of-health-care/index.htm>
- [9] DataFlok. 2016. Five ways Big Data in reducing healthcare costs. Web page as article. (March 2016). <https://dataflok.com/read/5-ways-big-data-reducing-healthcare-costs/89>
- [10] Datameer. 2017. The Role of Big Data in Preventing Healthcare Fraud, Waste, and Abuse. Web page as article. (Sept. 2017). <https://www.datameer.com/company/datameer-blog/role-big-data-preventing-healthcare-fraud-waste-abuse/>
- [11] Digitalist. 2016. Can Big Data Analytics Save Billions in Healthcare Costs. Web page as Article. (Feb. 2016). <http://www.digitalistmag.com/resource-optimization/2016/02/29/big-data-analytics-save-billions-in-healthcare-costs-04037289>
- [12] Forbes. 2015. How Big Data in changing Healthcare. Web page as Article. (April 2015). <https://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare/#427274d12873>
- [13] Forbes. 2016. How an Electronic Health Record can Save Time, Money and Lives. Web page. (Dec. 2016). <https://www.forbes.com/sites/robertpearl/2016/12/01/how-an-electronic-health-record-can-save-time-money-and-lives/2/#4445b8275f57>
- [14] Harvard Business Review. 2014. How Cities are Using Analytics to Improve Public Health. Web page as article. (Sept. 2014). <https://hbr.org/2014/09/how-cities-are-using-analytics-to-improve-public-health>
- [15] Health Catalyst. 2017. Big Data in Healthcare Made Simple: Where it Stands Today and Where its Going. Web page as Article. (Oct. 2017). <https://www.healthcatalyst.com/big-data-in-healthcare-made-simple>
- [16] Health Catalyst. 2017. Five Reasons Healthcare Data is so Complex. Web page as article. (Nov. 2017). <https://www.healthcatalyst.com/>
- [17] Health Catalyst. 2017. Hadoop in Healthcare A no nonsense Q and A. Web page as article. (Nov. 2017). <https://www.healthcatalyst.com/Hadoop-in-healthcare>
- [18] Kayyali, Basel, Knott, David, Kuiken, Steve Van. 2013. McKinsey on Healthcare. Web page as Article. (April 2013). <http://healthcare.mckinsey.com/big-data-revolution-us-healthcare>
- [19] KQED Science. 2015. How San Diego is Using Big Data to Improve Public Health. Web page as article. (Aug. 2015). <https://www.kqed.org/futureofyou/2015/08/19/how-san-diego-is-using-big-data-to-improve-public-health/>
- [20] Liaison. 2017. Value Based Healthcare - The patient is the Center but Data is the Key. Web page as blog. (June 2017). <https://www.liaison.com/blog/2017/06/22/value-based-healthcare-patient-center-data-key/>
- [21] Managed Healthcare Execution. 2017. Five ways to reduce healthcare administrative costs. Web page as article. (April 2017). <http://managedhealthcareexecutive.modernmedicine.com/managed-healthcare-executive/news/five-ways-reduce-healthcare-administrative-costs>
- [22] McDonald, Carol. 2016. How Big Data is Reducing Costs and Improving Outcomes in Healthcare. Web page as Article. (June 2016). <https://mapr.com/blog/reduce-costs-and-improve-health-care-with-big-data/>
- [23] McKinsey and Company. 2013. The Trillion Dollar Prize. Web page as article. (Feb. 2013). <https://healthcare.mckinsey.com/sites/default/files/the-trillion-dollar-prize.pdf>
- [24] McKinsey and Company. 2017. How Big Data can Revolutionize pharmaceutical R and D. Web page as article. (Nov. 2017). <https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/how-big-data-can-revolutionize-pharmaceutical-r-and-d>
- [25] Patient. 2017. How Big Data Can Improve Health Care. Web page as article. (Nov. 2017). <https://patient.care/decks/privacy-technology/health-technology/how-big-data-can-improve-healthcare>
- [26] PBSO News Hour. 2012. Health Costs: How the US Compares with Other Countries. Web page as Article. (Oct. 2012). <https://www.pbs.org/newshour/health/health-costs-how-the-us-compares-with-other-countries>
- [27] Practice Fusion. 2017. EHR Adoption Rates 20 Must see stats. Web page as Article. (March 2017). <https://www.practicefusion.com/blog/ehr-adoption-rates/>
- [28] OECD Publishing. 2017. *Health at a Glance 2017*. OECD, Paris. [http://dx.doi.org/10.1787/health\\_glance-2017-en](http://dx.doi.org/10.1787/health_glance-2017-en)
- [29] Raghupahti Viju Raghupahti, Wullianallur. 2014. Big Data Analytics in Healthcare Promise and Potential. *Spinger Health Information Science and Systems* 2 (Feb. 2014), 2–3. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4341817/>
- [30] Rock Health. 2017. The Future of Personalized Healthcare: Predictive Analytics. Web page. (Nov. 2017). <https://rockhealth.com/reports/predictive-analytics/>
- [31] Search Technologies. 2017. Using Big Data Predictive Analytics to Improve Healthcare. Web page as article. (2017). <https://www.searchtechnologies.com/blog/predictive-analytics-in-healthcare>
- [32] Stephen B Thacker. 2003. HIPAA Privacy Rule and Public Health. *CDC* 52 (April 2003), 1–12. <https://www.cdc.gov/mmwr/preview/mmwrhtml/m2e411a1.htm>
- [33] The Common Wealth Fund. 2017. The Affordable Care Acts Payment and Delivery System reforms: A progress report. Web page as article. (Feb. 2017). <http://www.commonwealthfund.org/publications/issue-briefs/2015/may/aca-payment-and-delivery-system-reforms-at-5-years>
- [34] The datapine blog. 2017. Nine examples of Big Data Analytics in Healthcare that Can Save People. Web page. (May 2017). <https://www.datapine.com/blog/big-data-examples-in-healthcare/>
- [35] The Wall Street Journal. 2017. How to Research Medical Prices. Web page as article. (Nov. 2017). <http://guides.wsj.com/health/health-costs/>

- how-to-research-health-care-prices/
- [36] US Department of Health and Human Resources. 2017. EHR Basics. Web page. (Sept. 2017). <https://www.healthit.gov/providers-professionals/learn-ehr-basics>
  - [37] Wikipedia. 2017. Evidence Based Medicine. Web page. (Nov. 2017). [https://en.wikipedia.org/wiki/Evidence-based\\_medicine](https://en.wikipedia.org/wiki/Evidence-based_medicine)
  - [38] Wikipedia. 2017. Telemedicine. Web page. (Nov. 2017). <https://en.wikipedia.org/wiki/Telemedicine>
  - [39] Zane Benefits. 2017. FAQ - How much does Individual Insurance cost. Web page. (Nov. 2017). <https://www.zanebenefits.com/blog/bid/97380/faq-how-much-does-individual-health-insurance-cost>