

**Data Integration and Analysis in Healthcare: A Pivot Within the Industry**

by

**Mabeth Diaz Arcos**

BS, Creighton University, 2020

Submitted to the Graduate Faculty of the  
Department of Health Policy and Management  
School of Public Health in partial fulfillment  
of the requirements for the degree of  
Master of Health Administration

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SCHOOL OF PUBLIC HEALTH

This essay is submitted

by

**Mabeth Diaz Arcos**

on

April 21, 2022

and approved by

**Essay Advisor:** Kevin Broom, Ph.D., Associate Professor, Vice-Chair for Education, and Director of MHA and MHA/MBA Programs, Health Policy and Management, School of Public Health, University of Pittsburgh

Essay Reader: Jenna Carlson, Ph.D., Assistant Professor, Biostatistics, Assistant Professor, Human Genetics, Health Policy and Management, School of Public Health, University of Pittsburgh

Essay Reader: Jacob Collins, MS, UPMC, Department of Marketing Intelligence

Copyright © by Mabeth Diaz Arcos

2022

# **Data Integration and Analysis in Healthcare: A Pivot Within the Industry**

Mabeth Diaz Arcos, MHA

University of Pittsburgh, 2022

## **Abstract**

Since the HITECH Act in 2011, there has been a shift in electronic health records (HER) and data platforms that focus on providing access and better-quality care to patients and impacting healthcare systems. As advancements with data continue, an increase in statistical analysis and big data within the healthcare industry has occurred, which has led to a rise in needing analytical background employees to join the healthcare industry—coupled with a need for careful compliance and security of data to avoid significant data breaches that affect the organization and the patient. The current healthcare industry is changing, and it is essential that organizations stay at the forefront of these advancements, or they may suffer the consequences of being left behind.

This essay analyzes the current literature to understand how Big Data has impacted essential healthcare functions and how the COVID-19 pandemic revolutionized our healthcare industry. The COVID-19 Pandemic has pushed for massive adoption of telemedicine and advanced analytics in healthcare. The literature will highlight specific adoption examples throughout the country and even within the Pennsylvania (PA) market, reviewing analytical work conducted within the Department of Marketing intelligence at UPMC. Lastly, the literature will highlight how artificial intelligence (AI), predictive modeling, and powerful analytical tools impact the healthcare industry by providing greater access to patients, improving the quality of care, and creating potential cost savings. Holistically, the literature will give clear data integration and analysis examples that benefit healthcare entities and confirm that the healthcare industry is

changing. Organizations that lack sufficient data integration and analysis within their system may suffer consequences like suffering financial losses, low-quality reports, or even bankruptcy.

# Table of Contents

<b>1.0 Introduction.....</b>	<b>1</b>
<b>1.1 Big Data and Predictive Modeling in Healthcare .....</b>	<b>2</b>
<b>1.2 Process of Advanced Analytics.....</b>	<b>2</b>
<b>1.3 Benefits of Advanced Modeling in Healthcare .....</b>	<b>4</b>
<b>2.0 Data Impacting Healthcare.....</b>	<b>6</b>
<b>2.1 Utilizing Data Platforms for Strategic Efforts.....</b>	<b>6</b>
<b>2.2 Optimizing Data Sets.....</b>	<b>8</b>
<b>2.3 Advanced Analytics Improving Access .....</b>	<b>10</b>
<b>2.4 Advanced Analytics Improving Quality .....</b>	<b>12</b>
<b>2.5 Cost Savings through Advanced Analytics .....</b>	<b>14</b>
<b>2.6 Other ways Big Data and Predictive Modeling are Impacting Industries .....</b>	<b>15</b>
<b>3.0 Risks with Big Data and Predictive Modeling.....</b>	<b>17</b>
<b>3.1 Healthcare Data Breaches .....</b>	<b>17</b>
<b>3.2 Ethical Risks.....</b>	<b>18</b>
<b>4.0 Recommendations .....</b>	<b>20</b>
<b>4.1 Details of Recommendations .....</b>	<b>20</b>
<b>4.2 Detailed Benefits of Recommendation.....</b>	<b>24</b>
<b>4.3 Challenges with Recommendation .....</b>	<b>25</b>
<b>5.0 Conclusion .....</b>	<b>28</b>
<b>Appendix A UPMC Internal Cancer Competitive Analysis Example .....</b>	<b>29</b>
<b>Bibliography .....</b>	<b>30</b>

## List of Tables

<b>Appendix Table 1.....</b>	<b>29</b>
------------------------------	-----------

## List of Figures

<b>Figure 1. Three Ways to Reduce Implicit Bias in Predictive Analytics .....</b>	<b>23</b>
---	-----------



## **1.0 Introduction**

The literature review will focus on advanced analytics and big data in the healthcare industry as it impacts an organization's strategy by utilizing data to improve quality, cost, and accessibility. As data utilization grows, it is vital to comprehend how data have impacted healthcare organizations and review the varying risks of high data utilization. Therefore, the literature will analyze data breaches in the past while also reviewing best practices for what to do post-breach. Organizations must understand the future of healthcare concerning the amount of available data and the rapid movement of telemedicine utilization. COVID-19 has shifted telemedicine utilization; therefore, telemedicine data has emerged in which organizations can utilize AI and data analysis to optimize the data. To summarize, the literature will intertwine the points above in holistically analyzing the ways advanced analytics and data analysis have impacted organizations' ability to improve cost, increase access, and improve quality metrics.

In conclusion, the recommendation section will focus on the best strategies to meet such high demand for data utilization by recruiting analytical staff and having a strong compliance culture and knowledge. In addition, the recommendation will reiterate the importance of capital allocation to sustaining data warehouses, platform licenses, training, and other costs associated with having a robust analytical department. Healthcare organizations must stay ahead of their competitor by implementing such strategies or face falling behind in the analytical world. A weak analytical department within an organization could be a substantial roadblock for organizations in the long run; furthermore, organizations must comprehend this potential reality and adjust for the future.

## **1.1 Big Data and Predictive Modeling in Healthcare**

The literature section will review advanced and predictive analytics and its requirements while reviewing processes for conducting a solid analytical report. Advanced and predictive analytics deals with statistical methods, data mining, and game theory to analyze current and historical data collected at a medical establishment. These data models help improve patient care by analyzing; medical record information, age, social and economic characteristics, individual anatomy, and many other factors that reveal a patient's susceptibility to lifestyle-related conditions.<sup>i</sup> When building a predictive model, data analytics is used to find similar patterns in behavior and forecast people's responses or actions to occurring events. The end goal of using these practices is the optimization of decision-making, problem-solving, and identifying opportunities for improving a health system. The world's largest medical organizations are on their way to integrating advanced practices such as predictive analysis, simulation, and variable modeling. Specifically, in 2019, the Society of Actuaries (SOA) presented a report on predictive analytics in healthcare showing that 60% of respondents were already using predictive tools in their systems to improve key performance indicators (KPIs) in hospitals, clinics, and health insurance companies, and 20% had planned to implement predictive models within 2020.<sup>ii</sup>

## **1.2 Process of Advanced Analytics**

Advanced analytics is the use of different models to create forecasts or predictions to impact an organization's performance. Each respective model reviews a range of predictor variables or dimensions that affect future results, which gives organizations the ability to forecast

any variable within their business lines and strategically adjust. The specific models can use a simple linear regression or complex machine learning algorithms (e.g., neural networks) to perform advanced tasks to assess the different models to solve the defined problem. After the assessment, the model with the highest performance metrics (accuracy, etc.) is validated, tested, and rerun using an independent, real-world dataset. Predictive analytics and its main techniques, such as machine learning, data mining, and predictive modeling, stand at the next-to-last stage within both models. To fully utilize these models or practices, the necessary infrastructure must exist intertwined with a robust data analytics department that can sustain such tools. <sup>iii</sup>

Although there are differences in best practices for implementing advanced analytics within a department, research conducted reveals that, in general, following a process of acquiring the data, transforming data, then utilizing analytical applications for reports has been successful throughout the years. In detail, receiving the data refers to sourcing through either internal data via combing and centralizing platforms or externally by acquiring data sets from data mining companies. After acquiring the data, the next step is to transform the data by centralizing all data to a single data warehouse and utilizing different tools like HBase, JAQL, or Apache Hadoop to embed data into one platform. The data could include supply chain, digital imaging, infrastructure, outpatient/inpatient electronic medical records (EMR), claims, and consumer insight data. Lastly, after embedding data into a centralized data platform, analysts within the department can start utilizing applications to run queries, create reports, or data mine to conclude any trends. Moreover, the three-step process ensures the centralization and transparency of data, which provides an opportunity for organizations to utilize the platform or applications in the future.<sup>iv</sup>

Each section has subsections of components critical to the smooth adoption of the process. A report conducted by the Health Information and Management Systems Society (HIMSS) called

"HIMSS Analytics Adoption Model for Analytics Maturity (AMAM)" helps organizations address the critical components above. The implementation model introduces a roadmap for adopting analytics at healthcare organizations surrounding a seven-stage process. The process consists of standardizing vocabulary per data set, providing data transparency throughout the organization, and promoting a culture that surrounds itself by utilizing data to make better-informed decisions. Therefore, the first primary step holistically for an organization is to comprehend the process of advanced analytics and the best ways to adopt it within their respective organizations.<sup>v</sup>

### **1.3 Benefits of Advanced Modeling in Healthcare**

After reviewing the analytics modeling process and the various details surrounding data utilization and how to best implement it. The next step is to understand the benefits of such practices and how they impact both the patient and the organization by improving access, quality outcomes, and cost savings. As healthcare focuses on quality improvement and patient satisfaction scores through personalized medical care, predictive modeling can create specific models that enable personalized medical care and improve patient satisfaction scores. As the health experience becomes more holistic with integrated delivery systems and comprehension of the services being rendered per patient, predictive modeling can assist in comprehending and targeting the consumer. An organization needs to stay on top of the patient's experience as it deliberately continues to affect federal reimbursement rates like Medicare and Medicaid. Especially for an organization with a large population of Medicaid/ Medicare patients, utilizing these models to keep quality metrics high and consistent is necessary.

Research has indicated that the main benefits of predictive modeling in healthcare are improved diagnostics, high cost-effectiveness, enhanced operational efficiency, decreased readmission rates, and personalized medical care opportunities.<sup>vi</sup> Looking specifically at enhanced operational efficiency and high-cost effectiveness attributes to some of the most significant issues occurring within healthcare. Especially after the COVID-19 pandemic revealed systemic supply chain issues and a lack of cost-effective methods after many organizations had suffered losses during and even after the pandemic; as these issues continue to rise, healthcare administrators and organizations must comprehend how beneficial predictive modeling can best address the foreseeable issues. Moreover, the section below will evaluate how big data and advanced analytics comprehensively impact healthcare organizations by providing specific examples of utilizing data to improve a patient's experience or improve inefficiencies within the system.

## **2.0 Data Impacting Healthcare**

The literature below will elaborate on the benefits of predictive modeling and advanced analytics in healthcare as it explicitly impacts quality, access, and cost through data-driven strategies. This will be done by assessing the utilization of data platforms and optimizing data sets while also looking at specific examples of data positively impacting healthcare organizations' access, cost, and quality metrics.

### **2.1 Utilizing Data Platforms for Strategic Efforts**

Implementing data platforms has enabled healthcare organizations to review competition and reveal opportunities for growth, which indicates a shift within the industry. Studying various data sets containing claims data, market data, consumer insight data, and operational and financial data has given the ability for organizations to create data-driven strategic efforts, which have historically led to beneficial results. Various applications or platforms are accessible for organizations, like Microsoft Power BI, Tableau, STATA, and DOMO, allowing organizations to review large data sets in real-time.

For example, DOMO's business management platform allows for integrating data from any source, turns data into live visualizations, and extends BI into apps that empower an organization to make faster, better-informed decisions. Within the marketing intelligence department at UPMC, DOMO is the sole platform utilized which opens the opportunity to analyze clinical marketing, digital, operational, and financial data. A project conducted within the

department reviewed annual data reported by Hospital Data Insights (HDI) concerning Hospital US News Rankings. Every year, US News releases rankings for the best hospitals throughout the country broken down by 16 service lines and their respective rankings nationally. The rankings are a pervasive data process that mainly focuses on reputation and quality metrics; therefore, the project consisted of streamlining the data from the HDI database into DOMO. Specifically, it consisted of transferring the last five years of data to visualize and analyze outperformers or underperformers within the 16 service lines. Although this was not the first time the department has investigated the US News rankings, it is the first time the data would all be in one centralized area, promoting data governance and data transparency. Therefore, after a 3-month process, the data was transferred into DOMO, giving analysts, directors, and managers the ability to quickly review the data and create custom dashboards for service line leaders. For example, if the Chief of Heart and Vascular Institute at UPMC were to show up to the Marketing Department and ask how the cardiology ranking performance for the past two years has been? Analysts are then quickly able to create custom dashboards with DOMO and visualize the performance of cardiology nationwide within a two-year frame and report back to the Chief with any findings. These findings can report positive news revealing growth or negative information, which lays the path for an opportunity; nevertheless, real-time data allows the department to quickly analyze and adjust current strategic efforts to address findings within the report.

Lastly, another impactful way the business platform has impacted strategic efforts within the marketing department would be the creation of the “Hospital Performance Dashboard.” Directors, managers, and analysts worked together to create a Hospital Performance Dashboard that filters through UPMC hospitals throughout the PA region. The dashboard provides data transparency by enabling the Marketing Department to sift through the performance of each

hospital system and therefore strategize marketing efforts to improve an underperformer or initiate campaigns to address the futurist goals of UPMC. The dashboard consists of patient volume data, discharge data, consumer sentiment data, market share data, patient insurance breakdown, and surgery volumes. Having access to these performance metrics is critical to advancing the data as it strives to promote data-driven marketing campaigns in the long term.

## **2.2 Optimizing Data Sets**

Another critical aspect of utilizing data platforms and having a solid analytical department would be the comprehensive understanding of the benefits that come with optimizing data sets. Organizations nationwide have optimized single data sets to leverage themselves against competitors to gain market share or gain revenue. The most visible example of this would be Google; precisely, its search engine function, which amounts to trillions of data points coming from a single source in a day. Throughout the years, Google has optimized the data streamlined through its search engine functions to initiate strategic campaigns to increase its favorability or market share nationally. Within an array of data, there is room for creativity and innovation, which has led to utilizing a single source of data to comprehend a market entirety, from consumer favorability to futuristic demand for services. Google Analytics has continued to use data within its strategic efforts, which has led to substantial financial growth and market share favorability. According to Statista, Google has maintained 85.6% favorability within the past year, with Yahoo, Bing, Baidu, and Yandex RU below. In addition, within the past eight years, Google has never dropped below 81% favorability, reaching up to 94% in 2017. <sup>vii</sup>



To provide another example of the benefits of optimizing data set to utilize a single data source to reveal growth areas or comprehend a given market. The IQVIA referral analytic tool has been one of the most extensive projects conducted within the MI Department. Specifically, the IQVIA PI-360 referral analytical tool analyzes physician referrals across the country, given a data set with eight million rows of physician referral data. The first example of utilizing the referral analytical tool was when the MI team analyzed by leveraging the physicians' referrals and prescriptions using data backed by Medicare. The device allowed for analysis of prescription data reports and made correlations visible in finding a geographic concentration of patients based on specific patients identified with patient prescription consumption data. After a deep dive into the data and informative meetings, a report titled "The Use of Medicine in the U.S." concluded and publicized to UPMC executives and anyone interested in the data in June 2021. This type of referral analysis would not have been able to be conducted without the IQVIA affiliation with UPMC and their ability to provide data transparency with the tool and data set utilized.

Another example of utilizing the referral analytical tool in comprehending the Pennsylvania market would be the "Cancer Provider Referral Analysis." The "Cancer Provider Referral Analysis" custom dashboard was created after a cardiology service line leader wanted to look at the surgical oncology referrals being rendered or referred out of the UPMC network. Moreover, the dashboard synthesized the total referrals from in-network providers or out-of-network providers, revealing the number of referrals coming into UPMC facilities or being referred out. The dashboard even went a step further by showing the payer mix of the number of referrals and the specific insurance plans or groups associated with the referrals coming in or out of the UPMC network. Furthermore, the dashboard reported back to Marketing and Cardiology service line leaders, indicating areas of growth and strategizing marketing efforts to optimize the number

of referrals entering the UPMC network in the short term and long term. Overall, part of being a tremendous analytical organization is having complete comprehension of the organization's market; therefore, optimizing data sets can help organizations do just that.

The section has detailed optimizing data sets and how specific organizations have improved or opened room for innovation by utilizing data platforms to comprehend a market. It reveals how data has transitioned into every aspect of an organization's strategies, independent of its services or department. Moreover, the section below will go a step further by reviewing literature presenting specific ways that organizations have optimized data sets to create very innovative work that has been beneficial for organizations' access, quality, and cost metrics with predictive modeling or advanced analytics.

### **2.3 Advanced Analytics Improving Access**

The COVID-19 pandemic has quickly pushed the telemedicine movement forward, and healthcare organizations were forced into adopting this movement. Telemedicine nationwide was implemented promptly through the emergency policy framework, saving hundreds of patients. This is a fascinating time for healthcare institutions to join the telemedicine movement, especially as the industry's data utilization keeps increasing. Telemedicine has beneficially impacted patients' access to healthcare as it enables the opportunity to meet virtually and quickly make decisions. In addition, telemedicine has impacted another aspect of data for organizations to analyze, reviewing how patients prefer to render services and revealing the level of opportunity with telemedicine and primary care services. Data analytics firms like IBM, Oracle, and Cloudera have analyzed and quantified various metrics on how telemedicine has impacted a patient's experience by fully

comprehending the consumer.<sup>viii</sup> Research has revealed that 90% of healthcare leaders have implemented telemedicine technology, and 91% of consumers say that the health outcomes are as good or better via telehealth. Research specifically conducted at IBM reveals that 64% of healthcare consumers would be willing to render primary care services via telehealth. Furthermore, this shows that patients will continue to utilize these telehealth services. Healthcare organizations are already preparing to use these technologies for the long term, especially as telemedicine continues to improve health outcomes and the quality of care for patients.<sup>ix</sup>

Another aspect to comprehend is how telemedicine has impacted states' policy frameworks, especially as each state begins to vary in its telemedicine reimbursement rates. For example, Texas has recently passed a telehealth law affecting healthcare entities' reimbursing rates as of the last two months. This new law gives an outlook on how the future is moving and reveals an understanding of how there is variation in reimbursement nationally. The TexLa is a federally funded program designed to provide technical assistance and resources to new telehealth programs throughout Texas and Louisiana. The TexLa Telehealth Recourse Center (TRC) will continually evaluate telehealth programs in these two states and effectively deliver telehealth services, efficiency, sustainability, and patient satisfaction. The program will be funded by the U.S. Department of Health and Human Service's Health Resources and Services Administration (HRSA) Office for the Advancement of Telehealth, part of the Office of Rural Health Policy. They will ensure complete reimbursements for healthcare organizations that utilize telemedicine features for their services.<sup>x</sup>

Furthermore, this change is critical for understanding the advancement of telehealth in the future. Still, many equity issues can arise with telehealth expansion as some patients who live in rural areas do not have access to these features. Therefore, healthcare entities must have the

technical industry requirements and capital necessary to meet these needs. This requires having broad access to the internet through sufficient bandwidth to transmit audio and video and then having the correct devices and infrastructure to support those needs. A critical component to sustaining telemedicine usage in the long term is continually training staff and setting goals to upkeep the devices to ensure that the organizations provide the best care possible. Furthermore, the telehealth movement is expensive and requires high maintenance but also is the ideal broker for an organization's continual success or survival in the future.<sup>xi</sup>

Lastly, big data and predictive modeling can positively impact public health and communities by analyzing areas or community disease patterns and tracking disease outbreaks to improve surveillance and response speed. Turning large data sets into actionable information can identify a community's needs and predict a crisis, all of which benefit the population within the community. This impacts access, as organizations can utilize data to strategically improve the level of outreach their organization has and the level of access they can give to their community.

## **2.4 Advanced Analytics Improving Quality**

Advanced analytics can impact an organization's quality outcome with a better patient experience. For example, Columbia University Medical Center analyzed correlations of physiological data streams related to patients with brain injuries. Columbia utilized the entirety of the data collected throughout the years to optimize data reports best and adhered to proper predictive modeling standards. The goal was to provide medical professionals with critical and timely information to treat complications aggressively. The initiative did just that, as severe complications were diagnosed as much as 48 hours sooner in patients who suffered a bleeding

stroke from a ruptured brain aneurysm, giving administrators and physicians real-time information to best treat the patients and reduce certain health risks with such care.<sup>xii</sup> To add, one of the most prominent cases of using predictive analytics was Parkland Hospital System. In 2009, Parkland Hospital began applying models forecasting admission risks for patients with diagnosed diabetes, pneumonia, and heart attack. The models incorporated the entirety of the electronic medical record (EMR) database. The goal was to decrease total readmission rates for patients with congestive heart failure within the Parkland system. Moreover, within a three-year time frame, the organizations reported lower readmission rates and better-quality reports. This example also shows how early institutions started implementing predictive modeling in 2009 and how advanced modeling can improve a patient's experience.<sup>xiii</sup>

Lastly, looking specifically at clinical trials, predictive modeling can improve clinical trial design and find better matches with treatment and patient, speeding up the recruitment process and reducing trial failures through statistical tools and algorithms. Within clinical operations, models can be conducted to compare effective research and find appropriate, cost-effective ways to diagnose and treat patients. Models can analyze patient records to identify adverse effects of a product before it even reaches the market but, most importantly, can comprehend best practices for designing trials. With efficient designing and targeting of patients for clinical trials, the models can improve a patient's experience by achieving successful rates for the clinical trials and avoiding re-trials.

## 2.5 Cost Savings through Advanced Analytics

Another area of focus throughout the years has been the extreme amount of waste and continuing rise in costs within the medical industry. Specifically, McKinsey's consulting firm estimates that predictive modeling and big data can enable more than \$300 billion in savings per year in U.S. healthcare, approximately 8 % of national healthcare expenditures. In addition, McKinsey believes big data could lead to other potential savings in research and development (R&D) and clinical operations, with an estimate of \$165 billion in clinical operations and \$108 billion in R&D. To be specific, predictive modeling can produce a more targeted R & D pipeline in drugs and devices with research and development and run logistical models to find an efficient design for trials and reduce trials costs.<sup>xiv</sup>

To provide another example of how big data has impacted a large healthcare organization that operates nationally. Premier, a U.S. healthcare alliance network, which has more than 2,700 members, 90,000 hospital systems, and 400,000 physicians, has utilized the data within the system to provide advanced analytics reports. In detail, Premier has access to one in four patients discharged from any of their 90,000 hospital facilities, including clinical, financial, patient, and supply chain data. Because of Premier's extensive array of data, the network assembled a database that generated comprehensive and comparable clinical outcome reports, resource utilization reports, and transaction-level cost data. These outputs have informed decision-making and improved healthcare processes at approximately 330 hospitals, saving an estimated 29,000 lives and reducing healthcare spending by nearly \$7 billion.<sup>xv</sup> Therefore, this example strongly implies the varying benefits of big data and modeling. It impacts mortality rates nationally and significantly reduces healthcare spending, which is positive as healthcare spending and costs have increased yearly.

## 2.6 Other ways Big Data and Predictive Modeling are Impacting Industries

This part of the literature will highlight how big data and predictive modeling have impacted industries other than healthcare. Research conducted in 2022 reveals that healthcare lies 4th concerning advancement in predictive modeling and advanced analytics as Business Intelligence, Banking, and Supply Chain are among the top industries.<sup>xvi</sup> An industry of consistent modeling and predictive analytics remains in the finance and banking world. As research indicates, banking has remained the top hiring industry for analysts for the past ten years. Although the most visible form of modeling occurs with stock prediction and security investments, there are other forms that the finance industry utilizes advanced analytics to promote financial success and create lucrative cost estimations. For example, the FICO credit score uses statistical analysis to predict a consumer's behavior, such as how likely the consumer will miss payments. A consumer's score is based on how borrowers like the consumer have performed in the past. Therefore, banks utilizing credit scoring models use data to predict a person's creditworthiness. Lastly, another area of focus with modeling within the finance industry would be the implementation of analysts running complex SQL queries that investigators can investigate and use the model to detect fraud. Fraud is a massive issue within the finance industry; therefore, utilizing fraud detection models will increase a bank's ability to detect fraud and eliminate it immediately.

Another clear example of the level of analytical advancement that some industries have reached would be within the supply chain industry. Consumers become more food conscious by learning about specific food sources and processes. Specifically, with food safety assurance and sustainability, there has been scrutiny over the handling of foods over the years. Therefore, food companies have revamped their supply chain process by allowing advanced analytics in their strategic efforts. For instance, Loeul and Pirirot, a food delivery company specializing in the

processing of rabbit meat, aimed to track the efficiency of its supply chain from breeding to delivering the meat. The company uses two different systems to identify inefficiencies best; the first system utilizes flow modes reviewing data from the time orders were received to preparation, logistics, and product delivery. The other system solely focused on the purchasing process and examining the number of orders entering the system and delivering. Moreover, analysts within the company were able to bridge both systems and build a model to trace the entire journey of the livestock within its supply chain. The strategy proved to impact efficiency as there was a decrease in repeated and lost orders, creating cost savings. In addition, big data analytics helps plan for what-if scenarios within the supply chain world. It clarifies specific operational or marketing strategies that organizations can execute to address supply and demand concerns.<sup>xvii</sup>



### **3.0 Risks with Big Data and Predictive Modeling**

Throughout a healthcare administrator's career, it will be necessary to comprehend the legal implications of each action taken fully in any field or department. In addition, it is vital to an organization's success in meeting legal policies and meeting standards for U.S. compliance regulations. Therefore, this section of the literature will address and discuss risks that healthcare entities currently face within the healthcare industry, especially as big data and predictive modeling enter the field quickly.

#### **3.1 Healthcare Data Breaches**

The first risk implication that has continually increased throughout the years, especially with the mass movement of big data into the healthcare industry, would be healthcare data breaches. The legal compositions that intertwine the implications in 2020, entities reported healthcare data breaches of 500 or more records at more than 1.76 per day. In 2020, there were 642 breaches (a 25% increase from 2019) in data breaches reported by healthcare providers, health plans, healthcare clearinghouses, and business associates.<sup>xviii</sup> Healthcare data breaches continue to tarnish organizational financial stability and reputation, especially as the cost for healthcare data breaches rises to an average of \$9.3 million per healthcare data breach. To provide a specific example, the most significant healthcare data breach of 2020 was a ransomware attack on the cloud service provider Blackbaud Inc. Although the exact number of records exposed is unclear, it was reported that around 10 million records of 100 of Blackbaud's healthcare clients were affected.

The specific data included: Social Security numbers, driver's license numbers, health insurance, and financial information, which created massive financial repercussions for Blackbaud and a lack of trust from clients.

The causes of healthcare data breaches in 2020 and the central methodologies for hackers to abuse information systems at healthcare organizations vary. Hacking and other I.T. incidents dominated the healthcare data breach reports in 2020; 429 breaches were associated with this methodology which accounts for 66.82% of reported breaches. Specifically, these incidents included exploiting vulnerabilities and phishing, malware, and ransomware attacks, all of which impacted organization stability and reputation.<sup>xix</sup> Moreover, electronic data breaches have become a significant problem in all industries but can be particularly devastating to the healthcare industry due to the sensitivity of healthcare data. In response, nearly all states have developed laws to combat data breaches. Laws vary among states regarding the definition of personal information, how notice of violations must occur, and the number of fines for organizations found non-compliant with data breach laws. In addition, organizations have begun to implement data security strategies to prevent these issues from occurring, which will be addressed in the recommendation section of the literature.<sup>xx</sup>

### **3.2 Ethical Risks**

Big data research has also raised new ethical concerns despite the many gains. Big data research requires access to a large-scale amount of personal information and the relative tendency of analytics programs to reflect a human error. As healthcare organizations become reliant on predictive analytics to make decisions, these tools must be free of implicit or explicit bias that

could further drive health inequities. Implicit bias reflects explicitly the human tendency to use pre-existing knowledge of patterns to characterize new information, even though those patterns are rooted in social bias or cultural perceptions of an individual or community. Moreover, AI and predictive analytics tools are subject to the same kind of implicit biases, especially as configuring patterns is one of the core functionalities within predictive analytics. Therefore, developing flawless, bias-free algorithms is often challenging with the existing disparities that pervade the healthcare industry. In a 2019 study, researchers from the University of California Berkeley discovered racial bias in a predictive analytics platform referring high-risk patients to care management programs. After the discovery, Ziad Obermeyer, associate professor of health policy and management at UC Berkeley, stated, "We make so many choices when we train an algorithm that feels technical and small. However, these choices make the difference between an algorithm that's good or bad, biased, or unbiased."<sup>xxi</sup>

Moreover, healthcare organizations must comprehend the risk level and create strategies to best reduce biases within a given model. Ethical risks can also create financial risks through HIPAA compliance violations or privacy violations, creating legal suits that tarnish an organization's reputational and financial stability. Another matter that has consistently surrounded the ethical realm and big data would be privacy and the consumer, especially as there is a variation in each consumer's acceptability of data. Furthermore, as big data and analytics continue within the healthcare industry, organizations must adhere to proper privacy and HIPAA regulations and culturally comprehend that the analytics best improves the patients' experience and creates growth opportunities.

## **4.0 Recommendations**

The literature will now expand on recommendations touched on slightly throughout; specifically, discuss recommendations for data compliance, recruitment of analytical staff, and methods to stay ahead of the competitor. Historically, functionalities or adaptations have been implemented that have been a substantial roadblock for organizations, like the utilization of the internet or, even recently, the adoption of e-commerce platforms to supply goods in retail. In retail, an organization's sticking point had a robust e-commerce platform to sustain the advancements in the industry, especially as the pandemic led to the increase in the utilization of such services. With demand significantly higher than usual, if the retail companies did not have the necessary capital and infrastructure to meet needs, they would have substantial supply chain issues. There were clear winners (Costco, Walmart, Amazon) and clear losers (JC Penny, Dollar General, Pier 1 Imports). This provides an example of how industries face defaulting with these insurmountable issues, which healthcare now faces the same potential issue with advanced analytics. The detailed recommendation section below will best address strategies for sustaining an analytics department and reveal methodologies to stay ahead of the industry and be one of those potential winners in the future.

### **4.1 Details of Recommendations**

As organizations must comply with the HIPAA policies and regulations, each organization must have the understanding and methods to follow such regulations properly. Therefore, the first

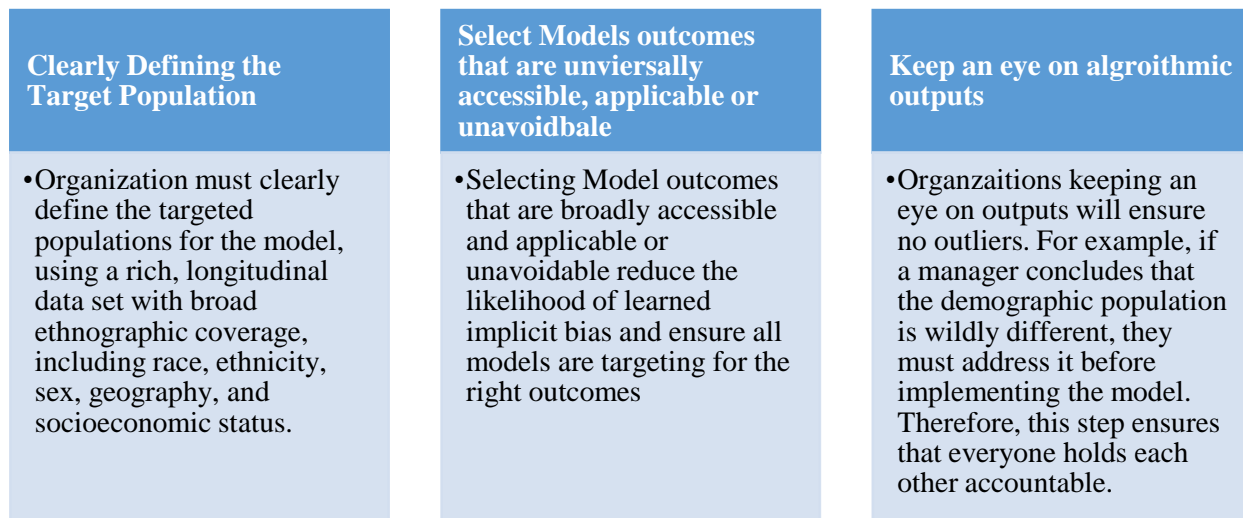
recommendation focuses on having a data compliance department. Every healthcare organization must have a compliance committee or department, from a small organization with one clinic to an integrated healthcare system with hundreds of physicians and facilities. Healthcare organizations must designate a specific compliance officer with at least a team or a committee working below to ensure the department's success and transparency. Compliance officers can use the committee to train and educate the organization on data security and build a culture that emphasizes this importance. Everyone must comprehend how data breaches can occur and the measures taken once they appear. Effective communication with everyone on the team is critical to the success of the process.

Although there are various methods to properly handle a data breach and steps to take after the breach occurs, the EU General Protection Regulation (GDPR) recently reported a "best method" that starts with the moment a personal data breach is identified. The first task is to inform the head of the compliance or compliance office at the organization. The second step consists of assessing the scope and impact of the breach by analyzing the personal data breached, the number of subjects, the type of data, and most importantly, notifying any third parties of the breach. The third step consists of a deep dive into the data and establishing security within a 72-hour frame to ensure no more violations in the short term. The final step would be once the personal data breach has been contained, the organization should review existing measures in place and explore the possible ways to prevent a similar breach from reoccurring.<sup>xxii</sup>

The second recommendation surrounds itself with the importance of recruiting analytical staff to the healthcare industry. The continuing recruitment of people with robust data security compliance knowledge and appreciation will ensure compliance with strong data regulations in the future. Healthcare industries have continued to recruit potential employees from other sectors

like banking, cloud computing, and tech firms to bring employees who have successfully dealt with large data sets and protected them. Overall, it is vital for employees on every level of the organization to comprehend these steps that need to be taken. It is critical to avoid more breaches in the future. Therefore, having a large scale of analysts or data-driven employees will allow for a cultural growth of data appreciation and comprehension of HIPAA regulations. In addition, hiring analysts will also promote the innovative and creative side of the organization, which would lead to benefits in costs, patient experience, and access improvement, as confirmed during the examples of advanced analytics above.

To add on, through the hiring of analytical staff, the staff could go through extensive training to review best practices for modeling and create bias-free models, which would directly impact the associated ethical risk mentioned. Analysts can work with experts and end-users to understand what clinical measures are essential to providers to remove bias from big data analytics tools. Although the communicative strategy improves algorithm development, this step is only half the battle when trying to create a bias-free algorithm. Collecting the high-quality data needed to develop unbiased analytics tools is time-consuming and difficult. Researchers at Columbia University have developed a learning module that focuses on correcting gender bias by building balanced data subsets of patients with equally distributed gender amounts to accelerate this process. Every researcher was required to do a module before any model or algorithm was created to ensure no biases in the model or reach the best potential as a group. Recent reports by the European Union on adhering to artificial intelligence provide a framework for organizations' best strategies to address the ethical issue. The report details three main steps organizations can take in creating bias-free algorithms; the figure below highlights the steps and benefits of developing models.<sup>xxiii</sup>



**Figure 1. Three Ways to Reduce Implicit Bias in Predictive Analytics**

The main point from the above is that through the recruitment of analytical staff and substantial training on best practices, an organization can address ethical and operational issues and utilize analytical tools without worry.

The last recommendation will be expanded in the next section but focuses on the importance of healthcare organizations staying competitive within the industry by consistently allocating resources to data via the implementation of technological devices, data platforms, and the consistent capital allocation to the upkeep of data warehouses and infrastructure. Research conducted by Forbes reveals that the cost of big data continues to rise as demand for services and data availability continues to increase throughout the years. Specifically, the prices are rising due to high licensure fees for data platforms like DOMO and Microsoft Power BI. The consistent maintenance, upkeep, and expansion of data warehouses are significant expenses for healthcare organizations as demand for data within the industry expands. Healthcare organizations need to

strategically prepare for these expenses by investing in proper infrastructure and adhering to federal policy and regulations while allocating available cash flow to the upkeep of data infrastructure and setting funds aside for expected costs in training employees. Without investment, organizations could suffer from budgetary constraints and lose analytical staff, creating a substantial roadblock in developing an analytical department.

## **4.2 Detailed Benefits of Recommendation**

With the efficient and strategic implementation of the recommendation mentioned above, healthcare organizations can stay ahead of their competitors and, most importantly, impact communities with population health improvements. By digitizing, combining big data, and adhering to compliance regulations, healthcare organizations varying from small to large are set to realize significant benefits. Potential benefits include finding disease patterns within a population, creating individualized care models, detecting healthcare fraud, and cost-effectiveness. Analysts could use historical data to predict and estimate the length of stay (LOS) for patients who will need surgery, patients who will benefit from surgery, complications, readmission rates, 30-day mortality, and even chances of a patient acquiring a hospital-acquired illness. In a sense, the sky is the limit with the analysis that can be conducted long-term as the availability and utilization of data increases worldwide. This is all the disposal of following the correct strategy, by implementing the data infrastructure and especially training staff on modeling standards. In healthcare, it is understood that the success of big data analytics tools depends on the value of the information used to train them. In addition, another critical aspect that comes with hiring analytical staff to conduct extensive advanced analytics is the opportunity to be a more cost-effective



organization. Through cost-effective models, an organization can double its cost-effectiveness within each department. For example, due to data-driven strategic efforts and analytics targeting consumers for a marketing campaign, the UPMC marketing department has doubled the effectiveness of its marketing spend, especially when looking at cost per acquisition (CPA) metrics.<sup>xxiv</sup>

Lastly, the healthcare industry utilizes evidence-based medicine and shifts its focus on the patient's experience; creating personal or individual care structures is another foreseeable positive impact on individuals. Advanced data analytics could help pinpoint patients who are the highest utilizing consumer of healthcare services or patients with the most significant risk for adverse outcomes. Therefore, providing individuals with the information, they need to make better-informed health decisions and more efficiently manage their health by tracking health behaviors. Patients will have access to data that will enable them to review treatment plans and their respective costs and benefits; specifically, EMR data examining environmental factors that increase trigger adverse health effects to decrease readmission rates. In addition, monitoring health in real-time via apple watches or vital monitoring devices allows patients to make data-driven health decisions. Applying advanced analytics to patients' profiles will efficiently identify individuals who would benefit from different care models like proactive, holistic, or preventive care.

### **4.3 Challenges with Recommendation**

This section will discuss some of the general challenges complicating the recommendations mentioned. Some industries like banking and insurance have recruited most of the talent pool for intelligent analysts or analytically minded people. As analysts continue to be high in demand

within the job market, industries such as banking can be a more lucrative field, and compensation packages differ significantly. As competition continues amongst industries, healthcare falls behind in recruitment compared to some more lucrative fields. On top of that, the higher salaries that analysts demand make it difficult for healthcare industries to compete in the market. Therefore, organizations must make tough decisions about whether an analyst can fit into the budget and if there is a level of ROI per employee.

Another evident challenge would be the extreme amount of capital needed to sustain data warehouses, data licenses, and other aspects of maintaining a solid analytical department. A challenge for every organization, but primarily small organizations with low cash on hand or poor financial stability, would be the required expenses to sustain data infrastructure. With low levels of liquidity, small entities might find it hard to invest in the data platforms licenses per employee or even a data warehouse. Moreover, a lack of investment could lead to a lack of innovation and being left behind in the process, which in the long run could lead to one of those insurmountable issue situations. On a similar path, large-scale organizations will continue to struggle with controlling costs with data, especially as companies like UPMC have thousands of employees paying for thousands of licenses to utilize data services. As prices continue to rise, both small and large organizations must make tough financial decisions to maintain competitiveness within a market.

Lastly, another challenge that arises with advanced analytics and data sets is understanding that there is still some unclarity on policy framework in the future as it impacts healthcare organization utilization of such services. The policy framework reveals a hurdle for such modeling advancements, as the lack of clarity can hinder industry growth and comprehension. Holistically organizations need to evaluate the various challenges with the recommendations and their

associated risks as they can impair an analytical department's growth. In addition, as noted in the risks section, there are ethical, operational, financial, and clinical risks associated with advanced analytics; furthermore, organizations must review the risks and ensure that they never become real situations for their organizations.<sup>xxv</sup>

## 5.0 Conclusion

Healthcare organizations must continue to allocate resources to recruiting analytical employees and begin to make advancements in the analytics or telehealth department as it continues to grow within the industry. As data breaches continue to rise, organizations must create robust data compliance plans and push for a culture and environment that cares about the importance of data security. As this movement continues to expand nationwide, it is critical for healthcare organizations, insurance groups, and more to focus on being ahead of the game in this aspect; otherwise, it could be an impasse for organizations. As discussed earlier, some industries like retail have revealed substantial roadblocks like e-commerce which have either led to considerable growth for an organization or led to bankruptcy and default. In the future, the lack of advanced analytics and big data could be one of those substantial roadblocks for any healthcare entity. Big data analytics can transform how healthcare organizations use sophisticated technologies to gain insight from data and improve their cost, quality, and access metrics. We will continue to see the widespread implementation and use of advanced analytics in the healthcare industry worldwide, which means healthcare organizations must stay ahead of the industry or be prepared to be one of those entities left behind.

## Appendix A UPMC Internal Cancer Competitive Analysis Example

**Appendix Table 1.**

Program	Locations	Clinical Trials	# of Spores	Research Programs & Spend	Faculty Size	Annual Unique Patients	Annual IP	Annual OP	NCI Ranking	Diagnostic and Treatment	Patient Support
UPMC	73- WPA & Central PA 3- MD, OH, NY 4 International- Ireland & Italy	450+	1-Head & Neck	7 Research Programs \$150 M	257	137,853	7,090	507,864 Provider Visit	Comprehensive	Exceptional	Exceptional
MD Anderson	10 (Houston) 15 Certified locations in 12 states 5 Partner Members 2 Associate Members (International) Affiliate members in Turkey and New Mexico	1,400	9 - Brain, Endometrial Cancer, GI Cancer, Hepatocellular Carcinoma, The Joe Moakley Leukemia, Lung Cancer, Melanoma, Ovarian, Prostate Cancer	37 Research Centers & Programs 1,364 active clinical research \$903.3 million	1,711	148,700	30,339	1.5 mil	Comprehensive	Exceptional	Exceptional
Memorial Sloan	25- 22 NY City and 3 in NJ	605	5- Bladder, Breast, Lymphoma, Prostate, Sarcoma	1,159 clinical protocols for pediatrics and adults 8 research programs & centers \$195,694,578	1,358	NA	25,597	839,073	Comprehensive	Exceptional	Exceptional
Mayo Clinic	14 total locations in Minnesota/Wisconsin, AZ and FL campuses	856	6 in MN- Breast Cancer, Hepatobiliary Cancer, Lymphoma, Multiple myeloma, Ovarian Cancer, Pancreatic Cancer 1 in AZ- Myeloma	10 Research Programs	216	150,000	N/A	N/A	Comprehensive	Exceptional	Exceptional

Rating	Diagnostic and Treatment	Patient Support
● Basic	Offers traditional diagnostics (X-Ray, CT, MRI, PET) and treatments (chemotherapy, radiation therapy including IMRT and IGRT, immunotherapy, surgery)	Provides very limited support for patients and families
●● Strong	Offers advanced services such as robotic surgery and tomotherapy but information on website is limited and does not clearly establish a cutting edge program	Offers a range of support services including nutrition, palliative and spiritual care, support groups, and social worker services
●●● Exceptional	Offers advanced services in chemotherapy, radiation therapy and surgery including newest techniques (proton therapy, targeted therapy, individualized gene-based treatment) and equipment (RapidArc™, TrueBeam STx®, TomoTherapy®)	Offers a wide array of services that support patient and family in a robust and clearly promoted program that targets every audience (parents of children with cancer, children whose parents have cancer, young adults, survivors, etc.); large volunteer network and multiple delivery methods (online, in person, via phone, etc.)

## Bibliography

---

- <sup>i</sup> Petrova, Bilyana. “Predictive Analytics in Healthcare.” *Reveal Embedded Analytics*, 13 Dec. 2021, <https://www.revealbi.io/blog/predictive-analytics-in-healthcare>.
- <sup>ii</sup> Petrova, Bilyana. “Predictive Analytics in Healthcare.” *Reveal Embedded Analytics*, 13 Dec. 2021, <https://www.revealbi.io/blog/predictive-analytics-in-healthcare>.
- <sup>iii</sup> Shugalo, Inga. “Predictive Modeling in Healthcare: How to Adopt It Painlessly.” *Itransition*, 8 Jan. 2020, <https://www.itransition.com/blog/predictive-modeling-in-healthcare>.
- <sup>iv</sup> Shugalo, Inga. “Predictive Modeling in Healthcare: How to Adopt It Painlessly.” *Itransition*, 8 Jan. 2020, <https://www.itransition.com/blog/predictive-modeling-in-healthcare>.
- <sup>v</sup> “Celebrating HIMSS Stage 7 Organizations Advancing Global Health.” *HIMSS*, HIMSS, 16 July 2020, <https://www.himss.org/news/celebrating-himss-stage-7-organizations-advancing-global-health#:~:text=The%20HIMSS%20Adoption%20Model%20for%20Analytics%20Maturity%20%28AMA%29,gained%20from%20installation%20of%20technology%20and%20surrounding%20processes>.
- <sup>vi</sup> Petrova, Bilyana. “Predictive Analytics in Healthcare.” *Reveal Embedded Analytics*, 13 Dec. 2021, <https://www.revealbi.io/blog/predictive-analytics-in-healthcare>.
- <sup>vii</sup> Johnson, Joseph. “Google Market Share Countries 2021.” Statista, 1 Mar. 2022, <https://www.statista.com/statistics/220534/googles-share-of-search-market-in-selected-countries/>.
- <sup>viii</sup> McManus, Eric. “Medicaid Telemedicine and Telehealth Services - TexLa.” *Texas Health and Human Services*, Telehealth Resource Center, 2021, <https://texlatrc.org/wp-content/uploads/2021/03/Texas-Medicaid-Update-slides.pdf>.
- <sup>ix</sup> “Impact of Telemedicine [Infographic].” Adroitent, <https://adroitent.com/blog/impact-of-telemedicine-infographic/>.
- <sup>x</sup> “About- TexLa.” *TexLa*, Telehealth Resource Center, 2022, <https://texlatrc.org/about/>.
- <sup>xi</sup> “Telemedicine Emergency Rule.” *Tdi.texas.gov*, Texas Department of Insurance, 14 June 2021, <https://www.tdi.texas.gov/news/2020/telemedicine-emergency-rule.html>.
- <sup>xii</sup> Bresnick, Jennifer. “10 High-Value Use Cases for Predictive Analytics in Healthcare.” *HealthITAnalytics*, 9 Aug. 2019, <https://healthitanalytics.com/news/10-high-value-use-cases-for-predictive-analytics-in-healthcare>.

- 
- <sup>xiii</sup> Petrova, Bilyana. “Predictive Analytics in Healthcare.” *Reveal Embedded Analytics*, 13 Dec. 2021, <https://www.revealbi.io/blog/predictive-analytics-in-healthcare>.
- <sup>xiv</sup> Balaji, Arjun, et al. “How Predictive Analytics Can Boost Product Development.” McKinsey & Company, McKinsey& Company, 13 Feb. 2019, <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/how-predictive-analytics-can-boost-product-development>.
- <sup>xv</sup> Bresnick, Jennifer. “10 High-Value Use Cases for Predictive Analytics in Healthcare.” *HealthITAnalytics*, 9 Aug. 2019, <https://healthitanalytics.com/news/10-high-value-use-cases-for-predictive-analytics-in-healthcare>.
- <sup>xvi</sup> Meltzer, Rachel. “The Top Industries Hiring Data Analysts Right Now in 2022.” *CareerFoundry*, 27 Dec. 2021, <https://careerfoundry.com/en/blog/data-analytics/top-industries-hiring-data-professionals/>.
- <sup>xvii</sup> Bresnick, Jennifer. “10 High-Value Use Cases for Predictive Analytics in Healthcare.” *HealthITAnalytics*, 9 Aug. 2019, <https://healthitanalytics.com/news/10-high-value-use-cases-for-predictive-analytics-in-healthcare>.
- <sup>xviii</sup> Mitchell, Hannah. “Healthcare Data Breaches up 55.1% in 2020, Report Finds.” *Becker's Hospital Review*, Becker Health, 22 Feb. 2020, <https://www.beckershospitalreview.com/cybersecurity/healthcare-data-breaches-up-55-1-in-2020-report-finds.html>.
- <sup>xix</sup> HIPAA. “2020 Healthcare Data Breach Report: 25% Increase in Breaches in 2020.” *HIPAA Journal*, 3 Mar. 2021, <https://www.hipaajournal.com/2020-healthcare-data-breach-report-us/>.
- <sup>xx</sup> Bowen, Mark. “A Cybersecurity Diagnosis for the Healthcare Sector with Breach-Likelihood.” *Intelligent CISO*, 4 Jan. 2022, <https://www.intelligentciso.com/2021/12/13/a-cybersecurity-diagnosis-for-the-healthcare-sector-with-breach-likelihood/#> .
- <sup>xxi</sup> Wade, Derick. “Ethics of Collecting and Using Healthcare Data.” *BMJ (Clinical Research Ed.)*, BMJ Publishing Group Ltd., 30 June 2007, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1906611/>.
- <sup>xxii</sup> Watt, Kirk. “Data Breach Handling Procedure.” *AylinWhite*, 2020, <https://aylinwhite.com/index.php/data-breach-handling-procedure/>.
- <sup>xxiii</sup> Palmer, Mark. “Seven Steps to Help You Reduce Bias in Algorithms in Light of the EU's Trustworthy AI Blueprint.” *Medium*, Towards Data Science, 29 Sept. 2021, <https://towardsdatascience.com/seven-steps-to-help-you-reduce-bias-in-algorithms-in-light-of-the-eus-trustworthy-ai-blueprint-b348dc3cf2ae>.
- <sup>xxiv</sup> Bresnick, Jennifer. “10 High-Value Use Cases for Predictive Analytics in Healthcare.” *HealthITAnalytics*, 9 Aug. 2019, <https://healthitanalytics.com/news/10-high-value-use-cases-for-predictive-analytics-in-healthcare>.

---

<sup>xxv</sup> Savitz, Eric. "The Big Cost of Big Data." *Forbes*, Forbes Magazine, 17 Apr. 2012,  
<https://www.forbes.com/sites/ciocentral/2012/04/16/the-big-cost-of-big-data/?sh=7c3edb35a3b7>.