**Zenas T. Njokom**

**University of Maryland, Baltimore County**

**Master of Professional Studies in Health Information Technology**

**HIT – 760**

**Health Informatics Capstone**

**John Paul Mulhern**

**Exploratory Analysis of Security Threats and Breaches Associated with unsecured Protected Health Information (PHI) during the COVID-19 pandemic**

**Summer 2021.**

**Table of Contents**

Abstract…………………………………………………………………………………...............2

[Introduction 3](#_Toc79052641)

[Background of Study 3](#_Toc79052642)

[*HIPAA Privacy Rule* 5](#_Toc79052643)

[*HIPAA Security Rule* 5](#_Toc79052644)

[Scope of the Project 8](#_Toc79052645)

[*Severity of the Problem Under Investigation* 8](#_Toc79052646)

[*Importance of the Topic and Researcher’s Point of Interest.* 9](#_Toc79052647)

[Stakeholders 10](#_Toc79052648)

[*Potential Causes of the Problem under Investigation* 11](#_Toc79052649)

[*Perverse business incentives* 11](#_Toc79052650)

[*Resource deficit* 11](#_Toc79052651)

[*Policy deficit or lack of government regulation* 12](#_Toc79052652)

[*Public Awareness* 13](#_Toc79052653)

[*Current Trajectory* 14](#_Toc79052654)

[Methodology 15](#_Toc79052655)

[*Theoretical framework* 17](#_Toc79052656)

[*Data/Evidence* 18](#_Toc79052657)

Results……………………………………………………………………………………………………19

[*Total Number of breaches under investigation* ....19](#_Toc79052658)

[*Analyzing Breach Location* 20](#_Toc79052659)

[*Analyzing the Number of Individuals Affected Over Time* 21](#_Toc79052660)

[*Investigating Types of Covered Entities Most Affected by Data Breach.* 22](#_Toc79052661)

[*Number of Business Associates Affected* 23](#_Toc79052662)

[*T-test at 95% Confidence Level* 24](#_Toc79052663)

[Discussion 25](#_Toc79052664)

[Conclusion 29](#_Toc79052665)

[Bibliography 32](#_Toc79052666)

[Appendix 37](#_Toc79052667)

**Abstract**

The rising influence of digital technology, especially the use of electronic health records (EHR), has created a huge potential to improve clinical outcomes and transform care delivery. However, there are increasing concerns relating to the security of healthcare data and devices. Increased connectivity and interoperability of electronic health records has exposed protected health information (PHI) to new cybersecurity vulnerabilities as witnessed during the COVID-19 pandemic. Healthcare is an attractive target for data breaches and cybercrime for two fundamental reasons: it is a rich source of valuable data, and it has weak security enforcement protocols, which center mostly around the Health Insurance Portability and Accountability Act (HIPAA), and the Health Information Technology for Economic and Clinical Health Act (HITECH). Security breaches include stealing health information and ransomware attacks on healthcare organizations, and this can reduce patient trust, destroy the integrity of health organizations, and negatively impact human life. This project analyzed major security breaches associated with unsecured protected health information faced by healthcare organizations in the United States during the COVID-19 pandemic. According to the analysis, healthcare data breaches are on a progressive rise with an average of 2,247,842 breaches reported per month, which means roughly each month, 250,000 patients have their health information compromised, from November 2019 through June 2021. Hacking and IT incidents are the top breach types as attackers sort to locate electronic health data for moratory motives. Securing electronic data and cybersecurity are critical to patient safety and must be considered an integral part of patient safety and the organizational structure. Adopting a security framework designed for healthcare would assist healthcare organizations to achieve a firm and adequate position on security. Healthcare organizations need to look at cybersecurity beyond the context of compliance on its own, but rather as an integral part of HIPAA and HITECH. This holistic approach to privacy and security ensures that all statutory obligations are met, as well as security measures and policies are implemented.

# **Introduction**

A reality for many organizations today is the challenge of ensuring proper security and ethical utility of enormous amounts of data. For most of these organizations, the data under their control is continuously accumulating, sometimes at an increasingly accelerated rate. The healthcare industry is highly impacted given that it deals with crucial and important personal (patient) information. A key challenge is healthcare data management is grappling with the increasing use of electronic medical devices that generate digital data that is stored in disparate formats and physical locations. Healthcare data, which often contains protected health information (PHI), is very important to both patients and medical practitioners. The important content of this information also makes it a lucrative point of attraction for on-premises and cyber (online) attacks.

The scope of this project is limited to analysis of major security breaches associated with unsecured protected health information (PHI) faced by healthcare organizations in the United States during the COVID-19 pandemic. According to the analysis, healthcare data breaches are on a progressive rise with an average of 2,247,842 breaches reported per month, which means roughly each month 250,000 patients had their health information compromised, from November 2019 through June 2021. Hacking and IT incidents are top breach types as attackers sort to locate electronic health data for moratory motives. At present, attacks on healthcare organizations’ data are 1.25 times higher compared to five years ago (IBM, 2021). This study was designed to investigate the growing concerns facing healthcare organizations and to propose some recommendations to help minimize the risk of attacks, in other to lower the chances of healthcare data breaches.

# **Background of Study**

Before the signing into law of the Health Insurance Portability and Accountability Act (HIPAA) of 1996, and the introduction of electronic health records (EHR) to be specific, most healthcare breaches of information and trust usually occurred on a small scale with a limited number of breaches. Attackers in that era relied on more primitive methods like using social engineering techniques to abuse human trust, shoulder surfing, breaking into record rooms (theft), misplaced document claims, undisclosed/unauthorized access, etc. During that era, a reporter could easily use social engineering to disguise and sneak into a facility as a staff to gain information on a high-profile patient. Similar incidents of this type were possible at that time because of very few established protocols as healthcare organizations were not regulated when it came to handling patient data.

Furthermore, prior to 1996, healthcare work settings, both clinical and administrative were left in the open when it came to information privacy and there was no federal rule governing the privacy and protection of health information. Most providers acted within their own established frameworks. A clear definition of what constituted protecting patient health information never existed at the federal level. Healthcare entities were fully in charge of managing patient information and health insurance updates. Patients could not access their own medical records unless they filed a lawsuit or involved law enforcement. Ironically, the best way back then to ensure your health data was safe (protected) was not to have it created in the first place. The reason being that patient paper-based health records were often seen exposed on doctor’s desk, unlock shelves, counters beside registration. This exposed information was often taken advantage of by individuals of nefarious intent, including sexual predators, identity thieves, insurance fraudsters, house robbers, and other criminals.

Another significant challenge which at that time was not noticeable was a mechanism to track who can access a patient’s information, what level of access, and how the accessed information is being used. All these were not possible since the healthcare industry at that time had no standardized and mandatory regulations on health data.

***The new era, (1996 - present):******The passage into law of The Health Insurance Portability and Accountability Act of 1996 (HIPAA).***

HIPAA is a federal law which requires the creation of national standards to protect sensitive patient health information from being disclosed without the patient’s consent or knowledge. The US Department of Health and Human Services (HHS) issued the HIPAA Privacy Rule to implement the requirements of HIPAA. The HIPAA Security Rule protects a subset of information covered by the Privacy Rule (CDC.gov)

## ***HIPAA Privacy Rule***

The Privacy Rule standards address the use and disclosure of individuals’ health information (known as “protected health information”) by covered entities (health plans, healthcare clearing houses, healthcare providers, and business associates). The Privacy Rule also contains standards for individuals’ rights to understand and control how their health information is used. One major objective of the Privacy Rule is to ensure that individuals’ health information is properly protected while allowing the flow of health information needed to provide and promote high quality health care and to protect the public’s health and well-being. The Privacy Rule strikes a balance that permits important uses of information while protecting the privacy of people who seek care and treatment. The HIPAA Privacy Rule applies to all forms of protected health information, whether electronic, written, or oral.

## 

## ***HIPAA Security Rule***

While the HIPAA Privacy Rule safeguards protected health information (PHI), the Security Rule protects a subset of information covered by the Privacy Rule. This subset is all individually identifiable health information a covered entity creates, receives, maintains, or transmits in electronic form. This information is called “electronic protected health information” (e-PHI). The Security Rule does not applyto PHI transmitted orally or in writing.

***Health Information Technology for Economic and Clinical Health Act of 2009 (HITECH): Breach Notification Rule***

Under the original HIPAA Privacy and Security Rules, millions of health records were at risk since the law did not require business associates to enforce HIPAA and only covered entities had a "contractual obligation" to comply with HIPAA. Prior to the HITECH Act of 2009, covered entities could avoid sanctions or penalties for a breach that involves any of their business associates by saying, they had no idea that the business associate was not HIPAA compliant. Because of this, many business associates could not be fined directly for HIPAA violations and HIPAA standards were not met.

The introduction of the HITECH Act of 2009 enforces the HIPAA Security and Privacy Rules for business associates and makes them directly liable for their own compliance with HIPAA. With this, business associates were obliged to sign a *Business Associate Agreement* with the covered entity on whose behalf they were processing PHI and had the same legal requirements as the covered entity to protect PHI and detect data breaches. Business Associates were also required to report data breaches to their covered entities.

In 2013, the *HIPAA Final Omnibus Rule* increased the commitment of business associates to compliance by subjecting them to HIPAA audits, civil and criminal penalties could be issued directly to business associates for the failure to comply with HIPAA Rules. With all these in place, then comes the *HITECH Breach Notification Interim Final Rule* requiring health care providers, health plans, and other entities covered by HIPAA to notify individuals when their health information is breached. According to this rule, as developed by the *Office for Civil Rights (OCR)*, healthcare providers and other HIPAA covered entities are required to promptly notify affected individuals of a breach, as well as the HHS Secretary and the media in cases where a breach affects more than 500 individuals. Breaches affecting fewer than 500 individuals are reported to the HHS Secretary on an annual basis. The regulations also require business associates of covered entities to notify the covered entity of breaches at or by the business associate (HHS.gov).

With the current framework in place, there has been an enormous number of breaches reported to the OCR over the years, something which covered entities could not previously account for. Many individuals have also received breach notification notices regarding breaches involving their PHI. Although this is seen as an act of accountability on the part of the covered entities, it also creates a feeling of mistrust in the eyes of the affected individual. It is in connection with this that the Acting Director and Principal Deputy Director of OCR Robinsue Frohboese had to say, “this new federal law ensures that covered entities and business associates are accountable to the Department and to individuals for proper safeguarding of the private information entrusted to their care.  These protections will be a cornerstone of maintaining consumer trust as we move forward with meaningful use of electronic health records and electronic exchange of health information.” (HHS.gov)

In 2019, the coronavirus (COVID-19) struck and was later declared a global pandemic in March 2020. The devastating consequences of this pandemic were especially hard on the healthcare industry given that it was a health crisis. Over a million lives have been lost coupled with economic hardship, and poor and vulnerable communities are affected the most. As the world battles to put the virus under control, organizations, governments, and researchers seek more and more information about the virus. Evidence suggests that the collection, use, sharing, and further processing of data can help limit the spread of the virus and aid in accelerating the recovery, especially through digital contact tracing. The World Health Organization (WHO), United Nations, and International Organization for Human Rights all call for human rights to be respected and healthcare entities to preserve human dignity to combat the virus, and that any data collection, use and processing should be rooted in human rights and implemented with due regard to applicable international laws, and UN Personal Data Protection and Privacy Principles (WHO). While the healthcare industry struggles with putting the pandemic under control, they also become a target for attackers as they sort to still the abundant data being generated within this period. For example, hacking incidents jump 42% while insider incidents affect 8 million patient records as reported by Protenus, a healthcare compliance analytics company. Furthermore, research shows ransomware attacks on healthcare organizations worldwide cost nearly $21B in 2020 (Berker’s Health IT), and in 2019, ransomware attacks on healthcare organizations all over the world each lasted an average of 287 days and cost an average of $8.1 million (Emsisoft).

# **Scope of the Project**

Even though there have been multiple reports of healthcare data breaches during the COVID-19 pandemic around the world as whole, this project shall be limited to the United States, and specifically only breaches involving covered entities in which 500 or more individuals have been affected and reported to the U.S. Department of Health and Human Services Office for Civil Rights within the last 24 months, as required by section 13402(e)(4) of the HITECH Act.

### ***Severity of the Problem Under Investigation***

In 2020 alone, in the heart of the deadly global pandemic (COVID-19) in the United Sates, healthcare organizations witnessed nearly 600 data breaches (healthdata.gov). The results of these breaches were negatively felt by both healthcare organizations (in terms of labor, finance, and reputation) and most profoundly by the patients (in terms of delay in receiving care, PHI leaks, emotional and psychological stress, stigma, etc.). Some of these attacks severely disrupted operations for hours or even days, putting patients’ lives at risk. For example, an attack on Scripps Health’s computer network disrupted care, forcing the healthcare provider to halt patient access to its online portal. The electronic medical records and the telemetry system meant for electronic monitoring of patient vital signs were shot down, forcing medical personnel to use paper records and manual recordings.

Understanding the consequences of a data breach is important, particularly in an industry like healthcare which deals with some of people's most confidential information. Healthcare organizations might easily recover back to operations from the consequences of a data breach, but long-term effects would need some time to fade away. The patients (public) may lose trust in the system, hence diminishing the organization’s reputation. A report by PwC shows that customer’s sentiment around cybersecurity and privacy risk reported that 92% of consumers agree that companies must be proactive about data protection and 85% of consumers will not request services at a business if they have concerns about their security practices (PwC, 2017). A good reputation is every healthcare entity’s most prized asset, and they must work constantly to build and maintain their integrity because just a single data breach can be very devastating.

Further, data security and privacy are essential to maintaining customers. In a Verizon survey, 69% of respondents would avoid a company that had suffered a data breach and 29% of those would never visit that business again (Verizon, 2019). From the above points, we see that the effects and implications of a data breach can be very challenging to overcome.

### ***Importance of the Topic and Researcher’s Point of Interest.***

The growing amount of data generated during the COVID-19 pandemic with the use of novel electronic devices and mobile applications calls for concern to investigate reported breaches during this period to see if healthcare organizations-maintained compliance with data regulations during this period. The analysis of these breaches will help unpack some useful information that could serve in addressing future challenges in protecting PHI and to give some recommendations on healthcare data protection during future pandemics.

The main point of interest in the topic under investigation is the increasing expansion of Health Information Technology to incorporate the full use of EHRs, Internet of Medical Things (IoMT) Telehealth/Telemedicine, the migration of healthcare data to Cloud, and the vulnerability of healthcare organizations to cyber attackers. With all these innovations, the need for tougher data protection protocols in healthcare organizations is inevitable.

Data breaches in healthcare are becoming rampant and more expensive. In 2020, healthcare data breaches totaled 599, up 55.1% from 2019, and affected more than 26 million people, according to Bitglass, a cloud security company. In addition, the average cost of a breach in healthcare increased by 10.5% from 2019 to 2020, and the cost per breach also rose to $499 in 2020, from $429 in 2019, a 16.3% jump (Bitglass, 2020). This means healthcare organizations lost close to $13.2bn because of breaches in 2020 alone. A recent report by IBM also identifies healthcare as the most expensive industry for a data breach with an average of $7.13 million per breach (IBM, 2021).

A report from Kaspersky Lab, a Russian multinational cybersecurity and anti-virus provider, highlights the human side of data breaches (cybersecurity incidents) by examining the discomfort and losses employees in charge of personal data face following corporate breaches. Around 30% of employees who have been involved in the aftermath of a data breach have missed an important personal event, 32% had to work overnight, 33% suffered additional stress, while 27% have had to cancel vacations (Kaspersky,2020). The Kaspersky report only considered the human cost on workers (healthcare workers). A closer look at the human side of a breach on the patient could be very dangerous; it might result in loss of life and delay in providing care due to closure of services at a care facility. A breach involving one’s HPI might also lead to psychological stress, trauma, social stigmatization, and even financial loss for the patient, since their personal information could be sold to fraudsters.

# **Stakeholders**

When it comes to addressing issues related to healthcare data breaches, HIPAA and HITECH are the lawful standards put in place by the federal government. The Department of Health and Human Service (HHS), through the Office for Civil Rights, is required to enforce these laws on covered entities: health plans, healthcare clearing houses, healthcare providers, and business associates. Since healthcare is an issue of national interest, each stakeholder, beginning with the patient at the base, through covered entities and HHS, and finally to the law makers, each has a part to play. From the formulation of regulations at the level of the senate and back down to the practical implementation by the business associates and the patients (since patients also have the responsibility to protect their health data or else face the direct consequences of their actions).

In summary, everyone enlisted above can be affected negatively when a breach occurs, but most of the burden is always felt by the patient whose information is being compromised. The federal government and covered entities may suffer mainly from financial loss and in maintaining their integrity, but the bulk of the direct human cost goes back to the patient.

## ***Potential Causes of the Problem under Investigation***

***Perverse business incentives***

The notion of perverse business incentives when it comes to data healthcare data breaches, most especially during the COVID-19 pandemic period, points to the gaps that exist regarding what unregulated medical device manufactures do with the data these devices collect, and how the use of technology can be very disastrous in the hands of bad actors. To a lesser extent, the introduction of EHRs to enhance healthcare productivity by increasing access to care using electronic records has tended to attract hackers as they keep trying to gain access to the sensitive information stored in these systems.

***Resource deficit***

Addressing resources in terms of available workforce, the U.S. Department of Labor’s Bureau of Labor Statistics predicts that the employment of medical and health services managers will grow by 32 percent between 2019 and 2029—far outpacing the average for all occupations in the United State (dol.gov). This data sounds good and promising, but the reality lies in the shortage of qualified and skilled health IT and health information management professionals who are aware of new technologies and regulations.

The problem faced with a workforce shortage can somehow be linked to funding. Healthcare organizations need to allocate more funds to step up training for their staff on best data protection practices and invest in the maintenance of their IT infrastructure (information system). A survey of healthcare IT professionals found 65% say their organization invests 6% or less of the IT budget on security. This is about half of the 12-15% of the IT budget other regulated industries spend on security according to an analysis of the survey (Symantec, 2021).

However, if the Bipartisan Infrastructure Framework announced on June 21, 2021, is confirmed by lawmakers, then government agencies and some private agencies (organizations) will have funds to invest in data protection infrastructures given that cybersecurity is also listed in the proposed bill as a major concern.

## ***Policy deficit or lack of government regulation***

There is a policy deficit that needs to be addressed. The rising number of healthcare data breaches lately can be partly attributed to open trades using cryptocurrencies. Because of the very little to no regulation on doing business using cryptocurrency, attackers (hackers mostly) have gained momentum knowing that they can sell stolen data or ask for ransoms over a comprised system using cryptocurrency without being tracked. For example, while hospitals were struggling with the pandemic, the University of California Hospital, San Francisco paid hackers $1.14 million in bitcoin after a ransomware attack in June 2020. (BBC News, 2020).

          Nevertheless, the attack on the Colonial Pipeline on May 7, 2021, triggered a major policy change in the nation’s cybersecurity policy. On May 12, 2021, the President signed an "Executive Order on Improving the Nation’s Cybersecurity" identifying prevention, detection, assessment, and remediation of cyber incidents as a top priority and essential to national and economic security (Whitehouse.gov, 2021). This executive order requires the federal government to partner with the private sector to make significant changes and investments to defend the vital institutions, mobilize resources to protect and secure computer systems including data processing systems (especially in federal offices), whether they are cloud-based, on-premises, or hybrid. Among other duties, the federal government must carefully investigate and examine major cyber incidents and apply lessons learned. Also, the private sector must commit to the continuously changing threat environment, ensure its products are built and operate securely, and partner with the federal government to foster a more secure cyberspace.

## ***Public Awareness***

Inasmuch as healthcare compliance can be seen as a process, of which it has been proven given the multiple updates on HIPAA since 2019, one can generally say there is still much to be done in the United States on consumer data as whole. In Europe, for example, privacy regulations are evolving, with a marked shift toward protecting consumers. The European General Data Protection Regulation (GDPR) of May 2018 gave consumers more choices and protections about how their data is used. The GDPR gives consumers easier access to data that companies hold about them and makes it easier for them to ask companies to delete their data (GDPR, 2018). One may ask why this is important for the Unites States healthcare system? The answer is simple: given the rise of Internet of Medical Things (IoMT) devices and other smart devices that collect one’s health data, e.g., Apple watch, the federal government will need to formulate regulations like the GDPR. This is particularly important because a huge deal of individual healthcare data ends up in the hands of noncovered entities. The regulations on data protection specifically on medical device manufactures need to be enhanced.

However, in the United States, the state of California has taken the lead by enacting the California Consumer Privacy Act (CCPA) which went into effect in January 2020. With this, residents have the right to know which data is collected about them and to prevent the sale of their data. As of now, the CCPA is the strictest consumer-privacy regulation in the United States and applies to profit making organizations that do business in California and satisfy one of the following criteria: earning more than half of their annual revenues from selling consumers’ personal information or holding personal information on more than 100,000 consumers earning gross revenues of more than $50 million, households, or devices. (McKinsey & Companies, 2020). Finally, as mentioned above, the "Executive Order on Improving the Nation’s Cybersecurity" signed by President Biden on May 12, 2021, marked a significant milestone at the federal level on protecting and defending the nation’s cyberspace and information infrastructure.

## ***Current Trajectory***

Interestingly, the concept of data breaches in healthcare organizations is moving towards the direction of the influence of technology in healthcare and society at large. There is a noticeable shift from traditional methods of physical or on-premises attacks (theft, missing files, and forced entry) to frequent online or cyber-attacks. More than 90% of attacks that involve breaches on healthcare organizations in the last 24 months are linked to the IT infrastructure of these organizations. Hackers have identified the healthcare industry to be amongst the most venerable industries and took advantage of that during the outbreak of the pandemic. In general, most data breaches in healthcare nowadays are related to cyber security or EHR systems, and often come as threats in one of the following methods.

**Malware:** Malware is any software intentionally designed to cause damage to a computer, server, client, or computer network. Malware is the contraction of “Malicious Software” and include a wide variety, including computer viruses, worms, Trojan horses, ransomware, spyware, adware, rogue software, wiper, and scareware. (Cisco)

**Distributed Denial of Service (DDoS):** A DDoS attack occurs when legitimate users are unable to access information systems, devices, or other network resources due to the actions of a malicious cyber threat actor. Services affected may include email, websites, online accounts, or other services that rely on the affected computer or network. A denial-of-service condition is accomplished by flooding the targeted host or network with traffic until the target cannot respond or simply crashes, preventing access for legitimate users. (cisa.gov)

**Advanced Persistent Threats (APT):** An adversary that possesses sophisticated levels of expertise and significant resources which allow it to create opportunities to achieve its objectives by using multiple attack vectors including cyber, physical, and deception. These objectives typically include establishing and extending footholds within the IT infrastructure of the targeted organizations for purposes of exfiltrating information, undermining or impeding critical aspects of a mission, program, or organization or positioning itself to carry out these objectives in the future. The advanced persistent threat pursues its objectives repeatedly over an extended period, adapts to defenders’ efforts to resist it, and is determined to maintain the level of interaction needed to execute its objectives. (nist.gov)

As noted above, the average cost of a breach in healthcare increased by 10.5% from 2019 to 2020, and the cost per breach also rose from $429 in 2019 to $499 in 2020, a 16.3% jump (Bitglass, 2020); healthcare organizations lost close to $13.2bn because of breaches in 2020. Also, IBM has rated healthcare as the most expensive industry for a data breach with an average of $7.13 million per breach (IBM, 2021). This staggering statistic has more meaning, especially when we look at the costs and the budget for preventing them. A survey of healthcare IT professionals found 65% say their organization invests 6% or less of the IT budget on security. This is about half of the 12-15% of the IT budget other regulated industries spend on security according to an analysis of the survey (Symantec, 2021).

# **Methodology**

Python (Version 3.8.6) was used, with Jupiter Notebook as the integrated development environment (IDE) for statistical computation, analysis, and visualization of the data.

The IDE was prepared for analysis by importing(installing) the following packages or libraries:

**NumPy*:*** this is the fundamental package for scientific computing in Python. NumPy arrays facilitate advanced mathematical and other types of operations on large numbers of data.

**Pandas*:*** an open-source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of NumPy.

**Matplotlib*:*** this is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose graphic user interface (GUI) toolkits.

**Seaborn*:*** this a data visualization library for statistical graphics plotting in Python. It provides beautiful default styles and color palettes to make statistical plots more attractive. It is built on the top of the matplotlib library and closely integrated to the data structures from pandas.

**SciPy*:*** this is a free and open-source scientific computation Python library used for technical computing.

After getting the IDE prepared by installing the packages, the dataset was then uploaded from the saved file location on the computer to the IDE. The dataset was prepared for analysis by cleaning it to remove unwanted characters, columns, and null values. This was done just to ensure the data was good for analysis without altering or adding any new data (entry) to the existing dataset that was collected, therefore the integrity of the data was maintained. *(See Appendix 1.b)*

A stratified sampling model was referenced to examine reported healthcare data breaches during the COVID-19 pandemic. This method allowed the data to be grouped into the various months. To better understand the trend in the incidents reported, the total number of affected individuals was filtered out from November 2019 through May 2020 and further subdivided into November 2019 – August 2020 and September 2020 – May 2021. These two new sub datasets were then used to perform a T-test at a 95% confidence level to determine if there was a significant difference in the means of the number of individuals affected over these periods. Note that June 2021 was left out because the data was incomplete at the time this project was done.

## ***Theoretical framework***

The purpose of this project was to investigate variations in data breach patterns and security threats of unsecured Protected Health Information during COVID-19. The study also examined the human and policy impact on PHI during the noted period and proposes some recommendations to address identified weaknesses.

       The official declaration of COVID-19 in March of 2020 as a pandemic by WHO and the Centers for Disease Control and Prevention (CDC) triggered adjustments to some healthcare regulations to mitigate the virus. Here in the United States, the Department of Health and Human Services Office for Civil Rights (OCR) issued waivers and notices of enforcement discretion related to HIPAA compliance. The COVID-19 blanket 1135 waiver issued in March by the OCR included, among others, the following:

* Allowed the patient to request confidential communication and privacy restrictions, to opt out of the facility directory, and receive a notice of privacy practices.
* Business associates whose contracts did not previously allow for data sharing, were given permission to share certain data for public health with the CDC.
* The waiver also gave guidance on how healthcare providers could share COVID-19 infection information with employees, first responders, or family members, to protect others from becoming infected.
* The OCR eased its stance on communications technology, which allowed flexibility with communication technologies providers could use to communicate with their patients. This permitted providers to use technologies that do not currently comply with the HIPAA Security Rule, like FaceTime and Skype for telehealth.

Even with the difficulties of the pandemic, healthcare organizations must comply with the privacy regulations. Covered entities were expected to understand how these waivers apply to them, and to ensure there were being used just to provide good patient care, and nothing else i.e., the minimum necessary standard which require healthcare professionals to make reasonable efforts to ensure any protected health information disclosed is restricted to the minimum necessary information to achieve the purpose for which the information is being disclosed. Organizations were also expected to maintain relevant documentation on how they implemented the minimum necessary standard.

With all these changes and pressure on healthcare organizations, people became stressed, distracted more susceptible to phishing and other online attacks. People worked remotely from unprotected environments(networks) with little or no security and privacy controls that were vulnerable to online HIPAA security threats. This gave hackers the opportunity to take advantage of people and systems as they were busy dealing with the pandemic crisis.

This project was designed based on the hypothesis that "The global pandemic has served to further underscore the consequences of health data vulnerability. There is an opportunity within the U.S to leverage publicly available data on breaches that have occurred over the course of the COVID-19 response, to draw actionable knowledge about human and technology patterns across these health data breaches".

## ***Data/Evidence***

The data used in this study came from the U.S. Department of Health and Human Services (HHS) Office for Civil Rights (OCR). Coupled with other duties, OCR enforces the Health Insurance Portability and Accountability Act (HIPAA) Privacy, Security, and Breach Notification Rules, and the Patient Safety Act and Rule, which together protect individual fundamental rights of nondiscrimination, conscience, religious freedom, and health information privacy at covered entities. The website contains a list of all reported healthcare data breaches involving covered entities over the years as required by section 13402(e)(4) of the HITECH Act, "the Secretary must post a list of breaches of unsecured protected health information affecting 500 or more individuals". The website also provides information and directives for individuals to report any violation relating to their health information privacy rights or any other violation under the HIPAA Privacy, Security, and Breach Notification Rules. *(See Appendix 1.a for dataset link)*

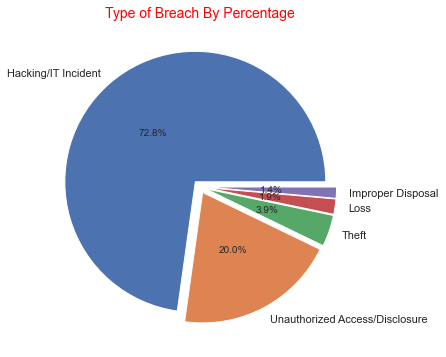
         This analysis was done based on reported breaches in the last 24 months with emphasis on when the COVID-19 pandemic started, November 2019 till June 9, 2021. The dataset contained the following columns: *Name of Covered Entity, State, Covered Entity Type, Individuals Affected, Breach Submission Date, Type of Breach, Location of Breached Information, Business Associate Present, and Web Application* with a total of 796 breaches.

**Results**

### ***Total Number of breaches under investigation***

According to the dataset, there were 796 incidents of breach under investigation dating from November 2019 through June 2021. These incidents amount to a total of 62,308,067 patients' health data. The types of breaches and the respective number of individuals involved is a follow. *(See Appendix.1.c & d for code)*

|  | **Type of Breach** | **Individuals Affected** |
| --- | --- | --- |
| **0** | Hacking/IT Incident | 58416763 |
| **1** | Unauthorized Access/Disclosure | 2994170 |
| **2** | Improper Disposal | 625866 |
| **3** | Loss | 144045 |
| **4** | Theft | 127223 |

****From the summary above, we see that most breaches (> 72%) under investigation that occurred during the COVID-19 pandemic were associated with hacking/IT incidents, 20% involved unauthorized access/disclosure, 3.9% for incidents involving theft, 1.9% involved data loss, and 1.4% for improper disposal of PHI.

*A Pie Chart showing location of breach by percentage.*

### 

### ***Analyzing Breach Location***

From the results above, most breaches involved hacking/IT incidents, therefore it was worth doing further analysis on the location of these incidents. Below is a summary of the findings.  *(See Appendix.1. e for code)*

**Location of breach**

Network Server: 329

Email: 265

Paper/Films: 54

Other: 36

Electronic Medical Record: 30

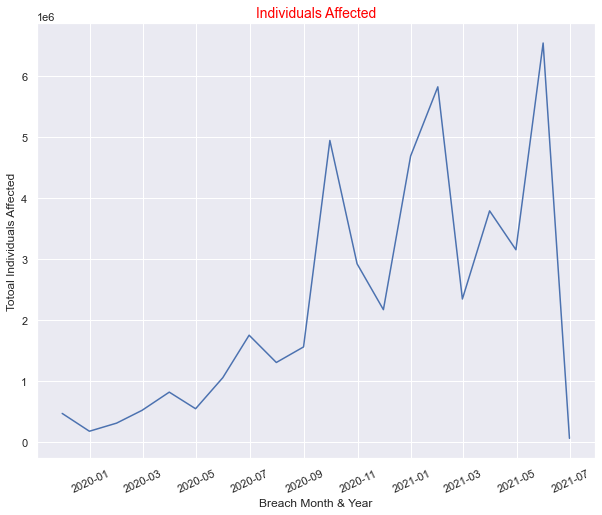
Other Portable Electronic Device: 10

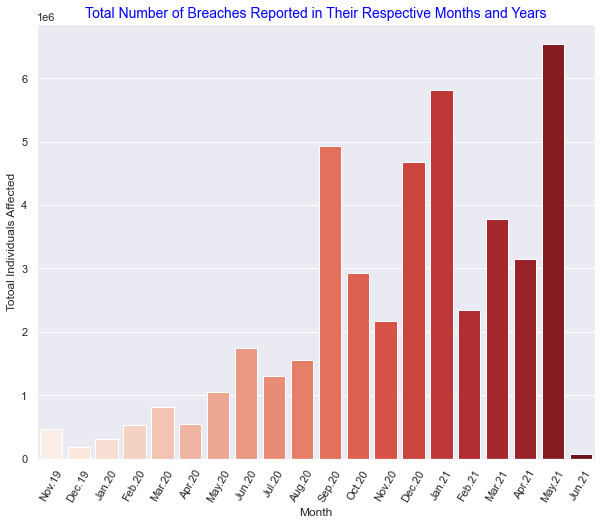
Laptop: 9

The analysis showed that network servers and emails were the main target breach locations, recording about 75%. Paper/films, electronic medical records, portable devices, and laptop locations were targeted 54,30,10 and 9 times, respectively**.**

### ***Analyzing the Number of Individuals Affected Over Time***

An analysis of the total number of individuals affected by data breaches from Nov 2019 through June 2021 was done using a Time Series as shown in the graph below*. (See Appendix 1.f & g for code)*

****

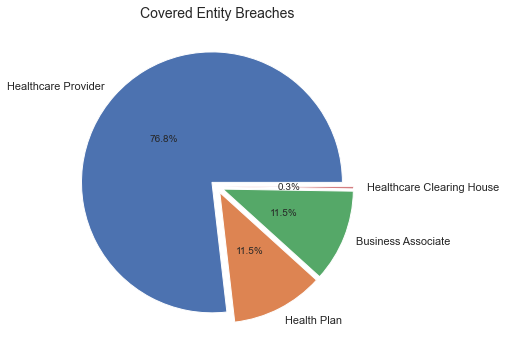
**** *A Time Series* *Showing Individuals Affected (Patients) by Data Breach Each Month.*

*A Bar Graph Showing Individuals Affected (Patients) by Data Breach Each Month*

To simplify the time series graph above, a bar plot was designed to visually show the breakdown by months. It shows a general upward trend in the number of individuals affected (patients) by data breach, with peak values reported in September 2020; December 2020; January 2021; February 2021; and May 2021.

### ***Investigating Types of Covered Entities Most Affected by Data Breach.***

Given how significant the reputation of a covered entity is at stake when involved in a breach, it was also worth looking at how they were affected by these breaches*. (See Appendix 1.h & I for code)*

 **Covered Entities Involved**

Healthcare Provider 610

Health Plan 91

Business Associate 91

Healthcare Clearing House 2

The data reveled that close to 77% of the reported breaches that took place from November 2019 through June 2021 involved healthcare providers, business associates and health plans, followed with a tie (91 breaches), and lastly, healthcare clearing houses with just 2 breaches.

### ***Number of Business Associates Affected***

**BA Involved in breaches**

No 512

Yes 282

282 of the reported breaches involved business associates, which represented about 35% of the total number of breaches. This finding was equally important to know how compliant business associate are with HIPAA. *(See Appendix.1. j for code)*

***States with the Most Breaches***

An analysis was done to know how different sates were affected by these breaches, and the result below represents a summary of states that recorded 20 or more breach incidents.

TX 70

NY 69

CA 65

FL 44

PA 38

OH 37

MI 31

IL 28

MA 28

MD 24

MO 22

GA 21

WI 20

MN 20

As seen, Texas, New York, California, and Florida were the most affected with breach incidents. Remarkably, theses states were also those most affected by the COVID-19 pandemic in the United States.

### ***T-test at 95% Confidence Level***

Lastly an analysis was done to determine the statistical significance in the trend on the reported number of individuals affected. *(See Appendix 1.l for code)*

Ttest\_indResult(statistic=-5.870720186402016, pvalue=0.00017243609971119003)

The result of the T-test showed the “t” statistic of -5.9 which indicated how much the sample mean deviated from the null hypothesis. Also, given the very low p-value (0.000173) from the result, therefore there was a significant statistical difference in the sample data hence we reject the null hypothesis. Thus, there was a general increase in the number of individuals affected by healthcare data breach from November 2019 through May 2021.

# **Discussion**

There are numerous reasons why HIPAA compliance is vital for every healthcare organization, but the following are the most crucial: its goals include ensuring privacy and confidentiality, giving patients access to their medical records, and reducing fraud by increasing accountability and improving data systems. Everything comes down to data security.

HIPAA provides several major benefits to healthcare organizations to aid in the transition from paper records to electronic health records. It helps in the streamlining of administrative healthcare activities, the improvement of healthcare industry efficiency, and the secure sharing of protected health information. HIPAA mandates require healthcare providers, health plans, healthcare clearinghouses, and business associates of HIPAA-covered entities to put in place several safeguards to protect sensitive personal and patient protected health information (PHI). While no organization would want its sensitive data to be exposed or stolen, without HIPAA, there would be no duty for healthcare firms to protect data – and no consequences if they fail to.

HIPAA Privacy rules require healthcare organizations to control who has access to health data, limiting who can view health information and with who the information might be shared. HIPAA ensures that any information disclosed to healthcare providers and health plans, as well as information created, transmitted, and retained by them, is subject to stringent security safeguards. HIPAA promotes patient-participation in their own care by providing easy access to their health records (in electronic format). Another important part of HIPAA is the Breach Notification Rule. Under this rule, a breach of unprotected PHI compels covered entities to notify affected individuals, HHS, and in some situations the media, and notifications must be sent out without undue delay no later than 60 days after the breach has been discovered.

Analyzing the increasing trend of healthcare data breaches in the United States during the outbreak of COVID-19, it became evident that the main motive of a data breach is to sell the compromised information. Medical data is unique in that it often contains vital information for hackers, such as credit card information, and Social Security and bank account numbers, which can give them access to one’s finances and investments. With this information, fraudsters can create fake IDs to buy medical equipment or drugs that can be resold, file fake insurance claims, or sometimes use this data to blackmail a patient to extract money from them.

Further analysis also indicates that most big breaches of healthcare servers result from hacking or IT, which is an indication that most healthcare IT systems are vulnerable. Hence, those in charge of healthcare servers should pay great attention to developing a secure framework to protect them. In most breaches involving healthcare organizations, they usually end up with a double penalty by having to pay ransom to the hackers to get their breached data (the system) back or to restore their hacked system and, secondly, they are expected to pay a government penalty for failing to safeguard patient information. This situation eventually leads to an increase in healthcare costs, a reduction in better healthcare delivery, and a loss of patient trust. When a healthcare organization experiences a breach of PHI, this alone can drastically affect its reputation and, thus, its business. Research shows that people might withhold their health information from healthcare providers because they are concerned that there could be a confidentiality breach, making them potential victims. This unwillingness to fully disclose information when receiving care could lead to delays in diagnosis, an increase in treatment costs, loss of life, spread in the case of a communicable disease, and hinder research activities.

***Some general recommendations to reduce the risk of health data breaches***

There are several measures that healthcare organizations can implement to successfully manage risks associated with data breaches involving PHI. Most of these measures seem common and familiar within most industries, yet effectively implementing them is a challenge.

1. De-identifying or, better still, hashing PHI in health information systems, backups, and health data warehouses, might reduce the risk of a data breach significantly. Because this renders de-identified or hashed healthcare data virtually useless to hackers. Doing this also protects the privacy of the patient in the case of a breach.
2. Software used should be updated and patched regularly to prevent hackers from identifying loopholes within the system. Only recommended updates from the software company should be implemented, and free software from unknown or un-trusted sources should not be used. Security patch management should be used to reduce the risk of compromise of systems, applications, and computers because of system flaws, adding to the fact that medical devices are at high risk after experiencing a security patch update.

1. Meaningful use should include evidence of IT training for doctors and nurses on the handling of the PHI of patients. Training should include best security practices when dealing with PHI on their own laptops, tablets, pen drives, workstations, and sending work-related emails. Training should include tips like creating stronger passwords using a mix of characters, avoid writing passwords on sticky notes on their computers or workstations, always signing out of their accounts when finished inputting patient information or after viewing patients' reports, using screen filters in exposed areas, proper password management rules- thereby reducing insider threats, etc.
2. Healthcare organizations should reinforce identification and authentication techniques such as biometric and two factor authentication. These mechanisms would help to identify system users as well as confirm that information is from a trusted source. Using these techniques will only approve access to permitted individuals and prevent restricted individuals, most of who have malicious intentions. Identification and authentication can help prevent intrusion into the health information system, loss of private information, and loss of services. Every healthcare organization should implement robust identification and authentication techniques, given how sensitive the information in their possession is.

1. Firewalls should be used to regulate access to the computer network for authorized users only. With the proliferation of the internet and medical devices, firewalls should be configured to block any unauthorized device from entering a healthcare organization's information network using built-in filters. Network logging attempts should also be analyzed because the biggest danger to your firewall is the public internet.
2. Isolating the network by separating and grouping computer networks into logically manageable entities (Virtual Local Area Networks—LANs) can help prevent system failure while enhancing management and making the network redundant. This isolation is very important in a healthcare organization since its network usually contains different electronic medical devices. This technique would also serve in part to address security challenges linked to the interoperability of medical devices due to their capability to communicate wirelessly. This wireless communication can permit unauthorized modifications of the device’s parameters, which may harm a patient, cause severe damage to their hearts, and even lead to death. Unprotected medical devices may also be used by an attacker as an access point into the organization’s network (Halperin et al, 2008).
3. Given that most data breach attacks that affect healthcare organizations nowadays are cyber related, the use of Malware Detection and Protection (MLDP) is equally necessary. The installation of a powerful MLDP system could provide well-defined technical controls. Such a system would be able to defend against malicious software such as viruses, worms, Trojan viruses, spyware, adware, and ransomware, which are often used by cybercriminals.
4. Putting in place standardized policies and systems within every healthcare organization would make it possible to identify and manage safety requirements, system effectiveness, security strategies, and guidelines. For example, in March 2011, Massachusetts General Hospital paid a million dollars in legal fees and penalties for the loss of 192 patient records after a hospital billing manager took these records (paper files) out of the hospital to work on them from home but accidentally left them on an MBTA subway train. Instead of spending a lot of money paying for breaches, healthcare organizations should increase their budget for IT infrastructure and healthcare compliance training to prevent such unnecessary spending.

# **Conclusion**

The use of EHR makes it easy to share healthcare data among providers, insurers, and researchers. It also allows the recognition of interesting medical patterns, reductions in medical errors, development of personalized and predictive medicine, elimination of insurance fraud, better disease management, predicting and preventing disease outbreaks, and identification of low-cost treatments just to name a few. On the other hand, unprotected health information poses a significant threat to patient privacy and the systems that manage this information might be accessible to hackers. HIPAA and HITECH are effective regulations when properly implemented on PHI in every healthcare organization. The failure to ensure such compliance poses a continuous problem for the integrity, management, healthcare cost, and reliability of healthcare organizations. Noncompliance can be very detrimental to patient’s privacy with significant effects on human costs.

This study unveils that during the COVID-19 pandemic most healthcare data breaches involved hacking IT incidents on servers and network systems, a clear indication that healthcare cybersecurity should be of top priority to every healthcare organization. A variety of strategies and policies can be used to thoroughly adapt compliance towards the goals and objectives of healthcare organizations to limit the number and severity of attacks on healthcare data. Healthcare organizations need to look at cybersecurity beyond the context of compliance on its own but rather as an integral part of HIPAA and HITECH. This holistic approach to privacy and security ensures that all statutory obligations are met, as well as security measures and policies are implemented.

Furthermore, healthcare organizations should not approach security just from a technical perspective. They should consider the socio-technical aspect of security. They should consider the critical role of “human elements” such as employee awareness, perceptions, and values in security management. Healthcare organizations should consider incorporating security into their organizational culture. This entails the design of a “system of collective moral concepts, mindsets and behavior patterns” (Brady, 2011). Integrating a security philosophy is a gradual and ongoing process, therefore, it is essential that management sets the tone for the rest of the organization. Employees must recognize that management considers “security as a core part of the business” (Johnson & Goetz, 2007). Healthcare organizations must continuously foster the concept of data security by making sure that employees are up to date with security issues and their respective roles in mitigating and minimizing risks. This will result in an institutionalized culture of security awareness that is consistently practiced by all employees. Such an approach will ensure that privacy and security compliance implementation become an essential part of the job function of every employee within healthcare organizations. When reactionary measures are supplemented with proactive analysis of evolving threat vectors, adherence to healthcare data security compliance can be easily achieved.

Lastly, adopting a security framework designed for healthcare particularly would assist healthcare organizations achieve a firm and adequate position on security. For instance, the HITRUST Common Security Framework adapts the ISO 27001 for the healthcare industry, identifies relevant security controls to meet all compliance requirements and presents a tailored, tested, and proven approach to address healthcare security concerns that takes into consideration HIPAA, HITECH, Information (and Cyber) security, and Risk Management. For example, the framework considers the complexity of the hospital’s information systems, the “maturity of the current security processes and controls” and its resource constraints (HITRUST Alliance, 2014).

# **Bibliography**

1. Solove, D. (April 2013). "HIPAA Turns 10: Analyzing the Past, Present and Future Impact" *Journal of AHIMA* 84, no.4 (April 2013): 22-28. Retrieved from <https://library.ahima.org/doc?oid=106325#.YNwe8uhKhyw>. [Accessed 10 June 2021]
2. Alder, S. (June 30, 2021). Government Watchdog Makes 7 Recommendations to HSS to Improve Cybersecurity*. HIPAA Journal*. Retrieved from <https://www.hipaajournal.com/government-watchdog-makes-7-recommendations-to-hss-to-improve-cybersecurity/> [Accessed 9 June 20121]
3. Alder, S. (June 01, 2021). Best Password Manager for the Healthcare Industry. *HIPAA Journal*. Retrieved from <https://www.hipaajournal.com/best-password-manager-for-the-healthcare-industry/> [Accessed 9 June 20121]
4. Alder, S. (June 18, 2021). May 2021 Healthcare Data Breach Report. *HIPAA Journal*. Retrieved from <https://www.hipaajournal.com/may-2021-healthcare-data-breach-report/> [Accessed 9 June 20121]
5. Mohammed, D., Mariani, R., Mohammed, S. (February 2015). Cybersecurity Challenges and Compliance Issues within the U.S. Healthcare Sector. *International Journal of Business and Social Research Volume 05.*
6. Moore, J. (July 24, 2014). Health care providers look to improve security incident response. iHealthBeat. Retrieved from <http://www.ihealthbeat.org/insight/2014/health-care-providers-look-to-improvesecurity-incident-response> [ 01 July 2021]
7. Schultz, D. (2012). As patients’ records go digital, theft and hacking problems grow. Kaiser Health News. Retrieved from http://www.kaiserhealthnews.org/Stories/2012/ June/04/electronic-health-recordstheft-hacking [29 June 2021]
8. Office for Civil Rights (OCR). (January 05, 2015). Submitting Notice of a Breach to the Secretary. Retrieved from <https://www.hhs.gov/hipaa/for-professionals/breach-notification/breach-reporting/index.html> [Accessed 12 June 2021]
9. Office for Civil Rights (OCR). (October 6, 2021). Breach Notification Regulation History. Retrieved from <https://www.hhs.gov/hipaa/for-professionals/breach-notification/laws-regulations/index.html> [Accessed 12 June 2021]
10. Office for Civil Rights (OCR). (July 25, 2017). HIPAA Enforcement. Retrieved from <https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/index.html> [Accessed 12 June 2021]
11. U.S. Department of Health & Human Services. (April 2, 2021). HIPAA and COVID-19. Retrieved from <https://www.hhs.gov/hipaa/for-professionals/special-topics/hipaa-covid19/index.html> [ Accessed 19 June 2021]
12. U.S. Department of Health & Human Services. (September 13, 2019). Omnibus HIPAA Rulemaking. Retrieved from <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/combined-regulation-text/omnibus-hipaa-rulemaking/index.html>[Accessed 19 June 2021]
13. Khan, S., Sayed, A., Hoque, L. (2016). Digital Health Data: A Comprehensive Review of Privacy and Security Risks and Some Recommendations. *Computer Science Journal of Moldova*, vol.24, no.2(71), 2016
14. (June 2021). Breach portal: Notice to the secretary of HHS breach of unsecured protected health information. Retrieved from <https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf> [Accessed 30 June 2021]
15. IBM and P. Institute, (n.d.). “2020 cost of data breach study: Global analysis,” IBM and Ponemon Institute, Research Report. Retrieved from <https://www.ibm.com/security/digital-assets/cost-data-breach-report/#/> [Accessed 14 June 2021]
16. Ponemon Institute (2016), The State of Cybersecurity in Healthcare Organizations in 2016.
17. Symantec (2016), Healthcare Internet Security Threat Report, Vol 21.
18. Scott, J. (November 2018). Protecting medical devices from hackers. *Saint John Patrick Publishers* Ltd
19. Health IT.gov. (n.d.). Top 10 Tips for Cybersecurity in Health Care. Retrieved from <https://www.healthit.gov/sites/default/files/Top_10_Tips_for_Cybersecurity.pdf>. [Accessed 26 June 2021]
20. Health IT.gov. (June 2021). HHS Announces $80 million in American Rescue Plan Funding to Strengthen U.S. Public Health IT, Improve COVID-19 Data Collection, and Bolster Representation of Underrepresented Communities in Public Health IT Workforce. Retrieved from <https://www.hhs.gov/about/news/2021/06/17/hhs-announces-80-million-in-arp-funding-to-bolster-underrepresented-communities-in-public-health-it-workforce.html> [Accessed 27 June 2021]
21. Yadron D. (2016), Los Angeles Hospital Paid $17,000 in Bitcoins to Ransomware Hackers, Retrieved from <https://www.theguardian.com/technology/2016/feb/17/los-angeles-hospital-hacked-ransombitcoin-hollywood-presbyterian-medical-center> [Accessed 21 June 2021]
22. Matt, B. (March 2020), What is GDPR? The summary guide to GDPR compliance in the UK. <https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018> [Accessed 14 June 2021]
23. State *of* California Department *of* Justice. (May 2018). California Consumer Privacy Act (CCPA). Retrieved from <https://oag.ca.gov/privacy/ccpa> [Accessed 20 June 2021]
24. The White House. (May 2021), FACT SHEET: President Signs Executive Order Charting New Course to Improve the Nation’s Cybersecurity and Protect Federal Government Networks. Retrieved from <https://www.whitehouse.gov/briefing-room/statements-releases/2021/05/12/fact-sheet-president-signs-executive-order-charting-new-course-to-improve-the-nations-cybersecurity-and-protect-federal-government-networks/> [Accessed 22 June 2021]
25. Shin, S., Lipton, J. (August 2018), Security researchers say they can hack Medtronic pacemakers. Retrieved from <https://www.cnbc.com/2018/08/17/security-researchers-say-they-can-hack-medtronic-pacemakers.html> [Accessed 21 June 2021]
26. Costello T. (2016). Hacking of Health Care Records Skyrockets, Retrieved from <http://www.nbcnews.com/news/us-news/hacking-health-care-records-skyrockets-n517686>[Accessed 07 June 2021]
27. Coventry, L., Branley, D. (2018). Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. *Maturitas*, *113*, 48–52. <https://doi.org/10.1016/j.maturitas.2018.04.008> [Accessed 11 June 2021]
28. Ogasawara, G. (May 2021). Why Ransomware Poses a Threat to Both Providers & Patient. Retrieved from <https://hitconsultant.net/2021/05/20/ransomware-healthcare-organizations-patient-health%E2%80%8B/> [Accessed 12 June 2021]
29. EMSISORFT. (2019). State of Ransomware in the U.S.: 2019 Report. Retrieved from <https://blog.emsisoft.com/en/34193/state-of-ransomware-in-the-u-s-2019-report> [Accessed 20 June 2021]
30. Carson, J. (2019). Key Takeaways from the 2019 Verizon Data Breach Investigations Report. Retrieved from <https://thycotic.com/company/blog/2019/05/21/2019-verizon-data-breach-investigations-report-takeaways/> [Accessed 18 June 2021]
31. Bitglass (2020), Healthcare Breach Report 2020. Retrieved from <https://pages.bitglass.com/rs/418-ZAL815/images/2020_Healthcare_Breach_Report.pdf?aliId=eyJpIjoicXZKN0ZJTmdSM0czc245MyIsInQiOiJCRkNHa3AwTFJ0NU5uZ29WK3pxYStRPT0ifQ%253D%253D> [Accessed 19 June 2021]
32. Algarni, A.M., Thayananthan, V., Malaiya, Y.K.(19 April 2021 ).Quantitative Assessment of Cybersecurity Risks for Mitigating Data Breaches in Business Systems. Appl. Sci. 2021, 11, 3678. Retrieved from https:// doi.org/10.3390/app11083678 [Accessed 19 June 2021]
33. World Health Organization (November 2019), Joint Statement on Data Protection and Privacy in the COVID-19 Response. Retrieved from <https://www.who.int/news/item/19-11-2020-joint-statement-on-data-protection-and-privacy-in-the-covid-19-response> [Accessed 19 June 2021]
34. Cybersecurity and Infrastructure Security Agency, (February 2021). Retrieved from <https://www.cisa.gov/tips/st04-015> [Accessed 10 June 2021]
35. Cisco(n.d.). What Is Malware? Retrieved from <https://www.cisco.com/c/en/us/products/security/advanced-malware-protection/what-is-malware.html> [Accessed 10 June 2021]
36. National Institute of Standards and Technology. (n.d.). advanced persistent threat (APT). Retrieved from [https://csrc.nist.gov/glossary/term/advanced\_persistent\_threat#](https://csrc.nist.gov/glossary/term/advanced_persistent_threat) [Accessed 10 June 2021]
37. The White House. (May 12, 2021). Executive Order on Improving the Nation’s Cybersecurity. Retrieved from <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/> [Accessed 30 June 2021]

# 

# **Appendix**

**1.a**

[**https://ocrportal.hhs.gov/ocr/breach/breach\_report.jsf**](https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf) **[Accessed 02 June 2021]**

**https://blog.pdffiller.com/life-like-hipaa-changed-healthcare-industry/**

**1.b**

*#Intalling/ importing packages*

**import** **os** *#This provide fucntion for the interaction with the operating system.*

**import** **pandas** **as** **pd**

**import** **numpy** **as** **np**

**import** **matplotlib.pyplot** **as** **plt**

**import** **seaborn** **as** **sns**

sns.set()

%**matplotlib** inline

**import** **warnings**

warnings.filterwarnings('ignore') *#this helps filter uneccesary warnings.*

**import** **scipy**

**from** **scipy** **import** stats **as** stats

**1.c**

*# Investigating the number of each Type of Breach that was reported*

data=Report["Type of Breach"].value\_counts()

data

**1.d**

*#Pie Plot Showing Type of Breach By Percentage*

pie, ax = plt.subplots(figsize=[10,6])

labels = data.keys()

plt.pie(x=data, autopct="**%.1f%%**", explode=[0.05]\*5, labels=labels, pctdistance=0.5)

plt.title("Type of Breach By Percentage", fontsize=14);

pie.savefig("dataPieChart.png")

**1.e**

*# Investigating breach location*

Report["Location of Breached Information"].value\_counts()

**1.f**

*# create the time series plot*

fig,ax=plt.subplots(figsize=(10,8))

sns.lineplot(x = "Breach Month & Year", y = "Totoal Individuals Affected",data = New\_data)

plt.title("Individuals Affected",color='red',fontsize=14)

plt.xticks(rotation = 25)

plt.show()

**1.g**

*# Bar plot showing the Type of Breaches and the Affected Individuals*

fig,ax=plt.subplots(figsize=(10,8))

ax=sns.barplot(y="Type of Breach",x="Individuals Affected",data=Report,ci=**None**,

hue="Breach Yr",palette = "Reds",estimator = np.mean)

plt.title("Types of Breach vs Number of individuals Affected in the various years")

plt.savefig("add\_annotation\_to\_bars\_in\_barplot\_Seaborn\_Python.png")

plt.show(fig)

**1.h**

*# Investigating the number breach associated with each Covered Entity Type*

data\_CET=Report["Covered Entity Type"].value\_counts()

data\_CET

**1.i**

*Using matplotlib*

pie, ax = plt.subplots(figsize=[10,6])

labels = data\_CET.keys()

plt.pie(x=data\_CET, autopct="**%.1f%%**", explode=[0.05]\*4, labels=labels, pctdistance=0.5)

plt.title("Covered Entity Breaches", fontsize=14);

pie.savefig("dataPieChart.png")

**1.j**

*# Inestigating number of Business Associates Involved*

Report["Business Associate Present"].value\_counts()

**1.k**

*#Investigating with most breaches States*

data\_ST=Report["State"].value\_counts()

data\_ST

**1.l**

*# Performing T-test*

*# We divide the newly created dataset that contains the total of infected individuals for each month into two sub datasets,*

*# Sum1 and Sum2"*

*# Sum1 = Nov 2019-Aug 2020*

*# Sum2 = Sept 2020-May 2021*

Sum1 = New\_data['Totoal Individuals Affected'].iloc[0:10]

Sum1

Sum2=New\_data['Totoal Individuals Affected'].iloc[10:19]

Sum2

*# T-test*

stats.ttest\_ind(Sum1, Sum2, equal\_var= **False**)