5-2023

HEALTHCARE DATA BREACHES: ANALYSIS AND PREVENTION

Nikita S. Dean
California State University San Bernardino

HEALTHCARE DATA BREACHES:
ANALYSIS AND PREVENTION

A Project

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Information Systems and Technology

by

Nikita S. Dean

May 2023

HEALTHCARE DATA BREACHES:

ANALYSIS AND PREVENTION

A Project

Presented to the

Faculty of

California State University,

San Bernardino

by

Nikita S. Dean

May 2023

Approved by:

Dr. Conrad Shayo, Committee Member

Dr. William Butler, Committee Chair

Dr. Conrad Shayo, Department Chair, Information and Decision Sciences

ABSTRACT

It is evident that the healthcare sector continues to experience data breaches. This culminating experience project focuses on the need to maintain patient data privacy, minimize financial risks, and address public health concerns. The study examined data collected from U.S. Department of Health and Human Services from 2018 to 2023 to answer the following research questions: Q1. How many individuals are affected due to data breaches in healthcare & which States had the most affected individuals? Q2. What are the most common causes of healthcare data breaches & what measures can be taken to prevent this? and Q4. What are the healthcare data breach predictions for 2023 & what suggestions could be given to minimize them? The findings are: Q1. Within the last five years, several individuals have been affected by healthcare data breaches per year. The States where individuals are affected highest are California, Florida, Minnesota, and North Carolina; Q2. The most common type of breaches are hacking/IT incidents. The compromised device/ location that affected the most individuals was the network server ; Q4. The prediction for 2023 is that although breaches did decrease in 2022, they will continue to rise again in 2023. The conclusion for each question is: Q1. The year 2019 had the highest peak, this is the year Covid-19 occurred, Landi (2020) also discussed that in 2019 phishing attacks plagued healthcare. Proper training for staff in identifying phishing attacks could help combat and reduce the amount of individuals affected each year. Q2. Home networks are not always secure. Work-from-home devices need extra precautions, such as an IT team employing automatic updates overnight, or creating a virtual environment to access confidential information and work within the virtual environment. And Q4. Cybercriminals will continue to evolve their methods and breach industries. Having a healthcare business association (advisors, experienced and supportive references) present can reduce the amount of individuals being affected. Further research would be to include data breaches affecting less than 500 individuals. More research would also be recommended to find methods of detecting and blocking data breaches faster.

ACKNOWLEDGEMENTS

I would like to thank Dr. Conrad Shayo and Dr. William Butler for guiding me towards the proper direction in the writing of this Culminating Experience Project. Their suggestions and recommendations helped tremendously in making this experience possible. I would also like to thank my family for helping me stay motivated throughout this journey. And lastly, I would like to thank Dr. Pamela Abell who met with me via zoom on my least productive days and helped me with writing strategies, staying focused, and managing a busy schedule.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 4: Data Analysis and Findings

CHAPTER ONE:

INTRODUCTION

Data breaches are becoming sophisticated and plentiful when it comes to the healthcare sector. This has started happening because, as stated by Seh (2020), “Advances in information and communication technology have helped the healthcare industry to replace paper-based systems with electronic health record (EHRs) systems to provide better and more cost-effective services to its customers” (p. 1). With electronic data storage, cybercriminals are attempting to acquire these records. Data breaches within the healthcare industry happen because “Compromised health records can be especially profitable to criminals who seek to not only exploit social security numbers for financial gain but use health insurance policies to file fraudulent claims and write counterfeit prescriptions” (Collins, 2011, p. 4). Electronic health records have exposed these threats, along with human errors.

Although cybercriminals play a large part within data compromising, another factor is the poor human security. “Investigators concluded that human factors, and implementation of security policies were strongly related to an increase in data breaches” (McLeod, 2018, p. 60). With human errors, we have unintentional human factors, and malicious human factors. Yeo (2022) concluded that “26 percent of all human factor-based breaches were due to an insider's carelessness, negligence, or apathy” (p. 9). Along with poor human security, healthcare is also faced with improper physical security.

Improper physical security can lead to mishaps such as theft. Theft can be another reason for data breaches. Elahi (2020) presented a table with breach types affecting different entities and the number of individuals affected. Under healthcare providers, 135,041 individuals were affected by data theft in 2020 (p. 21). Despite the availability of several health information regulations (Chernyshev, 2019), e.g., HIPAA 1996, Data Protection Bill 2017, GDPR 2018, and Privacy Act 1988; data breaches continue unabated. The main causes of data breaches are technological advancements, human error, and theft of physical devices.

Research Background

For the purposes of finding areas for further study in healthcare data breaches, chatGPT3 (AI (Artificial Intelligence)) suggested areas such as: Prevention strategies, impact on patients, detection and response, legal and regulatory frameworks, cybersecurity risk management, and cybersecurity workplace. Yeo & Banfield (2022) presented a study for data breaches resulting from human factors such as carelessness/ negligence, phishing/ ransomware, malicious insiders and other sources of breach and some prevention strategies. They said their analysis of the data breaches provided information in specifying directions for future research and areas to focus on mitigating cyber-attacks.

Seh et al., (2020) also suggested future studies in identifying preventive measures that should be taken to avoid healthcare data breaches. The authors discussed several areas for further studies such as identifying and addressing the main victims of cyber-attacks in the healthcare sector, where the impact on patients can be included. Khan et al., (2021) presented a conceptual framework and integrated risk model for data breach management in their study. They suggested that future research can refine the proposed integrated risk model. This would go hand in hand with detection and response, legal and regulatory frameworks, and cybersecurity risk management. They also stated that further studies should investigate specific industries and firms since they used an overall approach. Lastly, Nyakasoka, L., & Naidoo, R. (2022) presented areas for future research that addresses the development of dynamic cybersecurity capabilities in healthcare settings. This would address the cybersecurity workplace.

Problem Statement

Given the areas chatGPT3 and other researchers have suggested for further studies mentioned above such as prevention strategies, impact on patients, detection and response, legal and regulatory frameworks, cybersecurity risk management, and cybersecurity workplace. The outstanding research questions driven from these recommends are:

1. How many individuals are affected due to data breaches in healthcare & which States had the most affected individuals?
2. What are the most common causes of healthcare data breaches & what measures can be taken to prevent this?
3. What are some ways to better detect and respond to healthcare data breaches?
4. What are the healthcare data breach predictions for 2023 & what suggestions could be given to minimize them?
5. What are best practices for cybersecurity risk management in healthcare?
6. What are the skills and training needed for healthcare? How can they retain cybersecurity professionals in healthcare?

This culminating experience project will use the dataset chosen between 2018 to 2022 to answer questions #1, #2, and #4.

Organization of the Study

This culminating experience project is organized as follows: Chapter one provided an introduction, problem statement, and research questions. Chapter two will provide the literature review. Chapter three will consist of research methodology. Chapter four will contain the analysis of the data and the findings. Chapter five will provide the discussion, conclusion, and areas for further study.

CHAPTER TWO:

LITERATURE REVIEW

The literature presented in this project was found on Google Scholar, OneSearch, and Statista. The search terms used were: Healthcare data breaches, organization data breaches, and data breach management. In Google Scholar and OneSearch there were 46 relevant articles relating to healthcare data breaches, 9 were chosen for this project. In Statista, there were 96 relevant results relating to healthcare data breaches, 3 were selected. The ones selected were articles and datasets that are free to the public or accessed through the university.

This culminating experience project employs a content analytics-based research method. This is described by Kovanović as “a particular form of learning analytics focused on the analysis of different forms of educational content” (Kovanovic, 2017, p. 77). A combination of quantitative and qualitative content analysis has several advantages. These could be areas such as finding correlation and pattern. Both methods will be used: Quantitative to find focused results using the dataset (numbers), and qualitative to find areas that cannot easily be put into numbers to understand the human experience.

To analyze the main cause of data breaches in healthcare, we need to define what data breaches are. Data Breaches: Incident where information is stolen or obtained without the systems’ knowledge. Breaches expose confidential, sensitive, or personal information to criminals. “Majority of breaches are financially motivated; however, attacks do not necessarily target the richest or best-known organisms, but those that are not prepared to attack - like medical organisms” (Hammouchi, 2019, p. 3).

Before this project investigates a dataset, it is good practice to understand why healthcare data breaches occur. Healthcare Industry is a target according to Bhosale (2021) because of the following reasons:

- Contains patient information
- Medical devices are easier to hack
- The need for remote access to data, presenting more vulnerability
- Workers not wanting to disrupt convenient working methods with modern technologies
- Healthcare staff lacking training in online risks
- The number of devices within hospitals are too vast for security to stay on top
- Healthcare information needs to be open and shareable
- Smaller healthcare organizations are also at risk

Table 1. Summary of Database Search of Relevant Publications

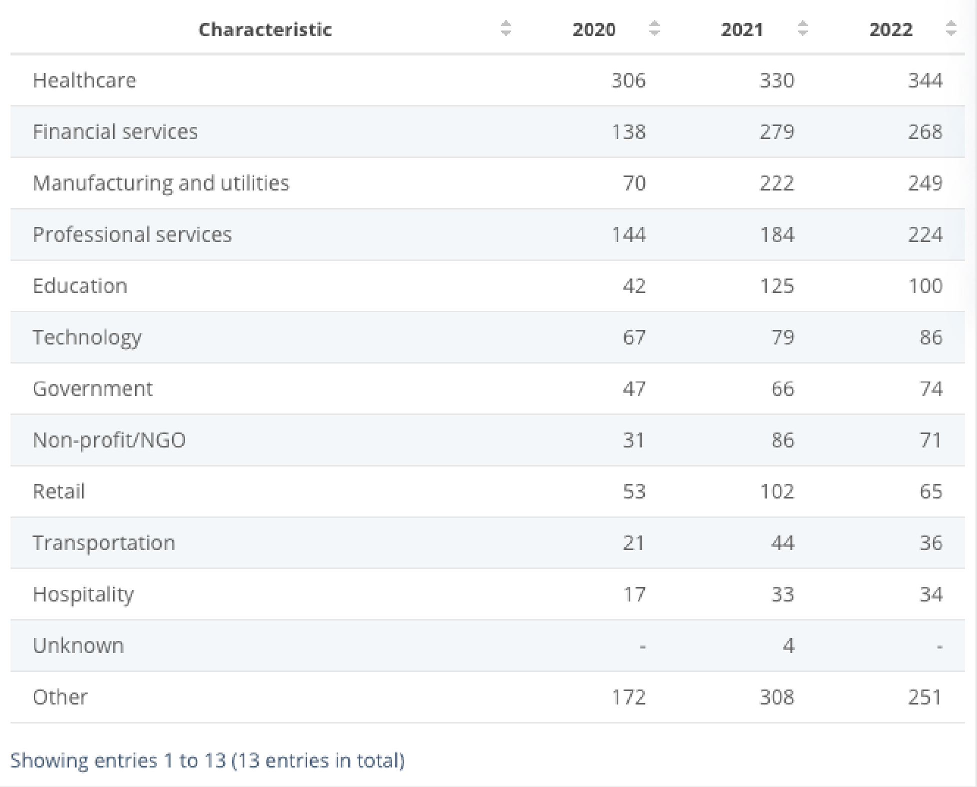
Database	Category	Number of Relevant Publications	Number of Relevant Publications Selected	Authors
Google Scholar; OneSearch	Healthcare Data Breaches	17	4	McLeod, A., & Dolezel, D., 2018. Seh, A., et al. 2020. Yeo, L. & Banfield, J., 2022. Nyakasoka, L., & Naidoo, R., 2022.
Google Scholar; OneSearch	Organization Data Breaches	7	2	Collins, J., et al., 2011. Dolezel, D., & McLeod, A., 2019.
Google Scholar; OneSearch	Data Breach Management	22	3	F. Khan et al., 2021. Lee, J., & Choi, S. J., 2021. Cheng, L., et al., 2017
Statista.com	Healthcare Data Breach	96	3	Ani Petrosyan, 2022.

The first question to be answered is: "*How many individuals are affected due to data breaches in healthcare & which States had the most affected individuals?*" The literature used to discuss this proposed question will be by Lee, J., & Choi, S. J. (2021), and Petrosyan, A. (2022).

Lee, J., & Choi, S. J. (2021) aimed to investigate the association between hospital data breaches and productivity by using a generalized difference-in-differences model with multiple prebreach and post breach periods. They sampled 2610 unique hospital-year observations, including general acute care hospitals.

Petrosyan, A. (2023) stated that the healthcare industry has been the most targeted sector by cyber-attacks resulting in data compromises. The healthcare industry has had more than 300 breaches consisting of data breaches, data expose, and data leaks. This is shown in the table below.

Table 2. Number of cases of data violation due to cyber-attacks in the United States from 2020 to 2022, by industry (Ani Petrosyan, 2023)



Characteristic	2020	2021	2022
Healthcare	306	330	344
Financial services	138	279	268
Manufacturing and utilities	70	222	249
Professional services	144	184	224
Education	42	125	100
Technology	67	79	86
Government	47	66	74
Non-profit/NGO	31	86	71
Retail	53	102	65
Transportation	21	44	36
Hospitality	17	33	34
Unknown	-	4	-
Other	172	308	251

Showing entries 1 to 13 (13 entries in total)

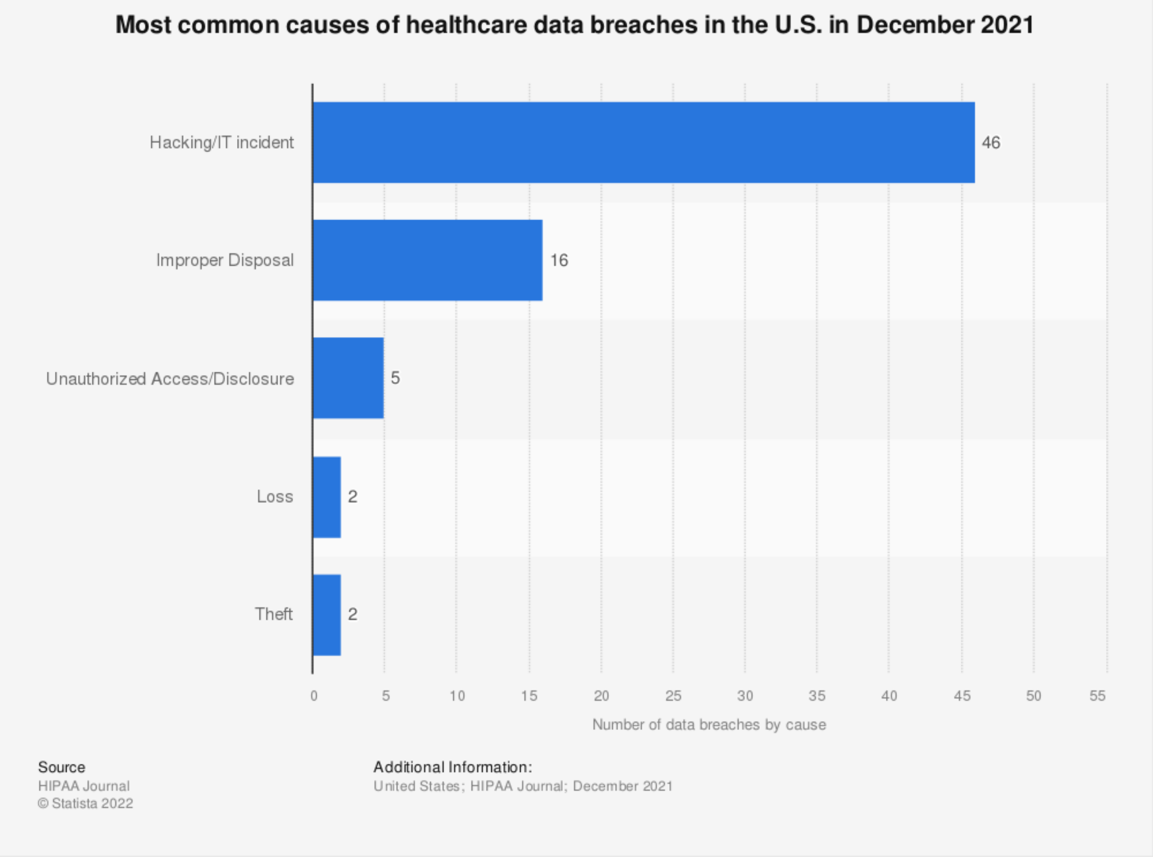
The second question that needs to be answered is: “*What are the most common causes of healthcare data breaches & what measures can be taken to prevent this?*” The literature used to discuss this proposed question will be by Yeo, L. H., & Banfield, J. (2022), Nyakasoka, L., & Naidoo, R. (2022), Khan, F., et al. (2021), Petrosyan, A. (2022), Cheng, L., et al. (2017), and Dolezel, D., & McLeod, A. (2019).

Yeo, L. H., & Banfield, J. (2022) studied and concluded that technical safeguards that protect health records need to be combined with human behavioral interventions to promote a robust cybersecurity plan. Their findings concluded that most data breaches were a result of unintentional insider threats - which was twice the amount of data breaches due to malicious intent. Their findings suggested that data breaches were occurring due to phishing frauds and that proper prevention would need to go beyond technical controls.

Nyakasoka, L., & Naidoo, R. (2022) presented a study titled Barriers to Dynamic Cybersecurity Capabilities in Healthcare Software Services. They studied areas such as inadequate user education, evolving threat landscape, and inability to attract or retain skilled cybersecurity staff. Their research recommended interventions to address these barriers.

Khan, F., et al. (2021) presented an integrated risk model for data breach management based on a systematic review of the literature. They extended on identifying and updating risks, resolutions, and provided a foundation for organizational response to data breach incidents. They presented a conceptual framework that identified three resolution categories for managing data breach risks: prevention, containment, and recovery. Adapted from this concept, Khan (2021) presents the actors involved as: Intentional data breach actors, or unintentional data breach actors. The intentional data breach actors could be a result of hackers, unauthorized access, malicious insiders, state-sponsored actors, or terrorists. The unintentional data breach could be a result of unsecure user behavior, loss or reuse of media devices, flawed software, unauthorized discloser, or unauthorized software.

Figure 1. Most common causes of healthcare data breaches in the U.S in December 2021 (Ani Petrosyan, 2022).



This figure shares the common causes of healthcare data breaches in the U.S. at a certain period. Hacking and IT incidents are the highest causes. The lowest number of causes are loss and theft; however, they do still exist and should be investigated. When it comes to training employees, it is critical to also disclose the causes of healthcare data breaches.

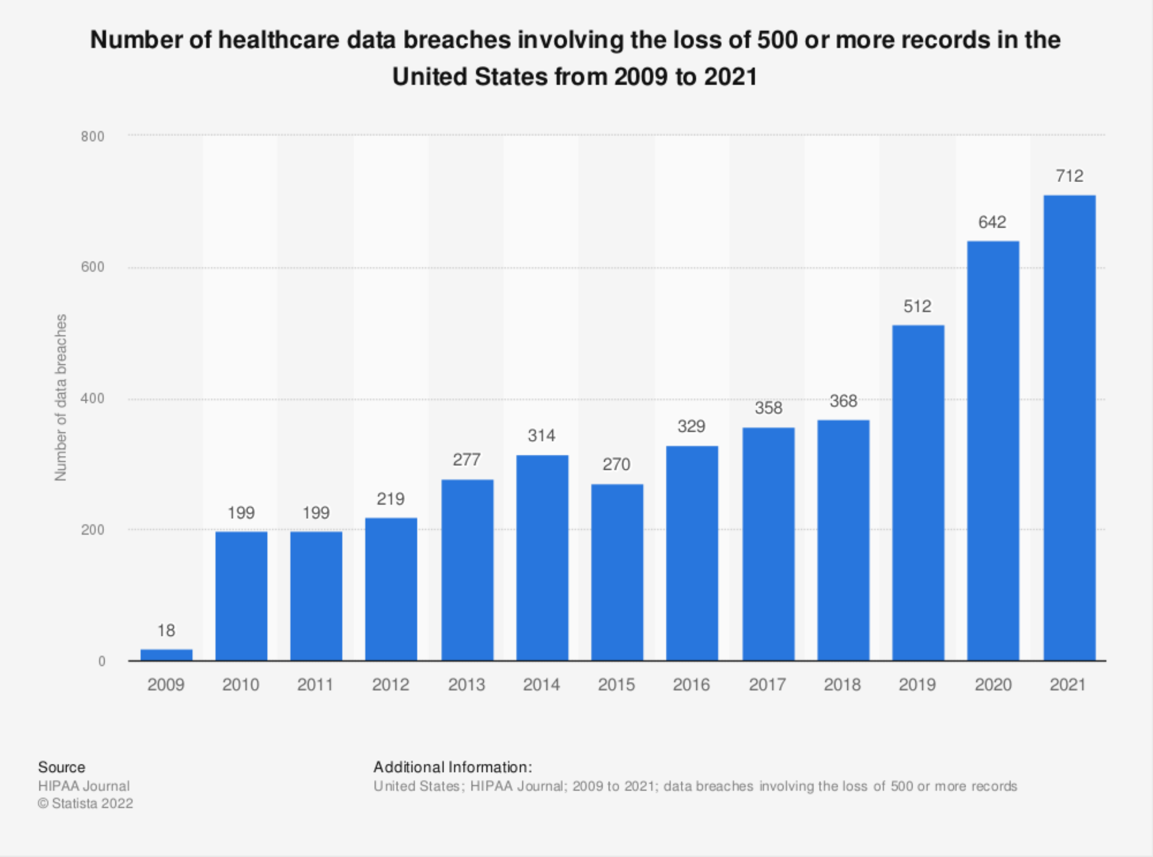
Cheng, L., et al. (2017) studied that data leakage posed serious threats to organizations such as reputation damage and financial loss. They shared the importance of detecting and preventing. Their review shared data leak threats, and prevention/ detection techniques.

Dolezel, D., & McLeod, A. (2019) worked on developing a model of factors associated with healthcare data breaches such as identifying discriminants of these breaches. They obtained data from the Department of Health and Human Services database of healthcare facilities reporting data breaches. They used a binary logistic regression to examine a representative data breach model.

The third question that needs to be answered is: "*What are the healthcare data breach predictions for 2023 & what suggestions could be given to minimize them?*" The literature used to discuss this proposed question will be by Seh et al. (2020), Petrosyan, A. (2023), Collins, J. et al. (2011), and McLeod, A., & Dolezel, D. (2018).

Seh, A., et al. (2020) worked on providing insights into the various categories of data breaches. Their main objective was to do an in-depth analysis of healthcare data breaches and draw inferences from them – and ultimately improve healthcare data confidentiality. They used two methods: the simple moving average method and the simple exponential soothing method of time series analysis to examine the trends of healthcare data breaches and their cost. They concluded that the simple moving average method provided more reliable forecasting results.

Figure 2. Number of healthcare data breaches involving the loss of 500 or more records in the United States from 2009 to 2021 (Ani Petrosyan, 2023).



Number of healthcare data breaches involving the loss of 500 or more records in the United States from 2009 to 2021

Source
HIPAA Journal
© Statista 2022

Additional Information:
United States; HIPAA Journal; 2009 to 2021; data breaches involving the loss of 500 or more records

One notices that the number of healthcare data breaches are increasing over the years. In 2018, there were 368 breaches, whereas in 2019 that number jumped to 512. It goes up to 712 breaches. “According to data published in January 2022, healthcare organizations in the United States saw the highest number of large-scale data breaches (resulting in the loss of over 500 records) to date in 2021” (Petrosyan, 2022). Data breaches within healthcare are growing rapidly. This rise in the number of data breaches is the reason prevention measures should be discussed.

Collins, J., et al. (2011) worked on organizational data breaches by reviewing the applicability of Situational Crime Prevention. Their analysis specifically addresses the variables: type of breach, reporting entity, year the breach was disclosed, and the geographic region in the United States. They concluded that the lack of a centralized reporting database for all data breaches prevents a definitive analysis of the field. They also concluded that situational crime prevention measures can be proactive in preventing future data breaches within healthcare entities.

McLeod, A., & Dolezel, D. (2018) considered organizational factors, business proves exposure factors, and technological security factors on occurrences where a breach may have taken place. They examined elements associated with data breaches and created a model of healthcare organizations experiencing data breaches. Their aim was to model factors that may be predictive of healthcare data breach weaknesses. They developed a testable healthcare data breach model.

This literature review studied the research done in this subject. Lee (2021) used the difference-in-differences model to find the association between hospital data breaches and productivity. Yeo (2022) studied the need for combining human behavioral interventions with technical safeguards to promote a robust cybersecurity plan. Nyakasoka (2022) addressed barriers that need to be eliminated to keep healthcare safer. Khan (2021) showed us an integrated risk model for data breach management. Cheng (2017) studies the effects of data leakage creating reputation and financial damage. Dolezel (2019) used binary logistic regression to examine their data breach. Seh (2020) presented us with two methods to improve healthcare data confidentiality. Collins (2011) addressed variables such as type of breach, reporting entity, year disclosed, and geographic region in the United States. McLeod (2018) developed a testable healthcare data breach model.

Lastly, it explored the visual content provided by Petrosyan (2022-2023). First, it showed the number of cases of data violation due to cyber-attacks where healthcare was one of the highest categories affected. The latter figures showed an increase in data breaches within healthcare. Lastly, we saw the most common causes of healthcare data breaches. Chapter 3, next, will provide the research methods used to answer the three selected research questions.

CHAPTER THREE:

RESEARCH METHODOLOGY

To answer the research questions, both studies and published dataset available on the U.S. Department of Health and Human Services Office for Civil Rights will be used. These questions will focus on the most recent breaches that occurred in healthcare, from the years 2018 to 2022.

To answer the question: *How many individuals are affected due to data breaches in healthcare & which States had the most affected individuals?* This project will create graphs from a dataset to visualize how many individuals are affected due to healthcare data breaches. It will also examine locations in the United States where the highest number of individuals are affected.

To answer the question: *What are the most common causes of healthcare data breaches & what measures can be taken to prevent this?* The project will investigate intentional and unintentional data breaches. Data breach could be intentional and unintentional. Khan (2021) explains these two as “Intentional data breaches are incidents caused by malicious acts in which one or more humans or technology threat agents exploit vulnerabilities to cause harm to an organization” (p. 4). & “Unintentional data breaches are accidental incidents caused by individuals or processes not acting with malicious intent” (p. 4). To answer the questions posed, this project will need to use a dataset and create graphs to determine the type of breach. This information will enable the study to determine prevention measures needed in determined specific areas. By learning where a data breach is targeted, it can help with navigating prevention measures.

To answer the question: *What are the healthcare data breach predictions for 2023 & what suggestions could be given to minimize them?* Using the dataset, it will be possible to create predictive graphs to see how many individuals may be affected in 2023. These visuals will also help with finding suggestions to minimize the number of individuals affected. Creating these graphs will help guide this focus towards the questions.

Dataset

Dataset Sources and Description

There were two different datasets from Kaggle.com, it was determined that both are relevant to this culminating experience project. The first dataset was the *Major US Health Data Breaches*. This dataset has detailed information on US health data breaches that affected at least 500 individuals. The data stands to be valuable for those interested in learning more about health care security best practices and prevention against future data-related incidents. However, this dataset reported breach submission dates from 10/09 - 10/16. The second dataset was the *HIPAA Breach Report* and is similarly reporting data breaches that affected at least 500 individuals. This dataset reported breach submission dates from 10/09 – 09/17.

These datasets come from direct reports to the breach portal in the U.S. Department of Health and Human Services Office for Civil Rights. Upon some consideration, it was decided to select data from this government website, the resolved breach reports. This helps researchers gain data from any breach submission date of their choosing, in turn, data selected for this project will be from 01/18 to 12/22. This gave 2218 results.

The table shown below will describe how the dataset collected will be presented. It will identify eight columns that will carry the data such as which company reported their data breach, the state they are in, individuals affected, location of breached information, among other areas covered in the table.

Table 3. Dataset columns and descriptions

Column name	Description
Name of Cover Entity	The name of the entity that experienced the data breach. (String)
State	The geographical state location of the entity that experienced the data breach. (String)
Covered Entity Type	The type of entity that experienced the data breach, such as hospitals or insurers. (String)
Individuals Affected	The amount of people affected by a data breach (Number)
Breach Submission Date	The date the breach was reported. (Date)
Type of Breach	The type of breach that occurred, such as unauthorized access/disclosure or loss of hardware/electronic media. (String)
Location of Breached Information	The location of the breached information, such as paper records or electronic devices. (String)
Business Associate Present	Whether a business associate was present during the breach event. (Yes/ No)

Data Cleaning

Ideally, data cleaning is performed when data is missing, if there are mismatched columns, or if there are duplicates. Upon reviewing the data obtained from the breach portal in the U.S. Department of Health and Human Services Office for Civil Rights, these issues were not present. This data quality does not present tangible human errors, no technical problems arose when the data was transferred. However, the entire column for name of covered entity was removed since the data analysis will not focus on which specific entity faced the breach.

Steps Involved in this Dataset Analysis

1. Get a dataset ready for use
2. Choose correct algorithm
3. Model
4. Evaluate model
5. Improve model
6. Save and load.

Tools for Visualization

Batt (2020) stated that “the five goals of student research: (1) develop a worthwhile research question; (2) connect it to the literature; (3) formulate a testable hypothesis; (4) collect, clean, and use appropriate tools to analyze data to test the hypothesis; and (5) present the findings” (p. 318). This can be done using several methods. This project will use graphs to present the dataset. “Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data” (Tableau.com 2019, online). The tools used will be Tableau and Excel.

CHAPTER FOUR:

DATA ANALYSIS AND FINDINGS

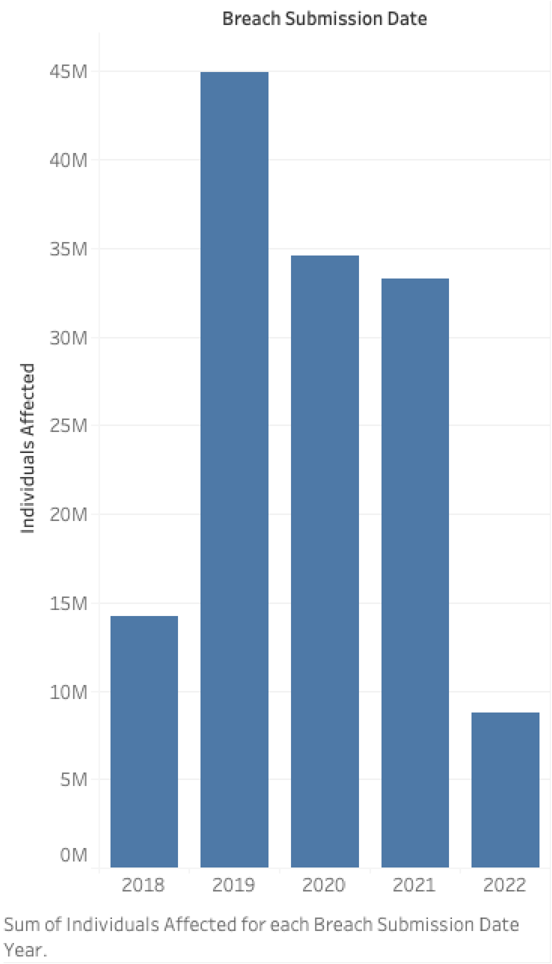
Chapter 4 provides the data analysis and findings for Questions #1, #2 and #4. Q1 will be answered using Figures 4.1 through 4.4. Q2 will be answered using Figures 4.5 through 4.8. Lastly, Q4 will be answered using Figures 4.9 and 4.10. The visuals organization for this chapter will be as follows: Year of breach and individuals affected, grand total of individuals affected, location within United States, type of breach (i.e., hacking, improper disposal, loss, theft, unauthorized access), number of individuals affected based on type of breach per year, compromised devices (i.e., laptop, network server, etc.), entity types, predictions for 2023, and whether they had business association or not. In this chapter, the focus will be on describing the analysis and findings only. Chapter 5 will focus on the discussion of the findings, conclusions, and provide areas for further study.

Data Analysis and Visualization

Figures 4.1 – 4.4 will answer question 1.

How many individuals are affected due to data breaches in healthcare & which States had the most affected individuals?

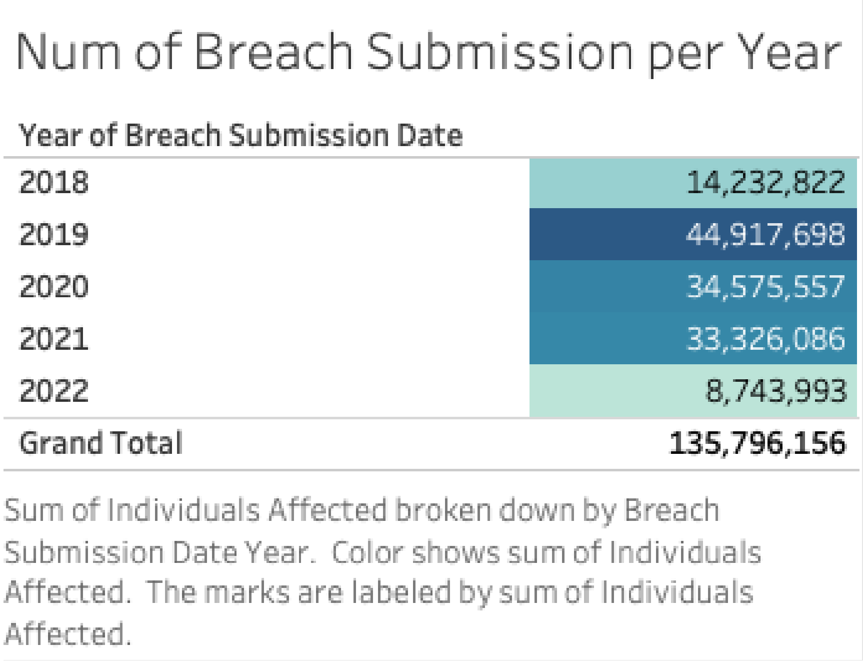
Figure 4. 1 Year of Breach Vs. Number Affected



Sum of Individuals Affected for each Breach Submission Date Year.

Figure 4.1 shows the results of a comparison of the Breach Submission Year and the Number of Individuals Affected. The findings are that the years 2019, 2020, and 2021 had the highest number of individuals affected.

Figure 4. 2 Specific Number and Grand Total



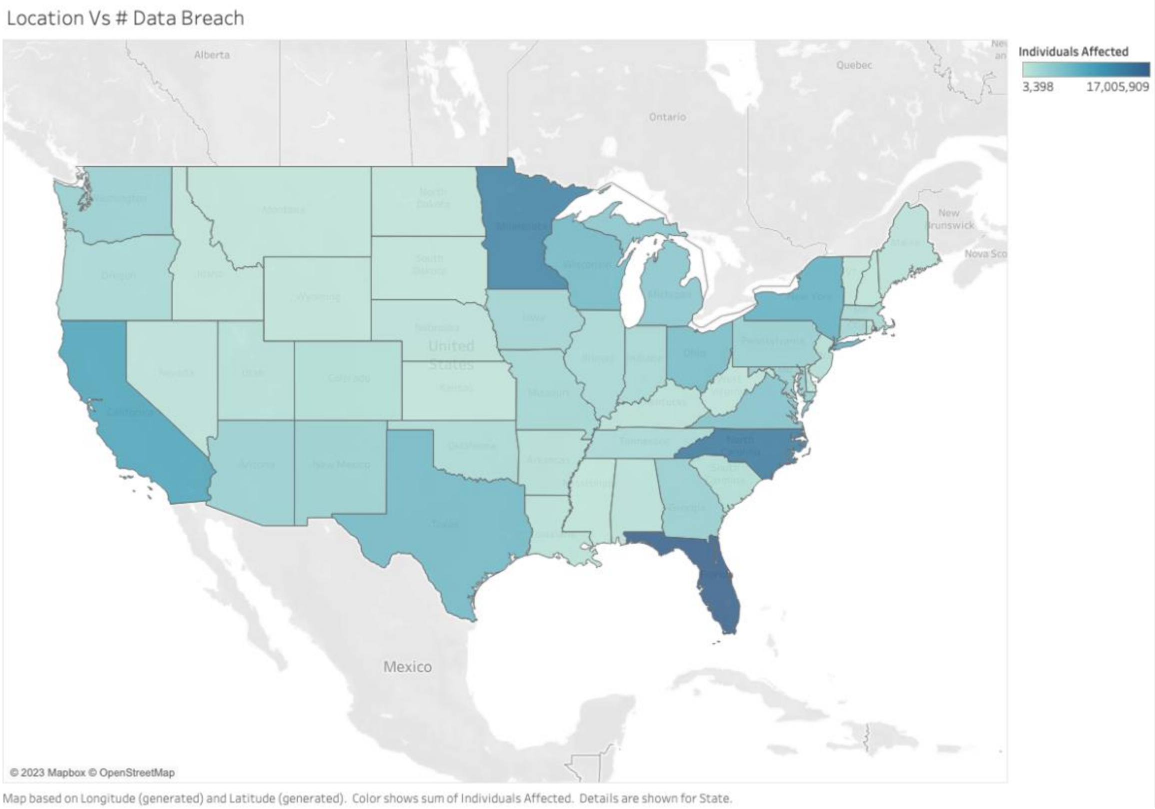
Num of Breach Submission per Year

Year of Breach Submission Date	
2018	14,232,822
2019	44,917,698
2020	34,575,557
2021	33,326,086
2022	8,743,993
Grand Total	**135,796,156**

Sum of Individuals Affected broken down by Breach Submission Date Year. Color shows sum of Individuals Affected. The marks are labeled by sum of Individuals Affected.

Figure 4.2 notes the specific number of individuals affected based on the years. The total number individuals affected from 2018 to 2022 was 135.8 million. The highest number of people affected occurred in 2019 (slightly above 44.9 million) followed by 2020 and 2021 (combined average slightly above 33.8 million). The lowest number of people affected occurred in 2020 (slightly above 8.7 million).

Figure 4. 3 Number Affected by Location



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Individuals Affected. Details are shown for State.

Figure 4.3 shows that the four States in the US with the highest number of affected individuals are California, Florida, Minnesota, and North Carolina.

Figure 4. 4 States with Highest Number of Breaches

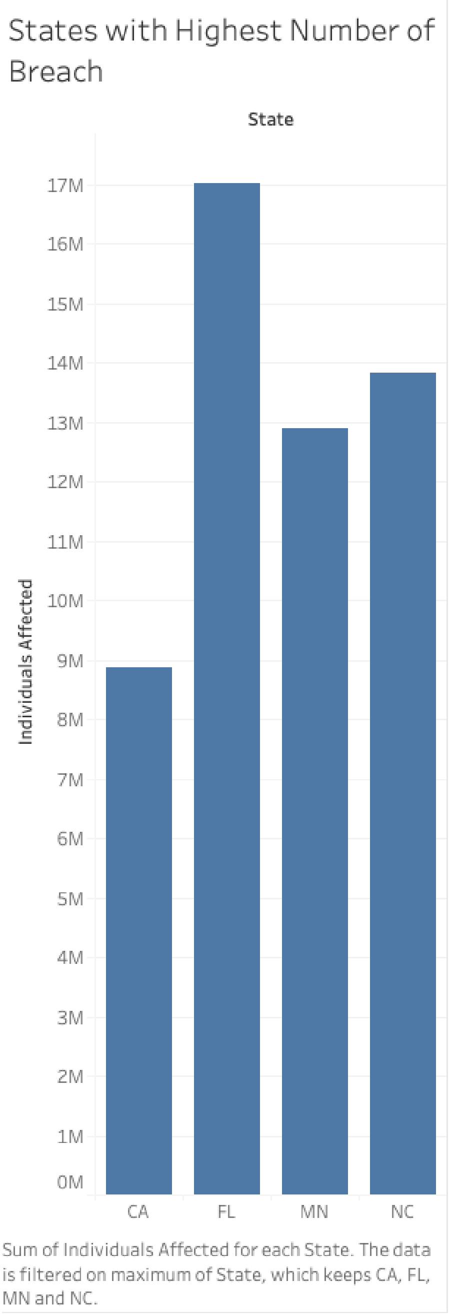


Figure 4.4 confirms that the four States in the US with the highest number of affected individuals are California, Florida, Minnesota, and North Carolina in a bar graph.

Findings:

It was determined in Chapter 4 through Figure 4.1 individuals affected by data breaches within healthcare rose significantly from 2018 to 2019. Figure 4.2 shared the specific number of breaches per year. The total number of individuals affected in 2018 was 14.2 million, in 2019 was 44.9 million, in 2020 was 34.5 million, in 2021 was 33.3 million, and in 2022 was 8.7 million. Figure 4.3 shows the four states where patients were most affected, these were California, Florida, Minnesota, and North Carolina. Figure 4.4 confirmed this with exact numbers – such as in California nearly 9 million individuals were affected, in Florida 17 million individuals were affected, in Minnesota 12.9 million were affected, and in North Carolina about 13.8 million were affected.

Figures 4.5 – 4.8 will answer question 2.

What are the most common causes of healthcare data breaches & what measures can be taken to prevent this?

Figure 4. 5 Type of Breach

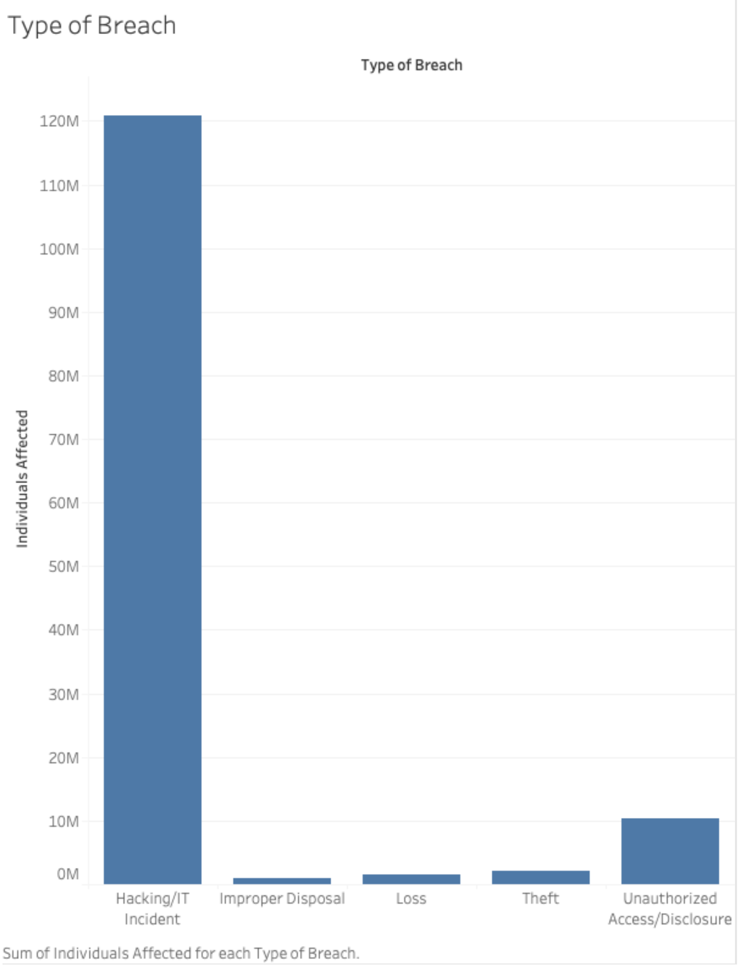
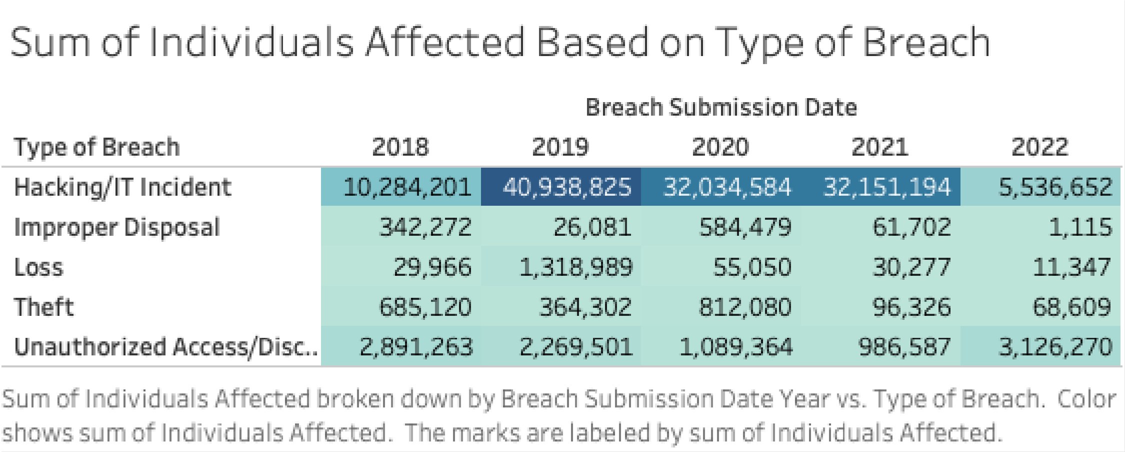


Figure 4.5 shows that the data breaches were occurring in diverse ways such as: hacking/ IT incidents, improper disposal, loss, theft, or unauthorized access/disclosure. It was determined that hacking/ IT incidents were the highest with over 120 million individuals affected.

Figure 4. 6 Number of Individuals Affected per year based on type of breach



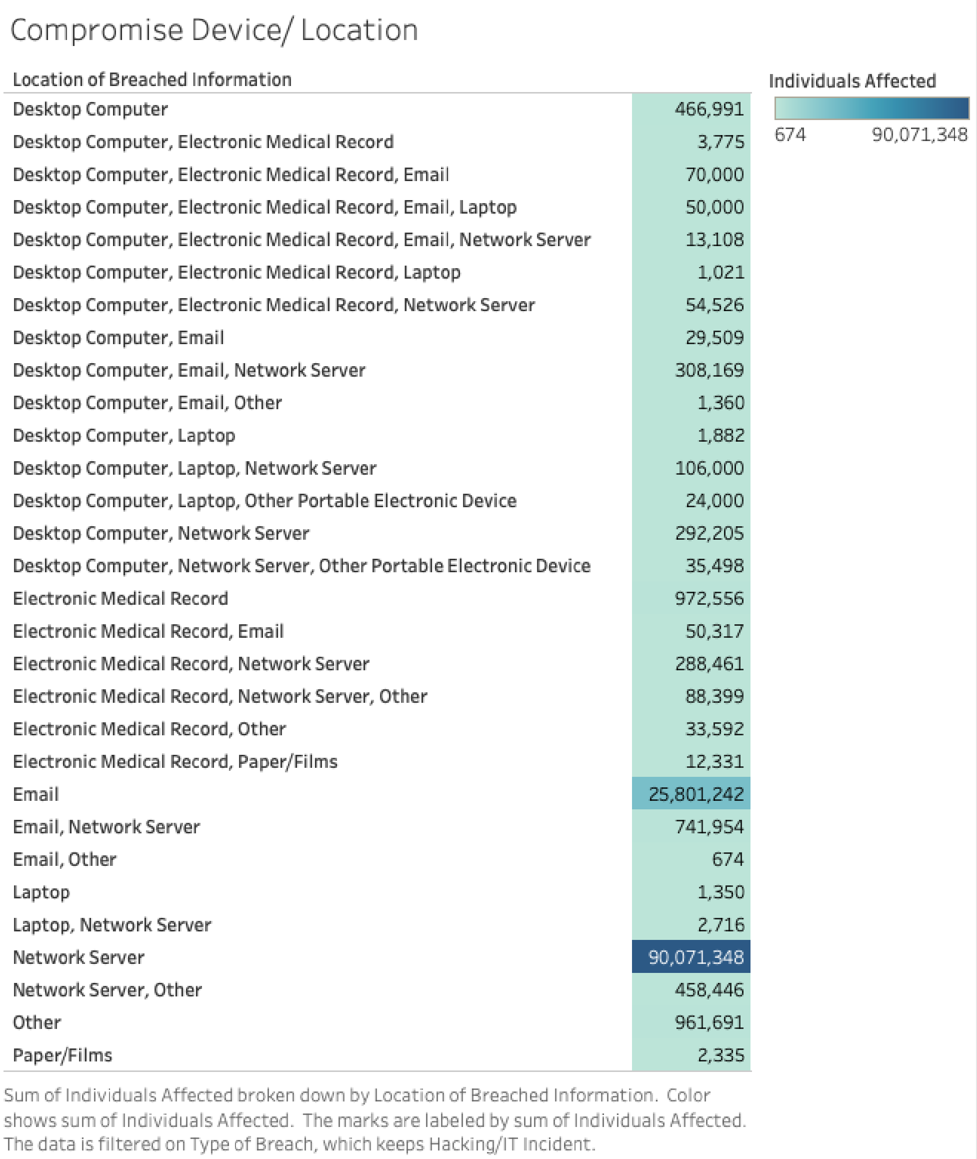
Sum of Individuals Affected Based on Type of Breach

	Breach Submission Date				
Type of Breach	2018	2019	2020	2021	2022
Hacking/IT Incident	10,284,201	40,938,825	32,034,584	32,151,194	5,536,652
Improper Disposal	342,272	26,081	584,479	61,702	1,115
Loss	29,966	1,318,989	55,050	30,277	11,347
Theft	685,120	364,302	812,080	96,326	68,609
Unauthorized Access/Disc..	2,891,263	2,269,501	1,089,364	986,587	3,126,270

Sum of Individuals Affected broken down by Breach Submission Date Year vs. Type of Breach. Color shows sum of Individuals Affected. The marks are labeled by sum of Individuals Affected.

Figure 4.6 shows the number of individuals affected per year due to breaches occurring from hacking, improper disposal, loss, theft, or unauthorized access. It is evident that the number of individuals affected by unauthorized access in 2022 increased drastically compared to the previous year.

Figure 4. 7 Where the Breach Happened



Compromise Device/ Location

Location of Breached Information	
Desktop Computer	466,991
Desktop Computer, Electronic Medical Record	3,775
Desktop Computer, Electronic Medical Record, Email	70,000
Desktop Computer, Electronic Medical Record, Email, Laptop	50,000
Desktop Computer, Electronic Medical Record, Email, Network Server	13,108
Desktop Computer, Electronic Medical Record, Laptop	1,021
Desktop Computer, Electronic Medical Record, Network Server	54,526
Desktop Computer, Email	29,509
Desktop Computer, Email, Network Server	308,169
Desktop Computer, Email, Other	1,360
Desktop Computer, Laptop	1,882
Desktop Computer, Laptop, Network Server	106,000
Desktop Computer, Laptop, Other Portable Electronic Device	24,000
Desktop Computer, Network Server	292,205
Desktop Computer, Network Server, Other Portable Electronic Device	35,498
Electronic Medical Record	972,556
Electronic Medical Record, Email	50,317
Electronic Medical Record, Network Server	288,461
Electronic Medical Record, Network Server, Other	88,399
Electronic Medical Record, Other	33,592
Electronic Medical Record, Paper/Films	12,331
Email	25,801,242
Email, Network Server	741,954
Email, Other	674
Laptop	1,350
Laptop, Network Server	2,716
Network Server	90,071,348
Network Server, Other	458,446
Other	961,691
Paper/Films	2,335

Sum of Individuals Affected broken down by Location of Breached Information. Color shows sum of Individuals Affected. The marks are labeled by sum of Individuals Affected. The data is filtered on Type of Breach, which keeps Hacking/IT Incident.

Figure 4.7 shows that there were several locations where the breach occurred such as: desktop, electronic medical record, email, laptop, network server, or paper. It was determined that within the last five years, the highest breaches occurred in the network server.

Figure 4. 8 Entity Type

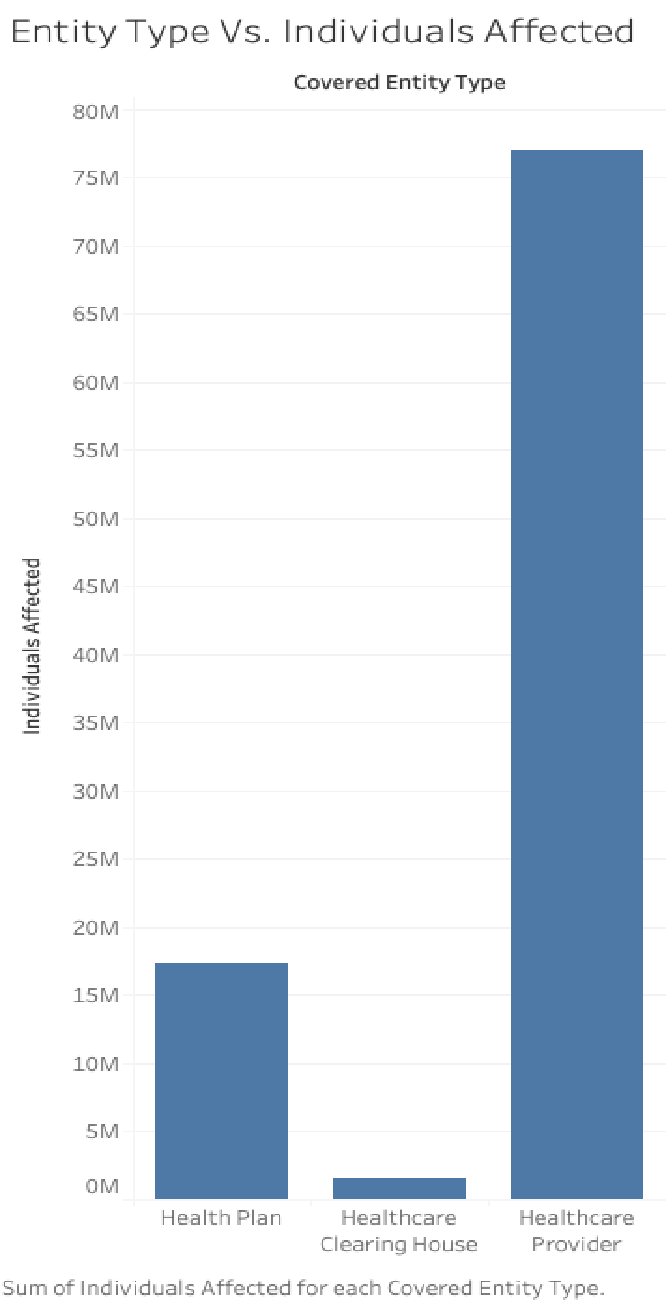


Figure 4.8 shows the number of individuals affected based on entity type (health plan, healthcare clearing house, or healthcare provider) mattered, with healthcare provider data breaches affecting up to 72 million individuals.

Findings:

Figure 4.5 shows the type of breach that occurred such as hacking/ IT incident, improper disposal, loss, theft, and unauthorized access/ disclosure. The highest number of individuals affected was a result of hacking and or IT incidents with nearly 120 million within the past five years. Figure 4.6 shows which specific type of breach was affecting how many individuals by year. In Figure 4.8 the entity types are shown with healthcare providers resulting in the highest number of individuals affected at 76 million. Figure 4.7 shows the compromised device such as laptops, network servers, emails, or paper/films. Here it was discovered that paper/films had the lowest number of breaches occurred, whereas network server breaches resulted in the highest number of individuals affected at 90 million. Some prevention measures for breaches within the network servers are role-based access controls, robust password policy, regular updates, secure routers, employee training, firewalls, encryption, antivirus, and continuous auditing.

Figures 4.9 and 4.10 will answer question 4.

What are the healthcare data breach predictions for 2023 & what suggestions could be given to minimize them?

Figure 4. 9 Prediction of Individuals that may be affected in 2023

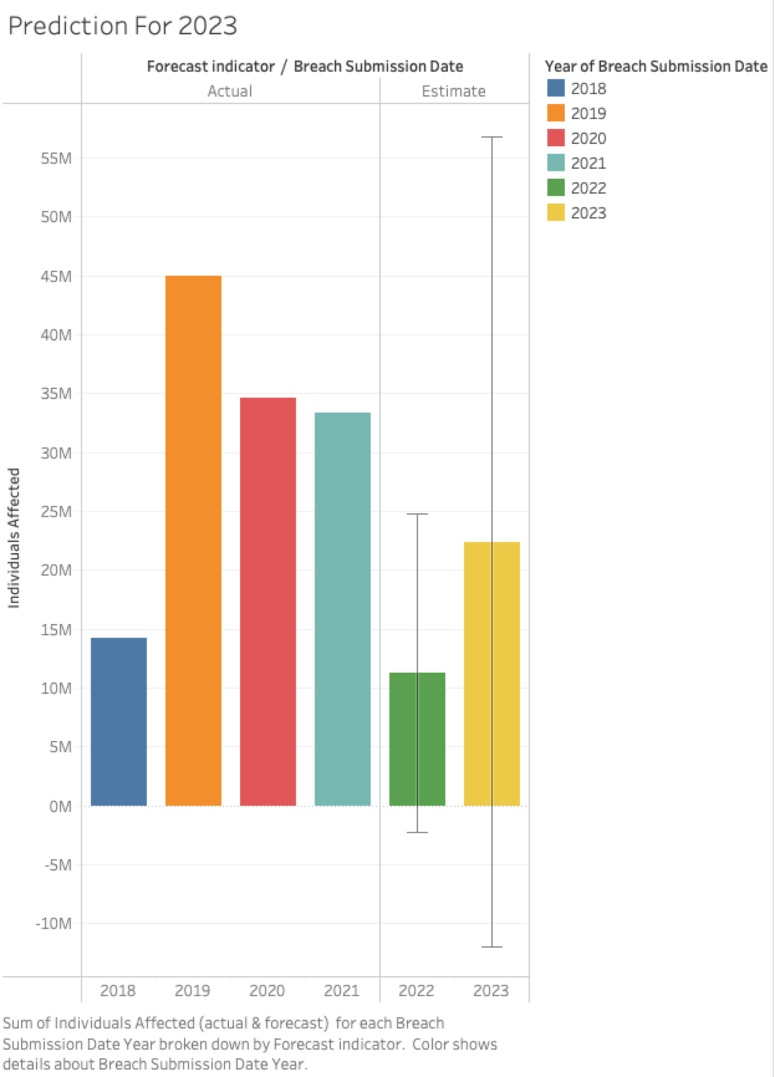


Figure 4.9 calculates the predicted number of individuals that would be affected. This calculation was driven from data from previous years.

Figure 4. 10 Whether they had Business Association

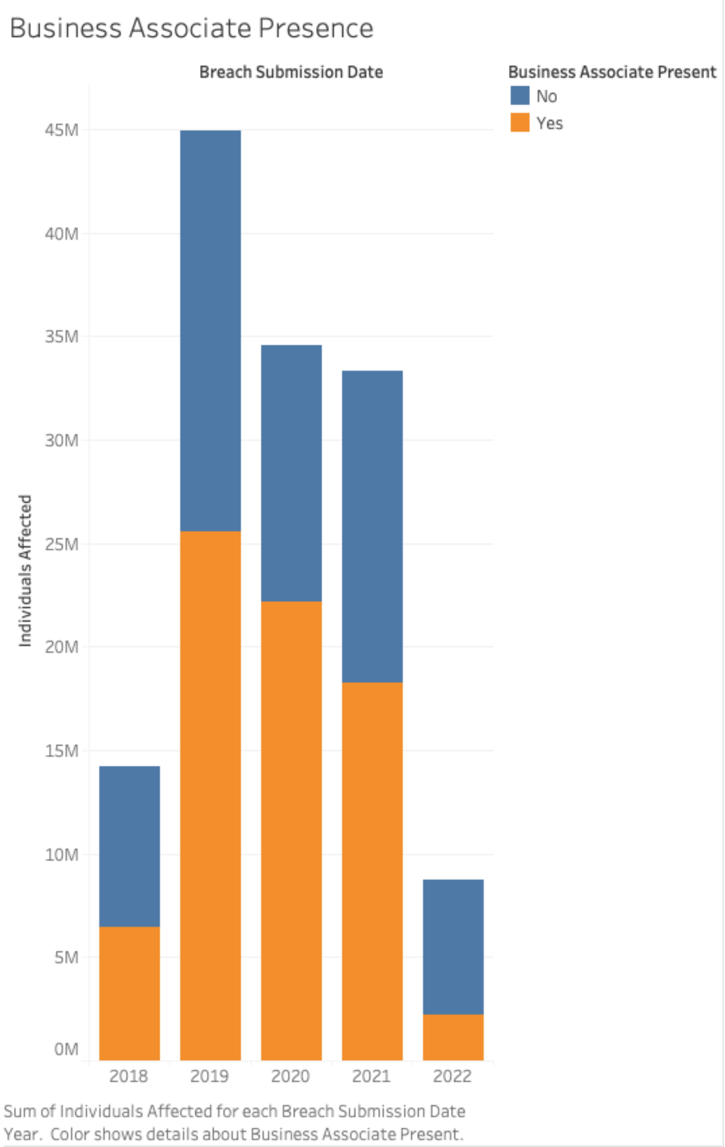


Figure 4.10 shows that healthcare organizations with business associations with cybersecurity business providers present had a much lower number of individuals affected than without one.

Findings:

In Figure 4.9 a predictive graph is shown to predict how many breaches will occur next year in 2023. Figure 4.10 that having no business association with cybersecurity business providers led to more data breaches. In 2018 healthcare organizations that had a business association present reported 6.5 million individuals affected whereas no business association reported 14 million affected, this is nearly double – a trend that is common in all years between 2018 to 2022. Based on the results from Figure 4.10 it would be beneficial for more healthcare entities to have a business association present.

CHAPTER FIVE:

DISCUSSION, CONCLUSION, AND AREAS FOR FURTHER STUDY

Discussion

This project posed the following questions:

Q1. How many individuals are affected due to data breaches in healthcare & which States had the most affected individuals?

Q2. What are the most common causes of healthcare data breaches & what measures can be taken to prevent this?

Q3. What are the healthcare data breach predictions for 2023 & what suggestions could be given to minimize them?

The findings and conclusions for each question are:

Q1. Within the last five years, millions of individuals have been affected by healthcare data breaches per year. The States where individuals are affected highest are California, Florida, Minnesota, and North Carolina. The year 2019 had the highest peak, this is the year Covid-19 occurred, Landi (2020). Ever since, the number of individuals affected has been extremely high. Proper training for staff in identifying phishing attacks could help combat and reduce the number of individuals affected each year.

Q2. The most common type of breaches are hacking/IT incidents. The compromised device/location that affected the most individuals was the network server. This indicates the need for stronger cybersecurity practices. Georgiadou, A., et al. (2022) suggested in their study of working from home during the COVID-19 crisis that organizations (such as healthcare) needed to be fully functional with certain limitations “require a deeper security culture” (491). Home networks are not always secure. Work-from-home devices need extra precautions, such as an IT team employing automatic updates overnight, or creating a virtual environment to access confidential information and work within the virtual environment.

Q3. The prediction for 2023 is that although breaches did decrease in 2022, they continue to rise again in 2023. Cybercriminals will continue to evolve their methods and breach industries. Having a healthcare business association (advisors, experienced and supportive references) present can reduce the number of individuals being affected.

Healthcare Data Breach Prevention

Healthcare data breaches can be prevented by taking several steps such as implementing strong security measures, limiting access, training employees, updating software and systems, conducting security assessments, and having a data breach response plan. Having a backup can ensure that in case of breaches, the backup be used where the information will be most credible. Converting security training into engaging activities will help employees be better prepared for instances of breach

Conclusion

In conclusion, this culminating thesis project reviewed work done by other scholars and then chose a dataset to examine by creating models and evaluating them. It was determined that within the last five years, several individuals are affected by healthcare data breaches per year. Employee education in phishing attacks could help reduce the number of individuals affected.

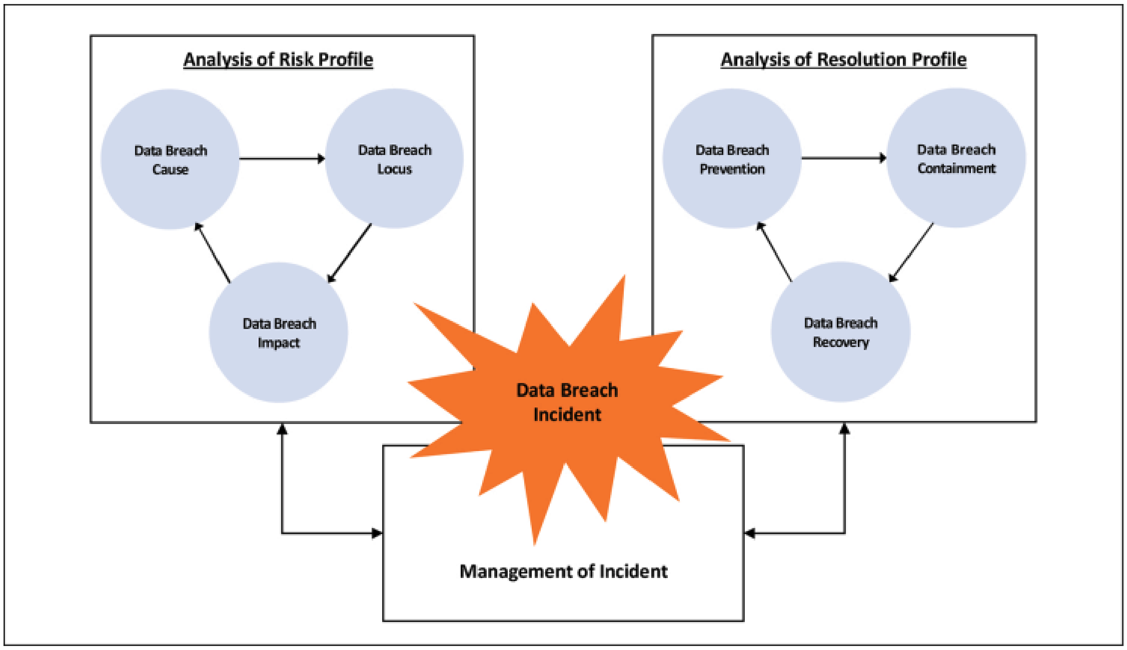
Areas for Further Study

This project analyzed data that was archived being all resolved breach reports and/or reports older than 24 months (about 2 years). For further study it could be beneficial if cases currently under investigation were also analyzed, which lists all breaches reported within the last 24 months (about 2 years) that are currently under investigation by the Office for Civil Rights. The dataset used in this project was a list of breaches of unsecured protected health information that led to affect 500 or more individuals per breach. For further study, a good start would also be including data breaches affecting less than 500 individuals. Further exploring ways of faster data breach detection is also recommended.

APPENDIX A:

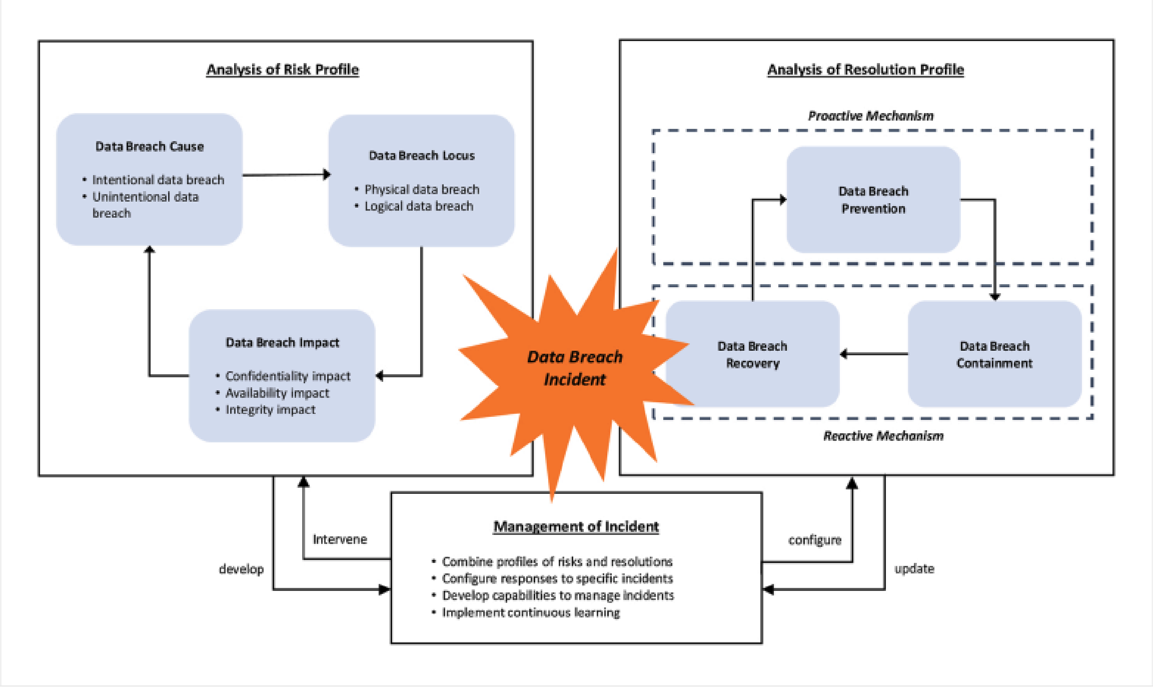
F. KHAN ET AL., 2021 CONCEPTUAL AND INTEGRATED FRAMEWORK

Conceptual Framework (adapted from F. Khan et al. 2021).



Khan (2021) proposed a conceptual framework that elaborated risk categories, resolution categories, and heuristics. “The framework analyzes an organization’s data breach risk and resolution profiles and subsequently applies them to manage data breach incidents” (Khan, p. 3).

Integrated Risk Model for Data Breach Management (adapted from F. Khan et al. 2021).



Khan (2021) offered this integrated risk model for data breach management based on the conceptual model developed.

REFERENCES

Batt, S., Grealis, T., Harmon, O., & Tomolonis, P. (2020). Learning Tableau: A data visualization tool. *The Journal of Economic Education*, *51*(3-4), 317-328.

Bhosale, K. S., Nenova, M., & Iliev, G. (2021, September). A study of cyber attacks: In the healthcare sector. In *2021 Sixth Junior Conference on Lighting (Lighting)* (pp. 1-6). IEEE.

ChatGPT. (2023, March). Six areas for further study in healthcare data breaches. Retrieved from https://chat.openai.com/c/b57a0265-6818-4a7d-8bd5-8b3ed5e2086c

Cheng, L., Liu, F., & Yao, D. (2017). Enterprise data breach: causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(5), e1211.

Chernyshev, M., Zeadally, S., & Baig, Z. (2019). Healthcare data breaches: Implications for digital forensic readiness. *Journal of medical systems*, *43*, 1-12.

Collins, J. D., Sainato, V. A., & Khey, D. N. (2011). Organizational data breaches 2005-2010: Applying SCP to the healthcare and education sectors. *International Journal of Cyber Criminology*, *5*(1), 794.

Dolezel, D., & McLeod, A. (2019). Cyber-analytics: identifying discriminants of data breaches. *Perspectives in Health Information Management*, *16*(Summer).

Elahi, H., & Geman, O. (2020). Recent Healthcare Information Breaches and their Lessons. *Archives of Surgical Research*, *1*(4), 17-23.

Georgiadou, A., Mouzakitis, S., & Askounis, D. (2022). Working from home during COVID-19 crisis: a cyber security culture assessment survey. *Security Journal*, *35*(2), 486-505.

Hammouchi, H., Cherqi, O., Mezzour, G., Ghogho, M., & El Koutbi, M. (2019). Digging deeper into data breaches: An exploratory data analysis of hacking breaches over time. *Procedia Computer Science*, *151*, 1004-1009.

Khan, F., Kim, J. H., Mathiassen, L., & Moore, R. (2021). Data breach management: An integrated risk model. *Information & Management*, *58*(1), 103392.

Kovanović, V., Joksimović, S., Gašević, D., Hatala, M., & Siemens, G. (2017). Content analytics: The definition, scope, and an overview of published research. *Handbook of learning analytics and educational data mining*, 77-92.

Landi, H. (2020). Number of patient records breached nearly triples in 2019.

Lee, J., & Choi, S. J. (2021). Hospital Productivity After Data Breaches: Difference-in-Differences Analysis. *Journal of medical Internet research*, *23*(7), e26157.

McLeod, A., & Dolezel, D. (2018). Cyber-analytics: Modeling factors associated with healthcare data breaches. *Decision Support Systems*, *108*, 57-68.

Nyakasoka, L., & Naidoo, R. (2022, July). Barriers to dynamic cybersecurity capabilities in healthcare software services. In *Proceedings of 43rd Conference of the South African Insti* (Vol. 85, pp. 231-242).

Seh, A. H., Zarour, M., Alenezi, M., Sarkar, A. K., Agrawal, A., Kumar, R., & Ahmad Khan, R. (2020, May). Healthcare data breaches: insights and implications. In *Healthcare* (Vol. 8, No. 2, p. 133). MDPI.

Yeo, L. H., & Banfield, J. (2022). Human factors in electronic health records cybersecurity breach: an exploratory analysis. *Perspectives in Health Information Management*, *19*(Spring).