Survey of the Relationship Between Understanding of Data Storage and Behavior when Deleting Confidential Information

Dissertation Manuscript

Submitted to National University

School of Technology and Engineering

In Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY IN TECHNOLOGY MANAGEMENT

by

JESSE SCHULMAN

San Diego, California

February 2026

Abstract

Deleted data often remains on storage media in a recoverable form. Users in the healthcare field are required by HIPAA law to delete confidential data before disposal, but because secure erase technologies are not always utilized, confidential data can often be recovered from discarded storage media. This quantitative study investigated the factors that may lead users to improperly dispose of drives containing recoverable confidential data. A survey of 112 respondents in the healthcare industry indicated that a lack of understanding and perceived value, as suggested by the Unified Theory of Acceptance and Use of Technology model, are primary factors in users failing to use secure erase technologies. Respondents to the survey indicated a lack of understanding of how storage media functions on a technical level, a belief that standard erasure methods, such as emptying the Recycle Bin, are sufficient, and a lack of understanding as to the benefits and usefulness of secure erase technologies. This suggests that social factors within an organization, such as training methods and policies, are a significant factor in why users fail to properly erase confidential information when disposing of storage media. This research can contribute to future studies in better user interface design, educational methods, or automated erasure techniques.

Acknowledgements

I would first like to thank my Dissertation Chair, Dr. Chris Schweigert, for mentoring me through the long process of designing and analyzing this study and for helping ensure it worked as well as possible. I would also like to thank my subject matter expert, Dr. Dan Daniels, and my academic reader, Dr. Goldstein, along with all the professors at National University who guided me along my doctoral journey. I would of course like to thank my family for their endless support and encouragement – it's thanks to you that I was able to stick it out through everything! I would also like to thank my friend Claudia for all the help when I was learning the SPSS software; the analysis would have taken much longer without your help. And of course, I would like to thank all the participants who responded to my survey – there would be no study without you. Finally, I wish to thank my colleagues and students at Metropolitan State University for your enthusiastic encouragement while I worked on this doctorate.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Despite what many users believe, deleting a file does not necessarily mean it has been removed from the storage medium. Instead, deleting a file merely removes the location information from that disk's file management system, equivalent to a library throwing away the index card for a book they no longer have in stock, without actually removing the book itself from the shelves (Carlton, 2005). These files are not truly deleted, merely invisible to the rest of the computer, until such a time as that segment of the disk is overwritten by new data. As such, file carving and other data-recovery techniques can be used to recover data that users believe they have deleted, thereby increasing security risks if they sell or dispose of their old storage media (Jones et al., 2016).

Because users lack understanding of the intricacies of data storage, they engage in unsafe practices. Studies have shown that many users, from the general public to critical organizations such as banks and governments, frequently discard drives containing recoverable confidential data (Jones et al., 2016). Malicious actors can therefore access confidential data by using discarded storage media.

The healthcare sector is particularly at risk for data theft due to the confidential nature of the data. Medical data is protected under the Health Insurance Portability and Accountability Act (HIPAA), which requires that medical information be securely stored and fully erased when a drive is discarded (Heath et al., 2022; Shamlawi, 2018). However, users frequently fail to follow these requirements, often without being aware of this failure. In several studies of second-hand drives, researchers recovered personally identifiable information (PII) from the majority of drives examined, even when the drives appeared to have had their contents erased or their file systems reformatted (Jones et al., 2016; Shamlawi, 2018). As a result, healthcare users are exposing their patients to risks and exposing themselves to potential HIPAA penalties for failing

to appropriately delete confidential data before disposing of old devices. This suggests that even users who have undergone HIPAA training are unsure how to safely remove data from old drives (Heath et al., 2022).

Simply changing how drives operate to always erase data is not feasible. Even ignoring the issue of wear and tear on the drive caused by the full erasure process, which would be especially damaging to modern solid-state drives, users often expect that data recovery should be possible, either to recover their data if it is accidentally removed or they change their minds, or because law enforcement may need to recover data from a suspect computer (Hadi, 2016). This results in a contradictory scenario in which users expect both that their data will be fully erased and that it will be recoverable. Despite the prevalence of this dilemma and its known harmful consequences, more research is needed on user expectations regarding data deletion and how these expectations inform user behavior. This lack of understanding results in a situation in which users can accidentally lose data they intended to keep, yet remain vulnerable to data theft by malicious actors who perform file recovery operations on data users believed had been entirely removed. A study that provides a better understanding of user expectations and behaviors could enable system designers to develop an improved file management system that balances privacy and availability. However, no such research has yet been conducted, rendering the root of the issue (and potential solutions) unknown and necessitating further research into this topic (Carlton, 2005; Hadi, 2016).

Statement of the Problem

The problem addressed by this study was the challenge of user handling of storage data in the healthcare industry (Heath et al., 2022). Users of digital devices often fail to properly dispose of data stored on their devices when donating or discarding them, resulting in the second-hand

market containing a large number of supposedly empty devices that can be quickly recovered. This is because users often assume that deleting data on a drive permanently erases it, unaware that the data is merely hidden rather than removed. Existing studies indicate that many second-hand devices still contain personally identifiable or confidential data that can be recovered using standard forensic tools (Jones et al., 2016; Osawaru, 2024; Sutherland et al., 2010).

In many of the cases studied, even after the drives had been formatted, researchers were still able to recover sufficient data to identify the former owners. This indicates that the original owners of the devices believed they had taken the necessary steps to dispose of their data before selling or donating the devices and were, as such, unaware that their personally identifiable data could be recovered. This suggests that users lack a proper understanding of how data storage works, or that existing data management tools fail to communicate how to delete their data, thereby implying that deletion or formatting is sufficient. This incorrect understanding may influence how users behave when dealing with confidential data. In the healthcare industry, when laws often dictate that certain information must be deleted thoroughly, leaving the data in this recoverable form exposes healthcare professionals and their patients to threats, from the loss of private data to fines or legal penalties (Heath et al., 2022; Jones et al., 2016; Shamlawi, 2018).

Purpose of the Study

The purpose of this quantitative, non-experimental, survey-based study was to determine healthcare workers' understanding of digital file storage when deleting data for HIPAA compliance and to address an existing research gap regarding the relationship between users' expectations for file systems and their perceptions of how the systems work. Users' needs for both privacy and convenience, as well as the desire to reverse accidental deletion, serve as an active dilemma for operating system designers, and a comprehensive study of what average users

require, has yet to be conducted. This issue affects users across the board, with both individual users and corporations facing threats from improperly deleted data (Hadi, 2016; Jones et al., 2016). This is the dilemma developers face when seeking to appease contradictory user goals. Further research is needed into users' needs and the broader social views on data privacy and retention to better understand how users want operating systems to handle their data and how these factors influence users' behavior when deleting HIPAA-protected files. Based on the results of such research, designers seeking to enhance their data storage code would have a firmer baseline to work from. To address this research gap, this study conducted a quantitative online survey of 112 healthcare workers in the United States, drawn from the current population of 17.4 million, according to the United States Bureau of Labor Statistics (U.S. Bureau of Labor Statistics, 2023). This survey is designed to determine the relationship between users' understanding of data storage and whether this understanding influences their behavior when working with HIPAA-protected data. The study examines one dependent variable, how users behave when erasing data, and two independent variables: the presence or absence of social factors regarding data deletion, and how a user properly understands how data storage works in a digital medium.

Introduction to Theoretical Framework

The framework for this study was based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model. The UTAUT model gauges how likely users are to accept and adopt a new technology or feature based on six factors: the perceived usefulness/performance expectancy, the perceived effort required to use the technology, social influences, the user's existing attitude toward the technology, facilitating conditions, and overall intention of use (Robles-Gomez et al., 2022; Venkatesh et al., 2003). For this study, the main attributes examined

are social influences, existing attitudes, and perceived usefulness and effort. This model is appropriate for the study because it allows the study to gauge how important users consider both permanent data deletion and the possibility of data recovery, how concerned they are about data privacy, and how difficult they find technologies such as file shredders and existing data deletion tools.

Introduction to Research Methodology and Design

The research employed a quantitative survey design. Participants were voluntarily recruited online and completed an online survey. SPSS was used to analyze results and identify relationships between responses to assess whether users' preexisting understanding (or lack thereof) of how data storage functions influence what they expect to happen when deleting a file on their computer, using the Pearson Correlation Coefficient.

Data Collection Strategy

Data was collected via an online survey to address the research questions. This data allowed the researcher to determine the relationships between users' beliefs and priorities. The results were then used to assess whether there is a relationship between users' understanding of how data storage works and their expectations for what should happen when deleting a file. According to the United States Bureau of Labor Statistics, there are approximately 17.4 million healthcare workers in the United States (U.S. Bureau of Labor Statistics, 2023). Taro Yamane proposed a formula for calculating sample size, shown in Eq. (1).

$$n = \frac{N}{1 + N(e^2)} \tag{1}$$

Equation 1 shows that n is the sample size, N is the overall population of the study, and e is the margin of error. With a 10% margin of error, this formula yields 97, indicating that the study required at least 97 participants to achieve a 10% margin of error (Israel, 1992). Users were recruited from social media platforms such as Reddit and LinkedIn, as well as via SurveyMonkey's own recruitment tool.

Data Analysis Plan

The data used to test the hypotheses were derived from the survey results and analyzed using SPSS. The data was collected via an online survey, with results anonymized to prevent identification of any specific participant. The results were examined to determine each question's mean and median response as well as using the Pearson Correlation Coefficient to measure the relationship, if one existed, between the user's understanding of data storage (the first independent variable), social factors influencing user behavior (the second independent variable), and how users behave when erasing data (the dependent variable).

Research Questions

Based on the purpose and problem statements, the study is intended to answer the following research questions.

R1

Do a statistical majority of healthcare workers understand how to erase HIPAA-protected files?

R2

To what degree does a user's understanding of how data storage works influence their behavior when deleting HIPAA-protected files?

R3

Is there a statistically significant relationship between healthcare workers' understanding of how files are deleted and how they should be deleted?

Hypotheses

$H1_0$

A majority of healthcare workers do not understand what happens when data is deleted.

$H1_a$

A majority of healthcare workers understand what happens to the data after deletion.

$H2_0$

A user's understanding of how data storage functions does not significantly affect their behavior when deleting HIPAA-protected files.

$H2_a$

A user's understanding of how data storage functions significantly affects their behavior when deleting HIPAA-protected files.

$H3_0$

There is no statistically significant relationship between healthcare workers' understanding of how files are deleted and how they should be deleted.

H3a

There is a statistically significant relationship between healthcare workers' understanding of how files are deleted and how they should be deleted.

Significance of the Study

This study aims to better understand the factors that may lead users to improperly delete confidential data. Users often sell, donate, or discard data storage media that still contain potentially damaging data in a deleted but recoverable form, exposing users to the threat of information theft (Hadi, 2016; Jones et al., 2016; Rowe, 2020). Breaches of HIPAA-protected data occur frequently and are increasing, with more than 15 million protected records exposed in April 2024 alone. While hacking is listed as the primary cause, improper disposal of stored data is identified as a factor in the breach. Poor user interface design or software failing to explain its behavior accurately could result in improper data disposal due to the user's incorrect impression of how data storage functions. As such, identifying which elements of the technology may lead users to an incorrect understanding of its functions could aid in developing better user interfaces (Heath et al., 2022; HIPAA Data Breach, 2024).

While the issue of users discarding drives containing potentially harmful data has long been recognized, no studies have examined why this occurs or how it could be mitigated (Carlton, 2005; Hadi, 2016). The intent of the study, therefore, was to determine what factors contribute to users failing to delete data they no longer wish to recover correctly, the results of which could be used by software developers and hardware manufacturers to design new data storage methods that better meet the desires of end users while reducing the risk of information theft by malicious actors, helping to protect both healthcare organizations and their customers. If the study finds that a lack of understanding is a factor in user behavior, future studies could focus on developing more effective methods of user education or on improving user interface design. On the other hand, a result indicating that user knowledge is not a relevant factor in user behavior could help future research determine where to focus when examining user behavior. As

such, this study provides a solid baseline for future research on improved data storage and file system design.

Key Terms

File Carving

File carving is a method for recovering deleted files without metadata by examining raw disk data for known code blocks that represent specific file types. Even without metadata, the hexadecimal data remains on the drive. Forensic analysts can recover individual files by locating the hexadecimal code corresponding to the data's file signature (Hadi, 2016).

File Shredders

Traditional data deletion methods are insufficient for erasing HIPAA-protected data (Certilman & Wiechmann, 2020; Lee, 2009). While metadata indicating the location and type of file fragments may be erased, the file itself remains recoverable. As such, highly confidential data is often handled using specialized tools known as file shredders. Traditional deletion of files by removing only the metadata and marking the stored data as unused is known as logical deletion, whereas fully erasing the data is known as physical deletion. Whether the deletion is logical or physical distinguishes traditional file deletion from a proper file shredder tool (Weijers, 2022). Physical deletion operates by overwriting the deleted data with new data, typically performed multiple times to ensure complete overwrite and prevent recovery. Multiple shredder algorithms exist, following different principles, ranging from overwriting the deleted file with zeros once to complex algorithms that perform 35 passes, alternating between random character overwrites and specific magnetic bit flips to ensure that no remnants of the data remain. Various

file shredder programs offer different algorithms; some allow users to select among multiple algorithms (Nahar et al., 2018; Weijers, 2022).

File Signature

A file signature is a specific, unique string of hexadecimal data at the beginning of a file, which denotes what type of file is present, even in the absence of metadata. File signatures are commonly used to identify recoverable files during file-carving operations. Forensic specialists can locate deleted files by identifying their start and end signatures (Poisel & Tjoa, 2013).

File Table

The file table is a special drive section that stores metadata for all files. The exact types of metadata and the table format vary by file system, but they generally store data such as the file name, file size, file type, and file location on the disk. File deletion removes the entry from the table, causing the operating system to treat the disk segment as empty. Still, the data remains on that disk sector until overwritten (Hadi, 2016).

Hexadecimal

Hexadecimal is a base-16 number system commonly used in computing, including in data storage. Data on a computer is stored in hexadecimal format and can be viewed by software known as hexadecimal editors, such as the free HxD software. Such programs are often used in file carving (Hadi, 2016).

HIPAA

The Health Insurance Portability and Accountability Act (HIPAA) is a law that governs the protection of personally identifiable medical information. It requires healthcare practitioners to protect patients' medical data and ensure that it is properly destroyed upon deletion to prevent

accidental disclosure. Failure to comply with HIPAA regulations can result in legal and financial penalties (Heath et al., 2022; Shamlawi, 2018).

Metadata

Metadata refers to data about data, such as a file's name, size, type, and location on the disk. When a file is deleted, this metadata is deleted rather than the file's contents. Metadata is used to determine additional information about data (Hadi, 2016).

Personally Identifiable Information

Personally Identifiable Information (PII) refers to data that can be used to personally identify an individual, such as medical information, a Social Security number, or financial information. Unlike more general information, such as birthdays and zip codes, PII is legally protected. Healthcare data is considered PII under HIPAA law (Flair, 2023).

Summary

The purpose of this quantitative, non-experimental, survey-based study was to better understand the factors that may lead users to improperly delete confidential data. Users often dispose of drives they believe they have removed confidential data from, when, in reality, the data remains recoverable. Various methods are available for malicious actors to recover deleted data, many of which are freely accessible to anyone (Jones et al., 2016). Within the healthcare field, this poses special risks, as the discarded drives may contain HIPAA-protected medical information (Heath et al., 2022). While prior studies have examined the frequency with which drives containing recoverable data are discarded and have explored new techniques for recovering the data, no studies have examined the relationship between users' understanding of how data storage operates and how users behave when deleting their data (Hadi, 2016). To

address this knowledge gap, this study used an online survey and analyzed the results in SPSS using the Pearson correlation coefficient to determine whether a relationship exists between users' knowledge of data storage and their behavior toward deleted files, and, if so, the nature of that relationship.

Key terms relevant to the study are file carving, file shredders, file signatures, file tables, hexadecimal, HIPAA, metadata, and Personally Identifiable Information (PII). File carving is the recovery of deleted data from a drive by searching for file signatures. These unique hexadecimal (sixteen-bit integer) strings denote files of a specific type, such as JPEG images or WAV audio files. Usually, the computer stores metadata – data about data, such as the location, type, and size of a file – inside the file table, but when a file undergoes conventional deletion, that entry is removed from the table, leaving the file carving as the main way to access it. HIPAA (the Health Insurance Portability and Accountability Act) is a law mandating personally identifiable data, specifically medical data, must be securely deleted, necessitating the use of tools such as file shredders – secure-erase software that overwrites or encrypts deleted data to ensure it cannot be recovered (Flair, 2023; Hadi, 2016; Heath et al., 2022; Nahar et al., 2018; Poisel & Tjoa, 2013; Shamlawi, 2018; Weijers, 2022).

Chapter 2: Literature Review

The purpose of this quantitative, non-experimental, survey-based study was to determine healthcare workers' understanding of digital file storage when deleting data for HIPAA compliance and to address an existing research gap regarding the relationship between users' expectations for file systems and their perceptions of how the systems work. Despite HIPAA regulations requiring that confidential data be securely erased to prevent attackers from recovering it, previously published studies have recovered confidential, personally identifiable healthcare data from discarded, donated, or sold drives that had been improperly deleted (Jones et al., 2016; Sutherland et al., 2010). In healthcare, data breaches are a high-severity issue and are increasingly frequent. One such cause of these breaches is improper disposal, classified as an insider threat because it results from careless or negligent employee behavior (Ewan, 2023; Seh et al., 2020).

Little preexisting research exists on user behavior regarding file deletion; the largest notable study, by Carlton in 2005, concluded that further research was needed, although, based on a literature review, no direct follow-up was conducted. As such, the literature review focused on synthesizing research on related topics to develop a more comprehensive understanding of the issue in preparation for the study. The review, therefore, focused on research on data recovery theories and techniques, HIPAA regulations governing confidential data, studies of various secure deletion tools, and past data breaches caused by improper disposal. The review also examined the literature on various technology adoption models, including the three TAM variants and two UTAUT versions, to determine which model would be most useful for the study. As such, the key terms used in the search were “Data Recovery,” “HIPAA Data Disposal,” “File Carving,” “File Shredder,” “Secure File Deletion,” “Technology Acceptance Model,” “UTAUT,” “UTAUT2”, “Digital Forensics,” “Healthcare Data Breach,” and “HIPAA

Regulations." The search engines and databases used were the National University Library, Google Scholar, and ProQuest, which included scholarly journals and doctoral dissertations. Most research papers were from the past five years, although older papers were cited where applicable when no more recent research was available.

Organization of the Study

The literature review is organized as follows: after an explanation of which theoretical frameworks were considered and how the final decision on which framework to use was reached, the review first discusses how data storage and deletion functions work, how data recovery can be performed, and the potential malicious applications of data recovery along with the standard non-malicious reasons and methods of secure data deletion. The review then examines HIPAA regulations and laws regarding secure data disposal, data breaches caused by a failure to erase data securely, and what factors impact employee behavior regarding organizational data security, including training, corporate culture, and existing biases, to demonstrate the necessity of additional research into how employee knowledge and methods of employee education can affect behavior when working with confidential or HIPAA-protected documents.

Theoretical Framework

The framework for this study was based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model. The UTAUT model gauges how likely users are to accept and adopt a new technology or feature based on six factors: the perceived usefulness/performance expectancy, the perceived effort required to use the technology, social influences, the user's existing attitude toward the technology, facilitating conditions, and overall intention of use (Robles-Gomez et al., 2022; Venkatesh et al., 2003). For this study, the main attributes examined are social influences, existing attitudes, and perceived usefulness and effort. This model is

appropriate for the study because it allows the study to gauge how important users consider both permanent data deletion and the possibility of data recovery, how concerned they are about data privacy, and how difficult they find technologies such as file shredders and existing data deletion tools.

The Technology Acceptance Model

The UTAUT model evolved from the prior Technology Acceptance Model (TAM). The TAM model was first developed in the 1980s to predict whether users would adopt a given technology. To determine this, the TAM model focuses on three primary factors: perceived usefulness, perceived ease of use, and user attitude toward use. Perceived ease of use influences the perceived usefulness, and both factors influence the user's attitude toward use, which determines the actual system use. Perceived ease of use refers to how difficult the user believes the new technology to be to learn and use – the less effort that is required for them to use the technology (or, at least, the less effort the user anticipates will be required), the more likely the user will be to try a new technology. Perceived usefulness, meanwhile, refers to how useful the user believes the new technology will be. The two factors are related: if perceived usefulness is low but perceived effort is also low, the user may still be willing to try the new technology because the perceived effort is low. These two factors combine to influence the third factor: the attitude toward use. The user's attitude and preconceived notions regarding the technology substantially affect whether the user will make an effort to use it and whether their efforts will be successful (Jan et al., 2022; Kalayou et al., 2020; Marangunić & Granic, 2014).

The Technology Acceptance Model: Version Two

The TAM model was later expanded by adding factors, resulting in TAM 2. In TAM 2, the model examines the variables influencing the perceived usefulness: subjective norm, image,

output quality, result demonstrability, and job relevance, as well as two moderating variables, voluntariness and experience. The first two factors refer to the user's social perceptions. Subjective norm refers to whether the user believes that people they consider important want them to use the new technology (or, conversely, that those individuals reject it). Likewise, the image refers to whether the user believes adopting the new technology will improve their social status. The three other factors relate to direct job performance. As the name implies, job relevance refers to whether users believe the new technology directly applies to their jobs. Users are unlikely to take advantage of new technologies if they do not perceive them as directly related to their normal job functions; as such, they are less likely to adopt technologies that lack perceived relevance. Output quality, meanwhile, refers to how well the user believes the technology performs the intended task. Even if the tool is relevant to their job, it has little perceived usefulness if it does not perform the task well, especially if the user's current tools already perform the job better. Likewise, result demonstrability refers to the extent to which users perceive the outcomes of using a new technology as tangible, visible, and demonstrable to others. Because the purpose of a new technology is to benefit users by enabling them to perform their duties more quickly and effectively, users will be less interested in the technology if they lack a means to demonstrate its efficacy. Finally, the two moderating variables, voluntariness and experience, affect the subjective norm. The willingness of users to engage with new technologies is affected by whether they believe they are choosing to engage or being forced to, and whether they have experience with the technology already (Jan et al., 2022; Kalayou et al., 2020; Marangunić & Granic, 2014; Park & Park, 2020).

The Technology Acceptance Model: Version Three

TAM 3 further examines the factors that influence the two primary variables. While the factors introduced in TAM 2 modify the perceived usefulness, TAM 3 adds the same consideration to factors that could influence the perceived ease of use. These are defined as four anchor factors and two adjustment factors. The four anchors are perceptions of external control, computer self-efficacy, computer anxiety, and computer playfulness, and the two adjustment factors are perceived enjoyment and objective usability (Tariq, 2024; Venkatesh & Bala, 2008).

Perceived external control refers to whether the user believes that organizational and technical resources exist to support the use of the new technology. Users are more likely to feel confident in their ability to use a new technology appropriately if they believe their organization will provide support if they encounter difficulties; as such, organizational support increases perceived ease of use. Meanwhile, computer self-efficacy refers to the degree of confidence a user has in their ability to perform required tasks using a computer or other technological tool. A user who is already confident in their computer skills will likely perceive new technologies and tools as easier to use than someone who lacks confidence. Conversely, computer anxiety refers to a user's fear of being forced to use a new technology. Users who lack confidence in their own abilities will perceive new technologies as more frightening and more difficult to use than those who are confident in their skills. Computer playfulness, meanwhile, is defined as the user's degree of natural cognitive spontaneity when working with computerized systems; individuals who are more creative and spontaneous when using computers are more likely to adapt quickly to new technology (Tariq, 2024; Venkatesh & Bala, 2008).

The two adjustment factors, then, are perceived enjoyment and objective usability. Perceived enjoyment is the extent to which a task or tool is considered enjoyable in itself,

regardless of any resulting performance or social benefits. Users are naturally more inclined to use tools they find enjoyable to use than those they regard solely as tools. Meanwhile, objective usability refers to the objective comparison of the effort required to accomplish tasks with the new technology versus the older one, regardless of users' perceptions. Under the TAM 3 model, these six variables are considered to influence perceived ease of use, an influence that is further affected by a user's prior experience (Kundu, 2022; Saleh, 2024; Tariq, 2024; Venkatesh & Bala, 2008).

However, whereas TAM focuses solely on the technological dimension, the evolved UTAUT model incorporates sociological factors. TAM views everything in terms of the technology, namely, how useful it appears and how difficult it is perceived to be. TAM2 and TAM3 attempt to consider some social factors as well, but only as modifying factors that influence the technological aspects rather than as separate factors considered equally to the technological factors, an oversight addressed by UTAUT. The UTAUT model examines these factors, the user's preexisting attitude toward the technology, and relevant social factors that may affect the user's likelihood of utilizing the technology. UTAUT was selected over TAM for this study because social factors (including HIPAA regulations and specific hospital policies) and existing user knowledge and education are relevant to how users may behave when deleting HIPAA-protected data and to whether they will properly engage with file-shredding tools. Additionally, while TAM 2 and TAM 3 do attempt to account for sociological factors, they treat these factors as modifiers for the technological matters, adding unnecessary complexity to the model as opposed to one which treats sociological factors as equal to technical factors (Jan et al., 2022; Kalayou et al., 2020; Lai, 2017; Loes, 2024; Marangunić & Granic, 2014; Park & Park, 2020; Saleh, 2024; Tariq, 2024; Venkatesh et al., 2003; Venkatesh & Bala, 2008).

The Unified Theory of Acceptance and Use of Technology

UTAUT specifically examines four factors: performance expectancy, effort expectancy, social influence, and facilitating conditions. The first three factors influence behavioral use intention, and this intention, along with facilitating conditions, influences actual use behavior. Performance expectancy correlates with perceived usefulness in the TAM model, referring to the extent to which users believe a new technology will benefit them. For example, it could allow users to complete tasks they already perform more quickly than they currently can, or it could refer to extrinsic factors, such as the perception that the technology will lead to social recognition from superiors and peers. These factors would, in turn, increase the intention to use the new technology, as users perceive it as valuable. Effort expectancy likewise correlates with the TAM model's perceived difficulty, i.e., the effort the user believes is required to use the new technology. For example, if the user believes the system requires excessive manual intervention, is excessively complex, or is difficult to control, the user will be less inclined to adopt the new technology because the perceived time or effort required outweighs any perceived benefits.

The third factor, social influences, is unique to UTAUT and is not examined in TAM. According to this factor, the user's perception of the people around them, the overall corporate culture, and how the new technology may impact their social status influence their likelihood of adopting it. If people with significant influence or perceived importance demonstrate that a new technology is valuable to them, other users are more likely to adopt it. Likewise, users are more likely to use a technology consistently if they are properly trained and feel supported by their managers and supervisors, or if using the technology is perceived as associated with a high-status role. On the other hand, if users do not observe their peers using the technology, do not feel

supported by their management in learning it, or view the tasks associated with it as low-prestige, they are less likely to intend to use it.

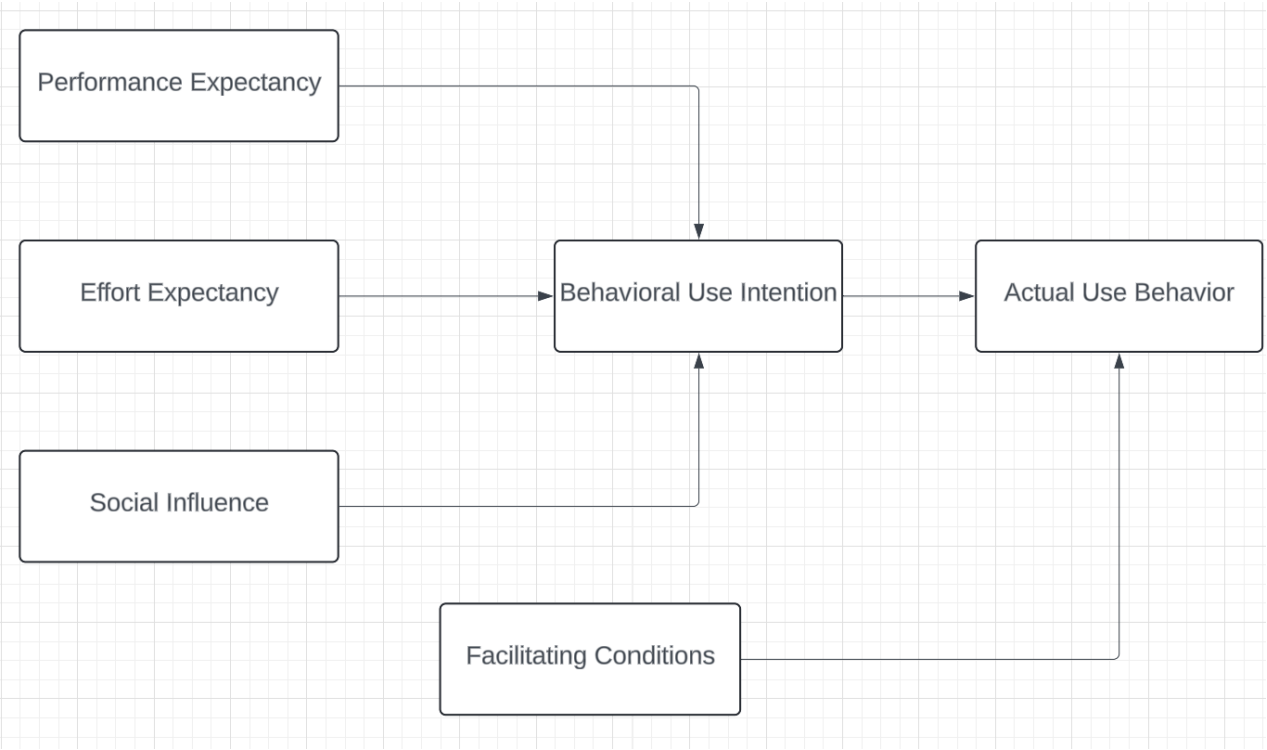
The three above factors influence what is known as behavioral intention – whether the combination of factors creates a perception for the user that they should use the new technology or not. If the overall combination of factors is a net positive, the user is likely to intend to use the technology in the future. However, how strong this inclination is may vary based on the exact nature of the three factors – a technology seen as only mildly useful, medium-effort, and not overly associated with positive or negative social factors is less likely to be adopted than a technology perceived as highly useful, low-effort, and widely supported by superiors within an organization.

Behavioral intention and facilitating conditions then combine to influence the final use behavior. Facilitating conditions refer to whether users believe that the technical and organizational infrastructure is present in a way to support the new technology. This includes whether the user believes they have been given adequate training to use the new technology, whether they have someone they can reach out to for help if they encounter problems, whether they feel they have the knowledge and resources needed to use the technology, whether they feel the system fits in well with their normal workflow and is compatible with their primary tasks, and whether they feel they have control over the new technology. If the usage intention is positive and the facilitating conditions are beneficial, this will result in a final use behavior in which the user properly utilizes the technology as intended (Venkatesh et al., 2003). UTAUT was selected over the TAM models due to the UTAUT model's unique usage of the social factors and facilitating conditions, which are expected to provide useful data regarding technology usage in large structured settings such as the healthcare field, as opposed to the TAM models, which

either do not account for social factors or only treat those social factors as modifying variables for the technological factors, rather than as unique, impactful factors in their own right.

Figure 1

Unified Theory of Acceptance and Use of Technology



Note. Adapted from "User acceptance of information technology: Toward a unified view" by Venkatesh et al., 2003.

Precedent for the use of UTAUT

Past studies have also examined the usage of other data security technologies concerning HIPAA-protected data using the UTAUT model. A 2021 study of healthcare professionals in New York City utilized the UTAUT model to examine the factors influencing whether medical professionals will properly use secure electronic medical records. In that instance, the study concluded that the primary factors were perceived performance and effort expectancy, driven by

a combination of insufficient professional training in using secure technologies and a lack of internal social factors that could have fostered the perception that the secure technology was beneficial. The technologies studied there are similar to the secure data erasure technologies examined in this study, indicating that UTAUT is an effective methodology for identifying social and educational factors that may influence whether users properly utilize secure data erasure tools (Sangurima, 2021).

UTAUT 2

Before deciding to use UTAUT, consideration was also given to using the subsequent UTAUT 2 model. UTAUT 2 builds on UTAUT by adding three new factors: hedonic motivation, price value, and habit. Hedonic motivation refers to any subjective positive emotional satisfaction a user receives when using a tool or technology, and price value refers to the return on investment of adopting the new technology that the user is consciously aware of (in other words, only the returns that the user directly perceives and understands are relevant. If the user has a substantial return but is unaware of it, it is not a factor UTAUT 2 considers). Habit refers to situations in which the use of a technology or tool has become automatic for the user (Alghatrifi & Khalid, 2019; Alqahtani & Braun, 2021; Tamilmani et al., 2021).

UTAUT 2 was excluded from the study because the price value and habit factors are not directly relevant to the planned study; the users in question will generally not be involved in purchasing the software or be aware of the cost and return on investment. The habit factor is generally relevant once a software or tool has been used for an extended period. Instead, to avoid adding unnecessary complicating factors, the study intends to focus on the three primary UTAUT factors- social influences (including coworkers as well as management and corporate policy), effort expediency (which could include developing a habit as well as any enjoyability factor

involved in using the technology), and performance expectancy (which could also include anticipated returns on investment) (Alghatrifi & Khalid, 2019; Alqahtani, & Braun, (2021); Lai, 2017; Sangurima, 2021; Tamilmani et al., 2021; Venkatesh et al., 2003; Venkatesh & Bala, 2008).

Research on data forensics is generally divided into two categories. One field of research focuses on techniques and tools for recovering deleted data. In contrast, the other focuses on studying the types of data recovered from drives to determine whether personal or compromising data can be recovered. This study synthesizes ideas from both research fields to create a holistic view of how data is actually handled versus how data owners believe it is handled, and to determine which organizational factors influence any potential gap between user belief and reality.

What happens to deleted data

The term deletion, when applied to data on a digital medium, is inaccurate. When data is removed, it is merely hidden from the user's and the operating system's view but remains on the disk as raw data (Hadi, 2016). Raw data refers to data stored on the disk but not referred to by any metadata (data about data) that would indicate the data's presence in the operating system. Normally, when a file is stored on a computer, the computer uses a database called a file table to track the file's location. The file table includes information on the name of the file, the type of file (for example, whether it is a JPEG image, a DOC file, etc.), the size of the file, and other information such as where on the physical drive the file is located or, if the file is broken into multiple fragments, where on the physical drive each fragment begins and ends and what order they are placed in (Carlton, 2005). This file table enables the computer and software to locate and read files. When a file is deleted, its record is removed from the file table, so the operating

system and software no longer know the file's type, location, or existence. However, the file remains on the drive as the raw data; it is simply invisible (Azeem, 2022; Poisel & Tjoa, 2013; Weijers, 2022). The operating system now believes that the drive segments containing the data are empty and can overwrite the existing data by writing new files there. Until and unless the data is overwritten fully, it remains on the disk, unknown to the drive's owner or the computer itself. The continued existence of data on the disk poses a threat to the drive's owner due to the risk of data recovery, including for malicious purposes.

Malicious recovery of data from victims' drives

Criminals can use data recovery techniques to cause harm. When drives containing deleted data are sold, donated, or discarded, criminals may be able to recover this data and use it for malicious purposes. The majority of data storage media obtained through secondhand methods have been found to contain valuable data, including bank account information, financial transaction logs, and usernames and passwords, all of which were deleted rather than erased from the drive (Jones et al., 2016). This data can often be recovered with little difficulty, meaning that criminals need only purchase used drives from second-hand sources and perform standard forensic operations to steal lucrative or damaging data. The tools used to recover this data are commercially available to the public and, in many cases, are free, meaning the barrier to entry is relatively low for prospective criminals (Jones et al., 2016). This means that stealing data from discarded drives is a relatively easy, low-risk, and low-cost method of data theft, making the second-hand market an attractive target for criminals. The same issue applies to drives suspected of being broken. In many cases, even when physical damage prevents the drive from booting normally, it can be repaired using specialized tools or by sourcing replacement parts from other drives, thereby allowing the data on the drive to be retrieved (Sutherland et al., 2010). Because

many damaged drives are either discarded or returned without further steps to destroy them, they can be recovered by criminals, who may be able to retrieve data that appears to have been erased.

Lost and Stolen Drives

Drives may also be lost or stolen. As with donated and discarded drives, malicious actors can employ file-recovery techniques to retrieve deleted data from lost or stolen drives. In these cases, the risks may be even greater, as while discarded, donated, or sold drives have often had some attempt at data erasure performed, such as reformatting, a drive that has been lost or stolen is unlikely to have had any data erasure techniques performed on it unless all deleted files were initially erased using a secure deletion/file shredding technique, which few organizations are likely to require as a matter of policy. As such, users performing standard deletion, from which files can be easily recovered, are at even greater risk in the event of a drive being lost or stolen, an event which happens with relative frequency between 2010 and 2022, theft/loss incidents accounted for 31.94% of reported healthcare data breaches (Almulihi, 2022; Jones, 2019; Sloane & Juhnke, 2016).

Intentional restoration of deleted data

Data may also be recovered for legitimate reasons. Users may accidentally delete data and seek to recover it; in such cases, a system that permanently deletes data immediately would be actively harmful to users (Carlton, 2005). Users require the ability to recover accidentally deleted data when they have control over their storage media. If the data were truly lost as soon as the drive reported it had been deleted, file-recovery services could not restore the data to its rightful owner.

Likewise, law enforcement officers may benefit from data that users believe to be deleted. While deleted drives that still contain data pose a risk to their owners, this also applies

when the user is a criminal attempting to hide evidence of a crime (Hadi, 2016). In these cases, the fact that data remains on the drive is beneficial to law enforcement officers, who can recover deleted data and use it as evidence in their investigations. Numerous forensic tools and techniques exist specifically to recover data for law enforcement purposes (Azeem, 2022).

File Recovery Through File Carving

Data recovery without any metadata is performed using file carving. This technique examines the raw hexadecimal data on the drive and reconstructs the data by manually extracting the hexadecimal data corresponding to the file in question. Traditionally, this is a signature-based approach that relies on file types having unique start and end hexadecimal signatures, such as all JPEG files beginning with FF D8 FF and ending with FF D9. However, this approach is insufficient for fragmented files, which are common. In file fragmentation, the system attempts to maximize the use of available disk space by breaking files into multiple noncontiguous fragments. In other words, if a drive has free space in cluster 300 and cluster 308 but not in any of the intervening clusters, and a file would require two clusters worth of space to store, the system will break the file into two fragments, placing one in cluster 300 and one in cluster 308, rather than continuing to search the drive until it finds two contiguous clusters in which to store the data. When the drive has functional metadata, this is not an issue, as the metadata will indicate to the system how many fragments the file has, where the fragments are located, and the order in which they should be read. However, when the metadata is erased, traditional file carving methods may fail to recover the entire file because they cannot locate all file fragments, or because fragments of unrelated files appear within the recovered data. On SSDs, in particular, data is likely to be fragmented because SSDs do not support disk defragmentation, as read/write

operations involved in defragmentation can degrade the drive's lifespan and provide no benefits (Ali et al., 2018; Pal & Memon, 2009; Ravi et al., 2016).

Recovering Fragmented Files

However, file fragmentation is insufficient to impede a determined analyst. There are multiple methods for reconstructing and recovering fragmented data. This commonly involves using graph theory to predict where other fragments may be located, using one of several algorithms, such as Shortest Path First, which attempts to reconstruct data based on the cost of the paths, calculated based on the sum of the weights between clusters of a recovered file divided by the number of clusters, working under the assumption that the best recoveries have the shortest costs. Another type of data recovery is Bifragment Gap Carving, which works for files split into only two fragments by identifying the header and footer fragments of the file, then exhaustively checking all possible combinations of data clusters between the two fragments until the data can pass a validation check based on the type of data in use (Pal & Memon, 2009).

A newer method of data recovery from fragmented systems relies on machine learning. For example, when recovering images, a technique known as Sequential Pixel Prediction can be used. This technique uses a neural network to predict the sequence of clusters from the known JPEG file structure. It can recover images from only the data within the deleted image, provided no data from other images is present in the file. Other techniques include fragment matching, which works by pairing the ending of one fragment with the start of another fragment based on known information about the structure of the file type, and fast object validation, which combines knowledge about a file system (such as knowing where on a disk sequences are allowed to begin) and gap carving with existing file type validator tools to quickly automate the recovery of deleted files in bulk (Ali et al., 2018; Garfinkel, 2007). Regardless of the recovery

method employed, modern tools can frequently recover even fragmented files, indicating that malicious actors can still recover data that has not been physically erased.

File Shredders

Traditional data deletion methods are insufficient for erasing HIPAA-protected data (Certilman & Wiechmann, 2020; Lee, 2009). Although metadata indicating the location and type of file fragments may be erased, the file remains recoverable. As such, highly confidential data is often handled using specialized tools known as file shredders. Traditional deletion of files by removing only the metadata and marking the stored data as unused is known as logical deletion, whereas fully erasing the data is known as physical deletion. Whether the deletion is logical or physical distinguishes traditional file deletion from a proper file shredder tool (Weijers, 2022). Physical deletion works by overwriting the deleted data with new data, typically repeated multiple times to ensure complete overwrite and prevent recovery. Multiple shredder algorithms exist, following different principles, ranging from simply overwriting the deleted file with zeroes one time to complex algorithms that run a series of 35 passes, alternating between random character overwrites and specific magnetic bit flips to ensure no remnants of the data are present at all. Various file shredder programs offer different algorithms; some allow users to select among multiple algorithms (Nahar et al., 2018; Weijers, 2022).

Drawbacks of File Shredders

While file shredders provide a method of physical data deletion, they are not without drawbacks. Aside from the fact that users may sometimes want their data to be recoverable, physical data deletion has several disadvantages relative to logical data deletion, making it impractical for non-sensitive data and routine use. First, even overwriting the data entirely may not be sufficient depending on the file system: file systems such as NTFS (used by Windows

operating systems) often store mirrored versions of data, known as shadow copies, for backup and recovery purposes in the event of data corruption. These backup copies are stored elsewhere on the drive to enable recovery of older versions of the file if the existing data is unintentionally damaged or altered. They are not pointed to by the same metadata as the primary data, meaning that when the main data is erased, the file shredder will not be aware of the existence or location of the shadow copies and, as such, will leave them intact. Not only does this mean that the data can be recovered by analysts who locate the shadow copies, but the presence of shadow copies without any sign of the original data could indicate to analysts that the data in question was considered confidential and that the original owner took pains to delete it, which could result in the analysts providing increased scrutiny to the recovered data that they might not have otherwise. Erasure procedures would therefore also need to know the locations of the shadow copies and manually erase them along with the original data, creating additional work for users and requiring greater technical knowledge (AlHarbi et al., 2022; Nahar et al., 2018; Weijers, 2022).

There is also another severe drawback to physical data deletion. The process of erasing data relies on multiple write operations across the drive to ensure complete overwrite, but these operations introduce additional risks. First, there is simply the time issue: these operations take a long time to run, and the system cannot be used during the process because of how resource-intensive the erasure is. As such, file shredder operations would be unusable during normal business hours, so it would not be feasible to perform file shredding every time a HIPAA-protected document is deleted. Users would instead need to schedule shredding tasks to run outside operating hours, which could pose difficulties in leaving systems to perform operations

unsupervised and could cause other technical issues if something were to go wrong during the operation (Weijers, 2022).

*File Shredders on Solid State Drive*s

The second issue with file shredders is more severe. Each read/write operation on a disk results in wear and tear on the drive. While this may be manageable on traditional magnetic hard drives (though such drives are also far slower, so the process will take substantially longer), on Solid State Drives (SSDs), each read/write operation can significantly reduce the drive's lifespan. On modern SSDs, regular physical data deletion can reduce the drive's lifespan by more than 50%, further shortening its already limited lifespan. As such, running file shredders on SSD devices will significantly shorten the drive's lifespan and likely render the drive unusable. In the case of wiping an entire drive using the standard seven-run physical deletion process (the minimum recommended by the German Federal Office for Information Security for their standards on data deletion), this would reduce the drive's lifespan by a full 85% and a 35-pass full-drive erasure would reduce the drive's lifespan by 97%, likely immediately destroying the drive if it had undergone any significant usage before (Weijers, 2022). Solid State Drives have an alternative called "Secure Erase," but this is not a standardized function and can mean anything from overwriting the data to purging the file table to encrypting the drive and then erasing the decryption key (Nahar et al., 2018; Weijers, 2022).

Secure Erase of Solid State Drives

In most cases, Secure Erase does not perform a physical deletion, and data recovery remains possible. At present, considering the dangers posed to SSDs by repeated write operations, the most feasible method of secure data deletion for SSDs is that of encrypting the entire drive (a function many professional-grade drives perform by default) and then erasing and

overwriting only the decryption key itself, leaving the data present on the drive but encrypted. While this is reasonably secure, assuming the drive is using a strong cryptographic method, the fact remains that the data is still present on the drive and could, in theory, be recovered if an attacker were to break the algorithm, a risk that increases as computers become more powerful and therefore better at brute-forcing decryption. As such, for HIPAA-protected data, the cryptographic erasure method may not be sufficient to fully purge the data. Combining this issue with the inability to perform overwrite-based physical data deletion on SSDs means that organizations needing to purge HIPAA-protected data fully may need to rely on older, slower magnetic hard drives instead of modern SSDs for their operations, at least when working with HIPAA-protected data, which could cause additional inconvenience for users that threatens their willingness to follow protocols (Nahar et al., 2018; Weijers, 2022).

If a drive is intended to be permanently erased, rather than simply removing files during normal usage to prevent malicious recovery of files from active drives, these concerns may not be an issue. Damaging an SSD's lifespan is not a relevant factor if the drive is not intended for additional use. In this situation, other full-disk erasure methods are also available to magnetic hard disks, such as degaussing, which erases all data on the drive using a strong magnetic field. Since degaussing does not apply to non-magnetic drives, however, the only method of erasure for SSDs remains either cryptographic erasure, the feasibility of which depends on organizational protocols, or the overwrite-based physical erasure method, which is only valid if the drive is not intended for reuse, sale, or donation, as that method is likely to destroy the drive permanently. As such, organizational protocols and policies are likely to influence whether such methods are viable when erasing older drives. If the drive is not intended for reuse, sale, or donation, then a

secure deletion method that damages the life expectancy of the disk is not an issue (Rodrigues et al., 2024).

User Behavior Regarding Data Deletion

Studies indicate that user behavior regarding data deletion does not align with how data deletion is implemented. Users frequently delete files or reformat their drives, then sell or donate those drives, suggesting that they believe this action erases data beyond what the average person can recover (Jones et al., 2016). As a result, users expose themselves to avoidable risks by making drives containing personal data publicly available under the false belief that the data has been deleted. Likewise, users often return or donate drives they believe are broken, even when they are often relatively easy to repair and from which data can be recovered (Sutherland et al., 2010). Once again, users unknowingly make their potentially compromising data available to others. Users generally assume that their data is irretrievable once it is deleted, the disk is formatted, or physical damage occurs, even when the data could be recovered using freely available tools or file-recovery services accessible to the users. This exposes users to easily preventable data theft by allowing malicious actors to access large amounts of confidential data that can be recovered using freely available tools and techniques.

Data Security and Privacy

Data security and privacy are key risk factors for deleted files. Users discard, donate, or sell drives they believe are empty, even though they may contain personal photos or government secrets (Sutherland et al., 2010). Failure to properly dispose of damaged or seemingly deleted drives exposes end users to criminal threats. Many drives still contain confidential, personally identifiable data that could be used to harm the original owner, including banking information,

passwords, personal addresses, and medical information (Jones, 2019; Sharma & India, 2021; Sloane & Juhnke, 2016; Sutherland et al., 2010).

Even when users have general knowledge that deleting a file does not permanently erase it, steps such as formatting the drive are insufficient to remove the data (Jones et al., 2016). This suggests that users, even those with sufficient technical knowledge to reformat a hard drive and understand the consequences, still need a proper understanding of how to erase files they want to ensure are gone. However, there is little research on these users' mindsets, and existing studies often yield conflicting findings: users desire both to have their data fully erased and to be able to recover it if they change their minds. As such, the root cause of this disconnect remains to be determined, whether it stems from a lack of education, unintuitive user interface design, file system structure, or other factors (Carlton, 2005).

Balancing Confidentiality with Availability

Data recovery techniques take numerous forms and can serve both benevolent and malicious purposes. Files can be recovered using advanced algorithms on functioning drives, even if the metadata is gone or the files are fragmented. Forensic specialists can recover substantial information as long as the data has not been completely overwritten, and data can be recovered even from damaged drives with simple repairs. Often, these repairs are neither particularly difficult nor expensive and are thus accessible to the average user. This same principle applies to data recovery from raw data, with many tools being free. This can be beneficial to users: files may be accidentally deleted, and an easy method for recovering data can prevent the loss of valuable or sentimental files. It can also be helpful to law enforcement, which may obtain critical evidence by recovering deleted files. In such a case, allowing the user to fully

erase their data at will would increase the risk to others by preventing law enforcement from completing their case (Hadi, 2016; Jones et al., 2016; Rowe, 2020; Sutherland et al., 2010).

At the same time, users have expectations of privacy, including the expectation that their deleted data is truly gone. Once again, the disconnect between expectations creates difficulties for file system engineers seeking to support both goals: users have conflicting expectations that their data is both secret and accessible when needed. At present, no research exists regarding the root cause of this disconnect, nor into methods of alleviating it. Rather, existing research focuses either on methods of data recovery or on the consequences, both beneficial and adverse, of recovery (Carlton, 2005; Hadi, 2016; Rowe, 2020).

HIPAA Laws Regarding Data Storage and Deletion

While HIPAA requires that sensitive records be disposed of, the regulations do not provide exact details. Instead, guidelines state that "the HIPAA Security Rule requires that covered entities implement policies and procedures to address the final disposition of electronic PHI and/or the hardware or electronic media on which it is stored" (U.S. Department of Health and Human Services, 2013, p. 4). As a result, healthcare organizations are left to determine for themselves what measures may be required, which can and does lead to organizations failing to dispose of data properly due to insufficient specificity and defined procedures in their policies. For example, in 2009, a study of used and discarded drives from various organizations was able to recover treatment records for numerous patients who had undergone cancer treatment, a direct example of potentially harmful and highly personal medical PII recoverable from a drive whose owners believed they had purged the data before disposal (Jones, 2019; Sloane & Juhnke, 2016).

Because regulations and guidelines may be vague and nonspecific, social factors within individual organizations influence employees' behavior regarding the deletion of sensitive data.

Prior studies have indicated that social factors, such as those examined by UTAUT, are a key factor in healthcare employees failing to perform adequately when deleting protected files. Training and funding are often limited in the healthcare industry, and many managers and other senior personnel within healthcare organizations often lack sufficient technical knowledge and experience in protecting confidential medical data (Basile, 2020).

As such, despite healthcare leaders being aware that data breaches are a severe danger and that they need to take steps to protect their organization, they lack the experience and education needed to know how to protect the data, a problem exacerbated by the lack of regulations or standardized industry guidelines that managers lacking technical knowledge could follow. This results in situations in which organizations in the healthcare industry may be exposed to a greatly increased risk of data breaches due to a combination of social factors and budgetary/technical limitations. The UTAUT model, applied to those factors, would predict that users would fail to use the technologies and tools needed to ensure that confidential HIPAA-protected data is properly erased.

Data Breaches Caused by Failure of Disposal

Users can expose themselves to data theft by discarding older storage media that have not been properly wiped. There have already been reported cases of confidential data being recovered through forensic methods by malicious actors, for example, with instances of the Chinese government utilizing forensic methods to recover deleted data from corporate targets (Rowe, 2020). These breaches also do not occur purely when malicious individuals are involved. Forensic investigators or other data recovery specialists may inadvertently recover potentially harmful data, and a poor judgment call or failure to follow proper guidelines may result in that recovered data being exposed or disclosed in a manner that could harm the owner (Carlton, 2005;

Rowe, 2020). The problem is exacerbated by organizations that do not know how to respond to such breaches; while many organizations will react to a forensic-related data breach by working to prevent future occurrences, others have been reported to have responded by denying the existence of any such issue. Organizations' failure to respond appropriately to a data breach can exacerbate the initial threat. If the organization does not acknowledge the cause of the breach, it is unlikely to take action to prevent future occurrences. As such, whatever shortcomings in training, policy, enforcement, or technology enabled the original breach will remain unresolved (Jones, 2009).

Consequences of Noncompliance

HIPAA noncompliance, including failures to delete data, has been a significant and rising problem in recent years. However, limited research has examined the factors contributing to non-compliance and the steps organizations can take to improve compliance (Basile, 2020; Heath et al., 2022; Khanijahani et al., 2022; Osawaru, 2024). Some limited studies have indicated that training in HIPAA regulations and requirements, as well as corporate data security policies and the penalties for noncompliance, can have a mitigating effect that reduces the likelihood of intentional noncompliance; however, additional research has yet to be conducted, especially regarding secure file disposal and employee training in utilizing secure disposal techniques. Additionally, training techniques used by many organizations are insufficient, or in some cases, outright counterproductive, indicating that many healthcare organizations are exposing themselves to risks of noncompliance even if their policies do require secure deletion due to negative social influences, such as limited training, within the organizations (Basile, 2020; McLaughlin, 2023).

Data breaches are a particularly severe concern in the healthcare industry, as many healthcare organizations are technologically under-equipped and often fail to use existing security technologies effectively. In 2020, the healthcare sector took nearly 100 days longer to identify and contain data breaches than the cross-company average. This excessively long delay in healthcare organizations' detection and response to data breaches exposes them and their customers to increased risk of harm, giving attackers substantial additional time to steal confidential data and use it for malicious purposes. This lends further evidence to the necessity for organizations to work to prevent data breaches from occurring in the first place, including using file shredders and other secure deletion methods for erasing confidential data once it is no longer required, rather than leaving the data on the storage media in a state that could be recovered and used for malicious purposes (Daggupati, 2020).

Social Influences in Technology Usage

According to the UTAUT model, social influences are a factor in whether users choose to adopt and utilize new technology, referring in this instance to the file shredders and other secure deletion methods necessary to securely erase HIPAA-protected data in a manner that ensures the data cannot be recovered through file carving or other standard forensic methods. The social influence is likely magnified in this situation because HIPAA regulations do not require the organization to use any specific technology or tool for securely erasing data, nor do they specify what qualifies as securely or insecurely erased. Because of this, the behavior of coworkers and superiors is likely to be a significant factor (Lai, 2017; Randolph, 2024; Venkatesh et al., 2003; White, 2023).

Social factors can include training, budgetary matters, and the actual attitudes of coworkers and superior officers. In the absence of official external guidelines, rules governing

data erasure must be established internally. These can refer to specific training regarding data security and policies created by organizational leads. Data security policies, including secure erasure protocols, must be communicated to employees during training to ensure they understand the required procedures. One issue presented is, therefore, the matter of training – how can employees be appropriately trained in secure data removal to ensure that employees will remember and choose to abide by the policy (Ewan, 2023; McLaughlin, 2023)?

The Impact of Training as a Social Factor

Frequently, training presents a problem for organizations. Training is constrained by budget and available time and must be presented in a manner that ensures employees understand and retain the information. In many cases, employees are disinterested in training they consider irrelevant, viewing it as a distraction rather than as valuable. As such, while cybersecurity training has become increasingly common, most attempts to measure its efficacy overlook that, in many cases, users may intentionally compromise their cybersecurity for convenience as long as the perceived threat remains hypothetical. Organizations often seek to improve the effectiveness of training programs to ensure that employees apply the training, but such efforts are frequently counterproductive. For example, many modern training programs employ gamification to transform a traditional program that may be perceived as dull into an engaging game that educates users through gameplay. In real-world studies of this method, however, gamification not only fails to improve training effectiveness, it also actively repulses employees, dramatically reducing their desire to take the training and implement what was learned. This demonstrates that improper training methods can foster an organizational climate that actively discourages adherence to policy and the use of available security tools (McLaughlin, 2023).

Organizations seeking to improve training generally reported better results when the employees considered the training relevant to their positions, meaning that the training had been personalized to their specific roles and related risks rather than a generic one-size-fits-all approach that often fails to justify its relevance to any particular function. Training must be personalized to the user's role and prior knowledge, as untrained users will likely make errors that reflect their prior knowledge and biases. However, many organizations surveyed in past studies appeared unwilling or unable to allocate the budget for effective personalized training. Despite training being a key preventive measure to prevent data breaches caused by improper data disposal, many organizations handling confidential data fail to allocate sufficient time and resources to ensure the training is effective. As such, to improve the adoption and usage of secure erasure methods, organizations may need to take a top-down approach to change the internal culture, beginning with higher-ranking executives allocating more time and resources to comprehensive, personalized training programs to ensure that employees understand both how and why to utilize the file shredder applications despite the potential additional effort required to use them (Ewan, 2023).

Studies of hospitals that had experienced data breaches often found that social factors were a dominant factor in the initial failure that led to the breach. However, studies have not examined how this may interact with the use of file-shredding software (or its absence). As such, it is unclear whether social or educational factors influence users' use of file shredders or whether the issue is more a matter of organizational policy and regulations (Heath et al., 2022; Khanijahani et al., 2022; Osawaru, 2024).

Summary

Data breaches are a major concern across industries. However, they are especially significant in the healthcare industry, where personal health information poses substantial risks to clients if the data is exposed. Despite the importance of data security in the field, HIPAA regulations and similar laws must provide adequate guidance on how organizations should ensure that data is kept secure and securely erased when necessary. Additionally, many organizations subject to HIPAA regulations fail to comply, resulting in frequent penalties and data breaches. When such breaches occur, the healthcare industry is also slower than most other industries in detecting and responding to incidents, by an average of nearly 100 days (Daggupati, 2020; Hadi, 2016).

The term "deleted data" is a misnomer, as most deleted data remains on the drive and is merely inaccessible due to the absence of file table metadata indicating where the data is stored and how it can be viewed. However, data recovery is a relatively simple and inexpensive process, with numerous free and open-source tools for file recovery and more advanced paid programs that further simplify the process. Even fragmented data or partially overwritten files can be recovered using advanced algorithms in newer tools, allowing attackers to retrieve almost any data that has not been physically erased. Even damaged drives can have data recovered by replacing components. Because modern data recovery tools are so powerful, even users who perform basic diligence in erasing and reformatting their drives before selling, donating, or discarding them remain vulnerable to having their data recovered from those drives. This is exacerbated by the risk that drives may be stolen, with users likely having performed no deletion steps prior to theft (Azeem, 2022; Poisel & Tjoa, 2013; Weijers, 2022).

To properly delete data, users must use a secure erasure method. However, securely erasing data is difficult. Secure erasure tools require much time to run, and the drive cannot be used while the tool is running. Additionally, secure erasure tools impose significant strain on the drive and can irreparably damage an SSD after a single run, owing to the limited read/write capacity of solid-state devices. SSDs offer an alternative method known as Secure Erase; however, this is not a standardized technology, and devices use it differently, making it unreliable for drive erasure. As such, secure erasure methods are primarily applicable to hard drives, whereas solid-state drives may require the less effective Secure Erase method if the user is unwilling to destroy the drive physically (Nahar et al., 2018; Weijers, 2022).

Existing studies have shown that training and organizational culture significantly influence employees' intent to comply with HIPAA regulations. Still, little research has been conducted on secure file disposal within the healthcare industry despite improper disposal being a recurring cause of data breaches. Existing studies regarding education and internal culture generally focus on compliance with storage policies but do not examine deletion behaviors. To fill this gap in research, this study examined how (and if) users' understanding of how data storage and data deletion functions affects how users behave when deleting HIPAA-protected data to better guide future research into improved methods of education and better policies regarding data storage and erasure to help prevent data breaches caused by improper disposal (Basile, 2020; Ewan, 2023; McLaughlin, 2023).

Chapter 3: Research Method

The problem addressed by this study was the challenge posed by the handling of storage data by users in the healthcare industry (Carlton, 2005; Shamlawi, 2018). Users of digital devices often fail to properly dispose of data stored on their devices when donating or discarding them, resulting in the second-hand market containing a large number of supposedly empty devices that can be quickly recovered. This is because users often assume that deleting data on a drive permanently erases it, unaware that the data is merely hidden rather than removed. Existing studies indicate that many second-hand devices still contain personally identifiable or confidential data that can be recovered using standard forensic tools (Jones et al., 2016; Osawaru, 2024; Sutherland et al., 2010).

In many of the cases studied, even after the drives had been formatted, researchers were still able to recover sufficient data to personally identify the former owners. This indicates that the original owners of the devices believed they had taken the necessary steps to dispose of their data before selling or donating the devices and were, as such, unaware that their personally identifiable data could be recovered. This suggests that users lack a proper understanding of how data storage works, or that existing data management tools fail to communicate how to delete their data, thereby implying that deletion or formatting is sufficient. This incorrect understanding may influence how users behave when dealing with confidential data. In the healthcare industry, when laws often dictate that certain information must be deleted thoroughly, leaving the data in this recoverable form exposes healthcare professionals and their patients to threats, from the loss of private data to fines or legal penalties (Heath et al., 2022; Jones et al., 2016; Shamlawi, 2018).

The purpose of this quantitative, non-experimental, survey-based study was to determine what healthcare workers prioritize in file management systems when deleting data for HIPAA compliance and to address an existing research gap regarding the relationship between users'

expectations for file systems and their perceptions of how the systems work. Users' needs for both privacy and convenience, as well as their desire to reverse accidental deletions, pose a persistent challenge for operating system designers, and a comprehensive study of what average users require has yet to be conducted. This issue affects users across the board, with both individual users and corporations facing threats from improperly deleted. This is the dilemma developers face when seeking to appease contradictory user goals data (Hadi, 2016; Jones et al., 2016). Further research is needed into users' needs and the broader social views on data privacy and retention to better understand how users want operating systems to handle their data and how these factors influence users' behavior when deleting HIPAA-protected files. Based on the results of such research, designers seeking to enhance their data storage code would have a firmer baseline from which to work.

Research Methodology and Design (Nature of the Study)

The research employed a quantitative survey design. Participants were voluntarily recruited online and completed an online survey. SPSS was used to analyze results and identify relationships between responses to assess whether users' preexisting understanding (or lack thereof) of how data storage functions influence what they expect to happen when deleting a file on their computer, using the Pearson Correlation Coefficient. The study focused on healthcare workers in the United States, examining how well they understand data deletion and whether this understanding affects their behavior when working with HIPAA-protected data.

This study employed quantitative methods rather than qualitative or mixed methods due to the nature of the research questions. The problem being examined is the issue of what factors, such as the technology failing to properly communicate its behavior to a user or the failure of organizational policies to address the data storage issue directly, result in improper data disposal.

The study used a structured quantitative survey method to gain measurable results to directly test the hypotheses proposed to answer the research questions, rather than examining subjective experiences or developing a hypothesis post-study, as is done in qualitative research (Cresswell, 2013; McCusker & Gunaydin, 2015)

Population and Sample

The population consisted of healthcare workers in the United States, as those are the workers directly working with HIPAA-protected data. According to the United States Bureau of Labor Statistics, there are approximately 17.4 million healthcare workers in the United States (U.S. Bureau of Labor Statistics, 2023). Following Yamane's formula (Eq. (1)), targeting a 10% margin of error, the study required a minimum of 97 participants (Israel, 1992). The final study surveyed a sample size of 112 participants. The total number of respondents who completed the surveys was used as the final sample.

Instrumentation

The survey instrument used a Likert scale of 1-5. Likert scales are a quantitative method for quantifying otherwise subjective data by asking respondents to rate their opinions on a numerical scale, typically between 1 and 5. However, one through seven are also common, the results of which can then be processed to gain information such as mean ranking and standard deviation on responses (Kolil & Achuthan, 2022). A five-point scale was selected given the nature of the questions, which focus on whether the user understands a specific topic (e.g., data storage, data deletion, HIPAA policy) rather than on a more nuanced continuum of opinions. In this situation, five-point scales allow for a basic degree of "strong" or "weak" yes/no, along with a neutral/unsure option, without having excessive nuance, whereas a seven or ten-point scale is

more frequently used for cases where the choice being made is a more subtle gradient with greater nuance (Joshi et al., 2015).

The survey asked participants several questions to assess their existing knowledge and understanding of how data storage and deletion work, what file shredders are, and how file shredders can be used to erase HIPAA-protected data, as well as to gather information on what factors (training, corporate policy, corporate culture) influence a user's understanding and attitude toward file shredders. Following this, the study asked additional questions to determine what influence, if any, user education and social factors have on the user's overall end behavior when erasing HIPAA-protected data. The study used quantitative rather than qualitative or mixed-method approaches to obtain more measurable results.

Pilot Study

A pilot study was conducted to evaluate the survey's efficacy. The survey was administered to approximately 10 participants (10% of the intended sample size of 100 respondents). SPSS was used to evaluate the results and determine whether the survey questions yielded usable data. To analyze the results, SPSS applied the Pearson Correlation Coefficient to the data gathered from the survey instrument, just as was done in the final analysis, to determine whether a consistent and measurable relationship can be found between the independent and dependent variables and thereby confirm the reliability and validity of the survey and the measurement methods.

Cronbach's Alpha was used to verify the reliability and validity of the pilot study. Cronbach's Alpha is a statistical metric frequently used to assess the reliability and internal consistency of surveys, focusing on whether a study's items effectively measure a single construct. The function assesses the relatedness between items in the questionnaire, yielding a

value between 0 and 1. A value above 0.75 is generally considered indicative of strong internal consistency, meaning that the survey items reliably measure the same construct (Izah et al., 2024; Tavakol & Dennick, 2011; Taber, 2018). Cronbach's Alpha is often used in surveys to assess the internal consistency of responses, making it an ideal tool for assessing survey validity (Izah et al., 2024). In the pilot study, Cronbach's Alpha was 0.941, suggesting high internal consistency.

Operational Definitions of Variables

Training and Social Factors Regarding Secure Data Deletion

The first independent variable was the influence of training on data deletion behavior. Based on the principles of the UTAUT model, social factors such as training can play a significant role in a user's decision to adopt or reject a given technology. As such, the purpose of this variable is to capture what impact this factor may have on user attitudes toward secure data deletion.

User Understanding of Data Storage

The other independent variable was the degree to which a user properly understands how data storage works in a digital medium. The study hypothesizes that users may adjust their approach to deleting secure data as their understanding of data storage and deletion processes improves; therefore, the survey examines whether this understanding influences their behavior when deleting files and, if so, how.

User Behavior When Erasing Data

The dependent variable to be analyzed was users' behavior when erasing data: whether they use secure physical deletion methods and, if so, whether they apply it in all cases or only

when specifically instructed. This is the dependent variable, as the hypothesis is that the two independent variables will influence user behavior in this situation. The study aimed to determine if the independent variables impact this variable and, if so, to what extent.

Study Procedures

The research employed a quantitative survey design, administered via SurveyMonkey, to recruit participants from the healthcare field. Participants were voluntarily recruited online and completed an online survey. The survey began with several screening questions to filter out bots or other potential invalid responses before progressing to the main questions. SPSS was used to analyze results and identify relationships between responses to assess whether users' preexisting understanding (or lack thereof) of how data storage functions influence what they expect to happen when deleting a file on their computer, using the Pearson Correlation Coefficient. The study focused on healthcare workers in the United States, examining how well they understand data deletion and whether this understanding affects their behavior when working with HIPAA-protected data. Yamane's formula suggested that a minimum of 97 respondents were required to achieve a 10% margin of error.

The study asked participants several questions to assess their existing knowledge and understanding of how data storage and deletion work, what file shredders are, and how file shredders can be used to erase HIPAA-protected data, as well as to gather information on what factors (training, corporate policy, corporate culture) influence a user's understanding and attitude toward file shredders. Following this, the study asked additional questions to determine what influence, if any, user education and social factors have on the user's overall end behavior when erasing HIPAA-protected data. Once the data was collected, the tool SPSS was used to examine the results, using the Pearson Correlation Coefficient to determine the relationship, if

any exists, between the independent variables (training and social influences regarding secure data deletion and user understanding of how data storage functions) and the dependent variable (whether the users properly utilize secure erasure tools).

Data Analysis

The software SPSS was used to conduct Pearson correlation testing on the Likert-scale data to assess whether the independent variables influence the dependent variables. The Pearson correlation coefficient measures the degree to which an independent variable changes the dependent variable. The data in question came from a survey that asked users questions to gather information on the three variables to determine whether either (or both) of the independent variables (Training and social influences within an organization regarding secure data deletion and existing user knowledge about how data storage works) impact the dependent variable (user behavior when deleting data) and to what extent. This helped to test the research questions and hypotheses, specifically whether most healthcare users understand what happens to data when it is deleted, whether their understanding or lack of understanding of how data storage functions affect their behavior, and whether there is a statistically significant relationship between a user's understanding and awareness of data storage functions and the actual policies regarding data deletion.

Assumptions

The researcher assumes that a relationship between users' preexisting attitudes and knowledge and their behavior may exist, and that this relationship can be detected and measured. The researcher also assumes that respondents answered the questions honestly, take the survey only once, and truthfully report meeting the eligibility requirements (being a healthcare worker in the United States). SurveyMonkey settings were be utilized to ensure that respondents could

only reply once, where available, and the sample size was intended to be large enough, based on Yamane's formula for calculating the minimum sample size, to prevent dishonest or fraudulent responses from influencing the overall results.

The research assumes that all respondents are real people, not bots. Standard anti-bot measures were implemented to ensure this, including repeating questions to assess consistency and including standard screening items, such as "select the fourth option in this list," in the initial screening before the Likert questions. The survey makes no assumptions about respondents' understanding of technical concepts, as one of its goals is to assess users' levels of understanding.

The survey was designed to avoid leading questions or extraneous details, as this would otherwise increase the risk that respondents might be biased in their responses. It is assumed that respondents understood the survey and its questions and answered them honestly to the best of their ability. Lastly, it is assumed that there may be a relationship between the dependent and independent variables.

Limitations

The study was limited only to healthcare workers in the United States. The individuals selected for the sample were required to understand HIPAA and the provisions governing the protection of patients' personal information. However, due to the need to preserve respondents' anonymity, it was difficult to directly verify that all respondents were healthcare workers in the United States. As such, several screening questions were used at the start of the survey to exclude invalid responses, and a sufficiently large sample size, based on Yamane's formula, was obtained to ensure that invalid responses do not unduly bias the results.

There are several potential threats to validity and reliability. Reliability is compromised if external factors the author did not account for skew the results, or if the measurement method is

inconsistent. If the results are unreliable, they may change in future tests. Failure to control external variables effectively can compromise internal validity, making it difficult for the researcher to determine whether the results are attributable to the variables being measured or to confounding variables (Slocum, 2022). If internal validity is compromised, the study results may not be relevant to the variables under investigation. If the study is not properly designed or employs an incorrect statistical procedure, the conclusion's validity may be compromised (García-Pérez, 2012). Without proper conclusion validity, whether the results indicate a link between the studied factors can be unclear.

To mitigate potential issues with external validity and internal validity, the survey included screening questions to ensure that respondents began from the same baseline: healthcare workers in the United States. To ensure validity, the results are available in an anonymized form, to protect the privacy of the respondents to a public database, along with the full survey questions and SPSS configurations, to support the validity of the conclusions by ensuring that author bias or statistical analysis errors are not present as the data and experimental setup are publicly available, and all steps of the process can be reproduced.

The study may have a lack of generalizability due to a potential selection bias among respondents, as, despite following Yamane’s formula to determine the necessary minimum number of respondents to ensure that the results can be generalized, the survey responses are still necessarily be limited to respondents who chose to register for an online survey program and to reply to the survey. However, some steps were taken to maximize potential generalizability despite potential selection bias. First, the study population has been defined as healthcare workers in the United States, thereby limiting the survey to a population that can be examined in detail. Second, the sample size was calculated using government-provided statistics on the

number of healthcare workers in the United States as input to Yamane's formula, yielding a minimum required sample size to ensure generalizability to that population.

Delimitations

In this study, the survey was intentionally limited to healthcare workers in the United States, as these users are responsible for securing and deleting files containing HIPAA-protected information. The survey uses closed-ended Likert-scale items to assess potential correlations between the two independent variables and the dependent variable. The variables "user knowledge" and "training and other cultural influences" were selected within the UTAUT framework to examine their potential correlation with user behavior when deleting confidential data.

Ethical Assurances

The study results were anonymized to minimize the risk of exposing personal information. Respondents' email addresses were not stored, and no identifying questions were asked. No personally identifiable information was gathered at any point in the survey. Additionally, the researcher did not interact with participants to reduce response bias.

Respondents were recruited via SurveyMonkey and other social media platforms, with consent to participate. The respondents were informed about the purpose and nature of the study, the data that was collected (including a notice that no personally identifiable data would be collected), and the measures that were taken to protect their privacy, and were asked to confirm their informed consent before beginning the survey. Respondents were additionally informed that they could discontinue the survey at any time and would not face any consequences.

Using SurveyMonkey for data collection and analysis, with IP Tracking disabled, eliminated the possibility of personally identifying survey respondents. The survey began with a

consent form explaining the purpose of the study, how the data would be used, and the steps taken to protect user privacy. It also ensured that all respondents provided informed consent and reminded participants that their responses were voluntary and that they could discontinue at any time. The survey's introduction also reassured participants that all data collected would be used solely for this research and would not be misused for other purposes.

The data were collected via SurveyMonkey, which does not attribute responses to specific individuals, ensuring that the research cannot identify individual respondents and thereby providing additional assurance of anonymity. The option to make all responses anonymous, meaning no IP Addresses would be collected, was checked. All collected data, including participant data, was securely stored on a BitLocker-encrypted drive and was accessible only to the researcher to ensure the confidentiality and privacy of respondents. Additionally, the data will be permanently deleted from the drive using a secure file shredder within three years following the completion of the study. Any responses observed to contain identifiable data, such as an email address or name, were immediately deleted using a secure file-shredding tool and are not included in the analysis.

To ensure replicability, the full survey questions and the analysis methodology are detailed in the study's appendix. Before data collection, the study, including all survey questions, data safeguards, privacy concerns, recruitment methods, and informed consent, received approval from the National University Dissertation Committee and the National University Institutional Review Board (IRB).

Summary

Failure to properly erase HIPAA-protected data can result in serious data breaches. To better understand what factors influence how users behave when working with confidential data

and how organizations can better influence their employees to ensure that HIPAA-protected data is securely erased in a nonrecoverable fashion, this study performed a quantitative survey to gather information on two independent variables (what social factors exist with organizations regarding secure data deletion, and what sort of understanding users have about how data storage and deletion functions) and one dependent variable (how users behave when deleting HIPAA-protected data) to determine what relationship, if any, exists between the dependent and independent variables. The survey was administered to 112 healthcare workers in the United States; a pilot study of 10 participants was conducted to confirm the validity of the survey items. The results were analyzed using SPSS, with Pearson correlation coefficients used to assess relationships between variables. The research results should help healthcare organizations adjust training and policies to protect confidential data from data breaches. In Chapter 4, the survey results are presented and analyzed.

Chapter 4: Findings

The problem addressed by this study was the challenge of user handling of storage data in the healthcare industry (Carlton, 2005; Shamlawi, 2018). Users of digital devices often fail to properly erase data when donating or discarding the device, resulting in the second-hand market containing a large volume of supposedly empty devices that may contain personal or confidential data that can be quickly recovered. This is because users often assume that deleting data on a drive permanently erases it, unaware that the data is merely hidden rather than removed. Existing studies indicate that many second-hand devices still contain personally identifiable or confidential data that can be recovered using standard forensic tools (Jones et al., 2016; Osawaru, 2024; Sutherland et al., 2010).

In many of the cases studied, even after the drives had been formatted, researchers were still able to recover sufficient data to identify the former owners personally. This indicates that the original owners of the devices believed they had taken the necessary steps to delete their data before selling or donating the devices and were therefore unaware that their personally identifiable data could be recovered. This suggests that users lack a proper understanding of how data storage works, or that existing data management tools fail to effectively communicate how to delete their data, thereby implying that deletion or formatting is sufficient. This misperception may influence how users handle confidential data. In the healthcare industry, when laws often dictate that certain information must be deleted thoroughly, leaving the data in this recoverable form exposes healthcare professionals and their patients to threats, from the loss of private data to fines or legal penalties (Heath et al., 2022; Jones et al., 2016; Shamlawi, 2018).

The purpose of this quantitative, non-experimental survey-based research was to determine what healthcare workers prioritize in file management systems when deleting data for HIPAA compliance and to address the existing gap in research regarding the relationship

between what users want from file systems and how they believe the systems work. Users' needs for both privacy and convenience, as well as their desire to reverse accidental deletions, pose a significant dilemma for operating system designers, and a comprehensive study of what average users require has not yet been conducted. This issue affects users across the board, including average home users and corporations, who face threats from improperly deleted data (Hadi, 2016; Jones et al., 2016). This is the dilemma developers face when seeking to appease contradictory user goals. Further research is needed into users' needs and the broader social views on data privacy and retention to better understand how general users want operating systems to handle their data and how these factors influence user behavior when deleting HIPAA-protected files. Based on the results of such research, designers seeking to enhance their data storage code would have a firmer baseline from which to work.

This chapter begins by presenting evidence of the data's validity and reliability. It then presents the study's results, followed by an examination of how these results relate to each research question. The chapter concludes with an evaluation and interpretation of the findings.

Validity and Reliability of the Data

Cronbach's Alpha was used to verify the reliability and validity of the pilot study. Cronbach's Alpha is a statistical metric frequently used to assess survey reliability and internal consistency, focusing on whether a study's questions effectively measure a single, unique concept. The function assesses the relatedness between items on the questionnaire, resulting in a value between 0 and 1. A value above 0.75 is generally considered indicative of strong internal consistency, meaning that the survey instrument's items reliably measure the same construct (Izah et al., 2024; Taber, 2018; Tavakol & Dennick, 2011). Cronbach's Alpha is often used in surveys to assess the internal consistency of responses, making it an ideal tool for evaluating

survey validity (Izah et al., 2024). In this study, Cronbach's alpha was 0.811, indicating strong internal consistency.

Results

Before evaluating the results, invalid responses (replies with no answers to any question) were deleted. After removing the invalid responses, 112 replies remained, more than the 97 needed by Yamane's formula. As such, Yamane's formula indicates that sufficient responses were obtained to ensure a 10% margin of error.

The first set of questions in the survey was intended to assess users' preexisting beliefs about data storage and whether they had received training that may have shaped these beliefs. The purpose was to assess the extent to which respondents agreed or disagreed with several statements regarding their prior knowledge and training. This data can then be compared to later responses about actual understanding, as well as to whether the statements were correct or incorrect, to establish how perceived knowledge is related to actual knowledge and user behavior.

The results of questions 1 through 5 were intended to establish a baseline understanding of how confident respondents felt in their knowledge of data storage and deletion, as well as whether they believed their existing training had adequately prepared them for dealing with data deletion, to obtain a better view of social factors – training and existing beliefs – that could influence behavior when working with data deletion. To accompany this, the second set of questions (6-10) was intended to elicit respondents' beliefs about how data storage and deletion functions. For the first research question, the combination of these two sets of questions was used to establish whether a majority of respondents appear to understand how data storage and

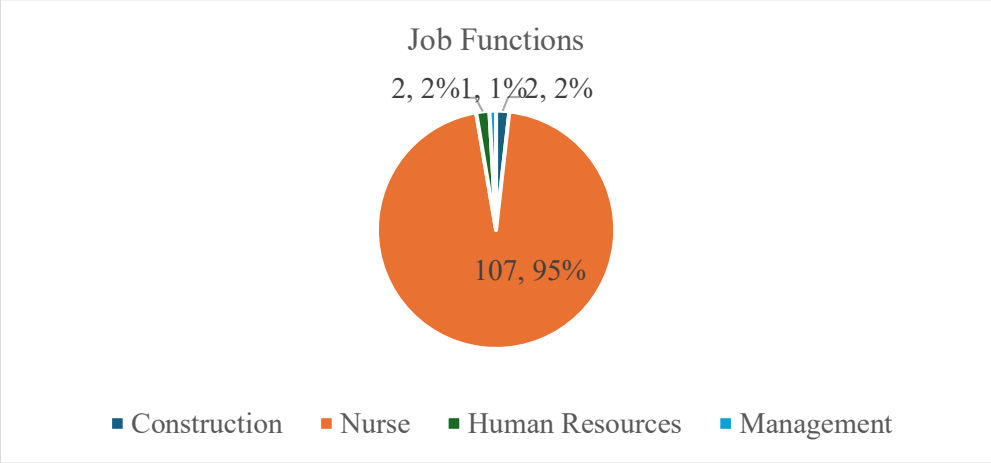
data deletion tools function, as well as whether existing beliefs in knowledge or training appear to influence the confidence of a respondent's belief in their knowledge.

Demographics

Of the respondents, forty-four (39.29%) were male, and sixty-eight (60.71%) were female. One hundred and seven (95.54%) worked as nurses, one (0.89%) worked in management, two (1.79%) worked in human resources, and two (1.79%) listed themselves as working in construction. All 112 (100%) listed their industry as healthcare. All 112 respondents were located in the United States. Fifteen respondents (13.39%) reported ages between 18 and 29, fifty (44.64%) between 30 and 44, thirty-two (28.57%) between 45 and 60, and fifteen (13.39%) over 60.

Figure 2

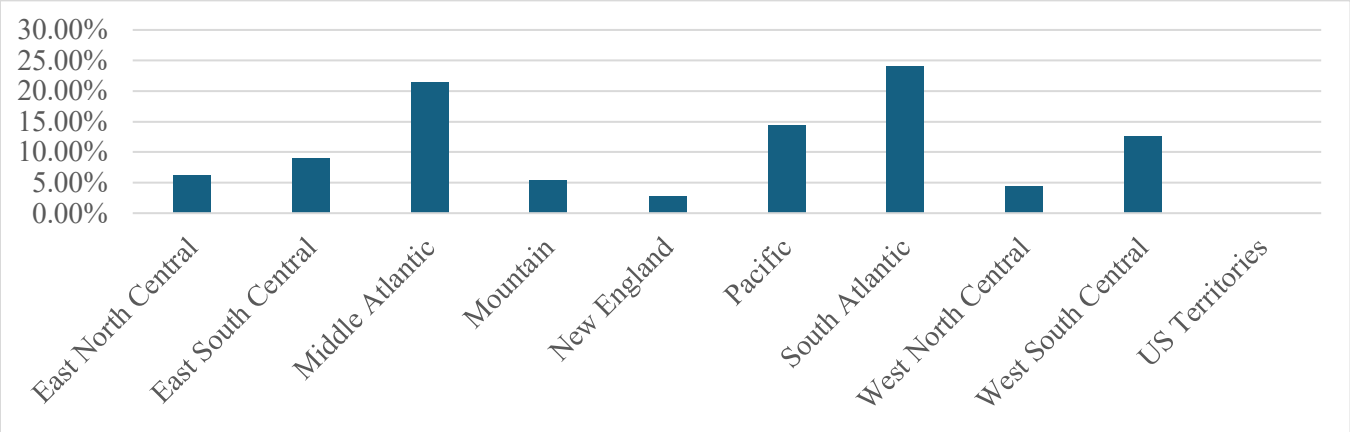
Respondents' Job Functions



This figure shows the reported job functions of the respondents. One hundred and seven out of the total one hundred and twelve reported themselves to be nurses. Two listed themselves as construction, two as human resources, and one as management. All worked in the United States healthcare industry.

Figure 3

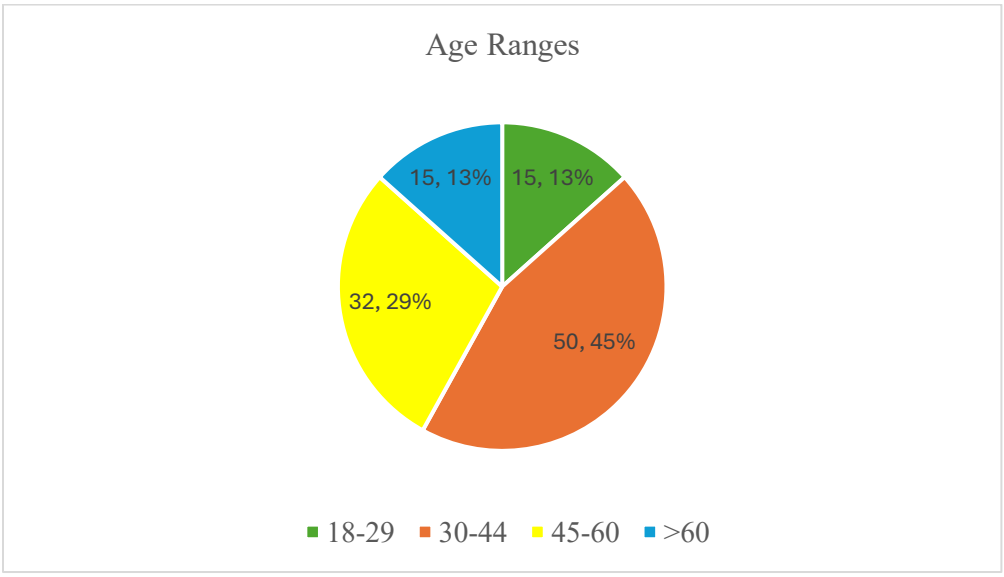
Respondents' US Region



This figure shows the reported regions in the United States where survey respondents lived. The majority were in either the South Atlantic (24.11%) or the Middle Atlantic (21.43%), followed by the Pacific (14.29%) and the West South Central (12.50%). No respondents were located outside of the United States or in the U.S. territories.

Figure 4

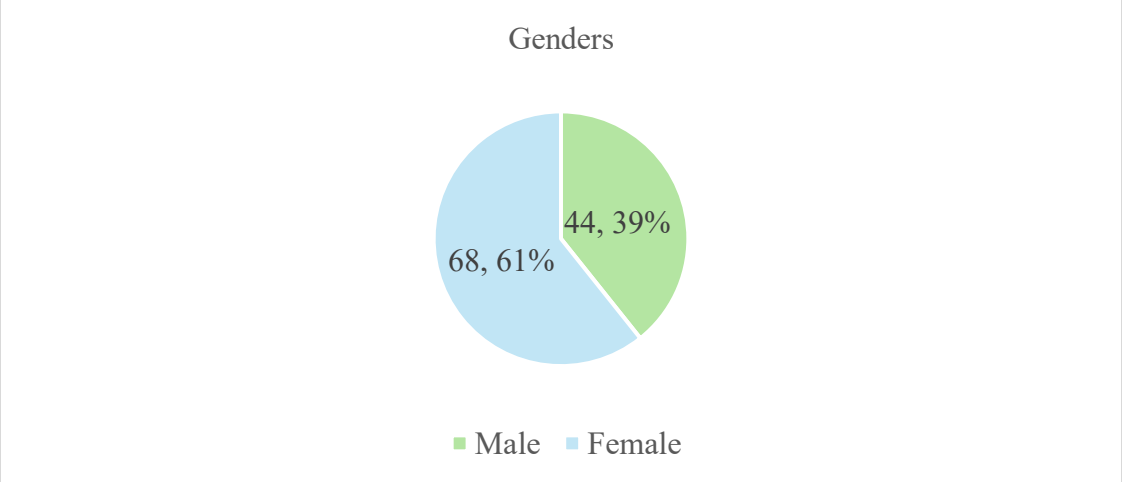
Respondents' Ages



This chart shows the age ranges of survey respondents. The majority of respondents (44.64%) were between ages 30 and 44. The second-largest group of respondents (28.57%) was between ages forty-five and sixty, and the remainder were evenly split between ages eighteen to twenty-two and over sixty.

Figure 5

Respondents' Genders



This chart shows the self-reported genders of survey respondents. The majority of respondents (60.71%) were female. The remainder (30.29%) were male.

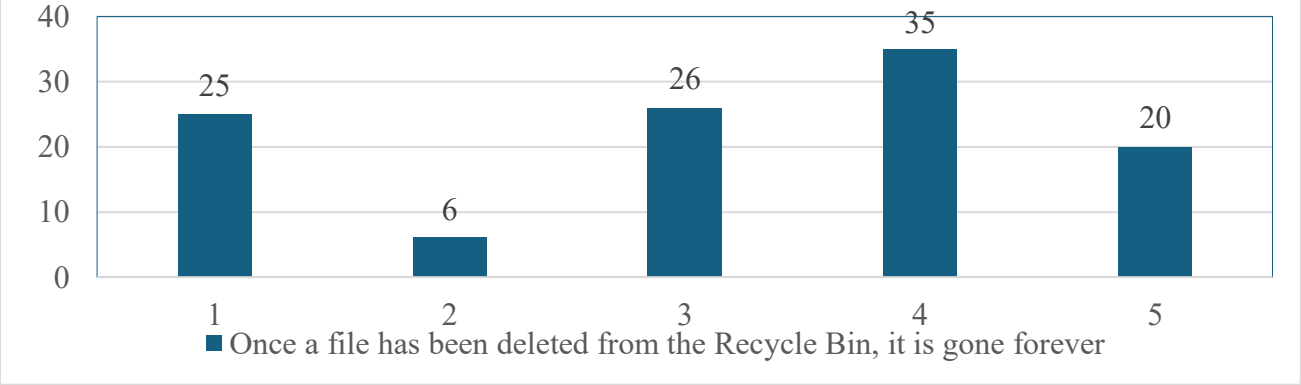
Research Questions

Research Question 1

Do a statistical majority of healthcare workers understand how to erase HIPAA-protected files?

Figure 6

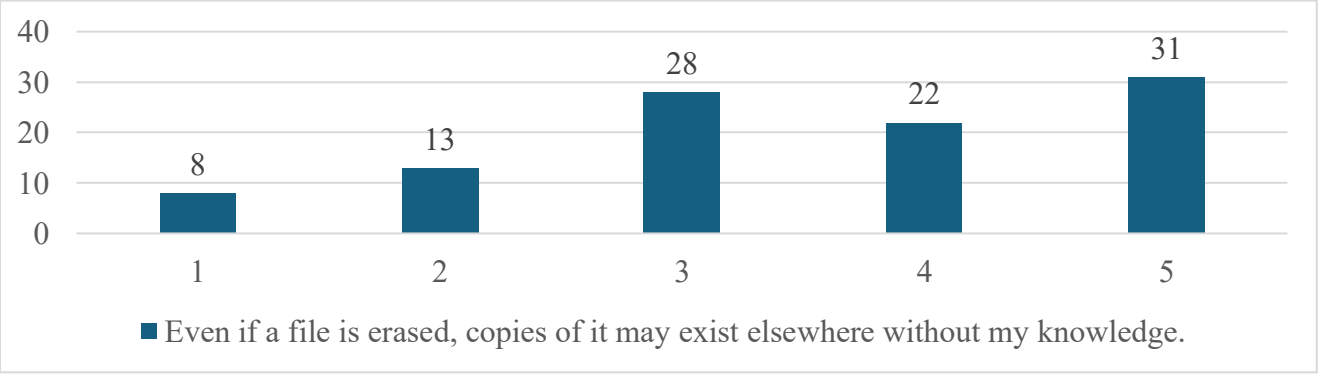
Belief that Copies of Files They are Unaware of may Exist Elsewhere on a Drive



As shown in Figure 6, 50% of respondents believe that deleting a file from the Recycle Bin removes it forever. Twenty respondents (18.18%) strongly agreed with the statement, thirty-five (31.82%) somewhat agreed, and twenty-six (23.64%) were unsure. Only six respondents (5.54%) slightly disagreed, and twenty-five respondents (20.96%) strongly disagreed, for a total of only 26.5% disagreeing. However, deleting a file from the Recycle Bin merely removes the pointer to that file and flags the sector of the drive containing it as "free," but does not overwrite or otherwise remove the file, meaning that files deleted from the Recycle Bin can be recovered with standard forensic tools (Al Sharif et al., 2014; Weijers, 2022).

Figure 6

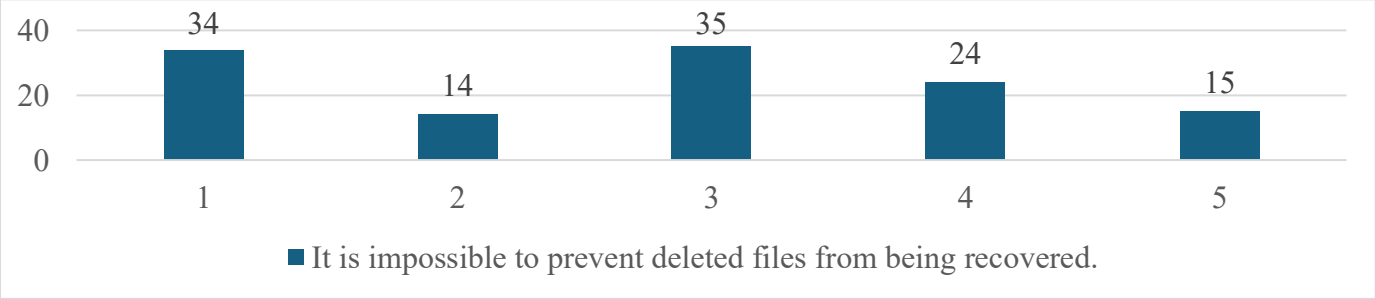
Belief that Copies of Files They are Unaware of may Exist Elsewhere on a Drive



As shown in Figure 7, 55.45% of respondents believe that even when a file is fully erased, other copies of it may exist elsewhere on the drive that they are not aware of. Thirty-two respondents (29.09%) slightly agreed with this statement, thirty-one (26.36%) strongly agreed, twenty-eight (25.45%) were unsure, thirteen (11.82%) slightly disagreed, and eight (7.27%) strongly disagreed. Often, drives create copies of the file that the user is unaware of, such as prefetch caches, mirrored backups, or shadow copies, and these copies are not deleted when the main file is deleted (Weijers, 2022).

Figure 7

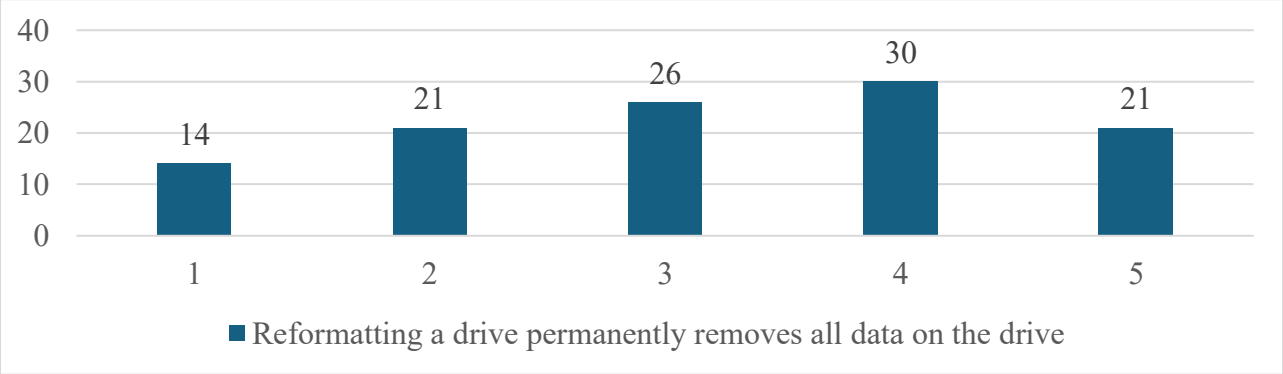
Belief that it is Impossible to Prevent Deleted Files from Being Recovered



As shown in Figure 8, 41.82% of respondents do not believe it is impossible to prevent data from being recovered. Thirty-four respondents (29.09%) strongly disagreed, fourteen respondents (12.73%) slightly disagreed, and thirty-five respondents (22.73%) were unsure. Twenty-four respondents (21.82%) slightly agreed, and fifteen respondents (13.64%) strongly agreed, for a total of 35.46% agreeing. Deleted data can be fully purged using methods such as file shredders, which involve repeatedly overwriting the data with other data (Nahar et al., 2018; Weijers, 2022).

Figure 8

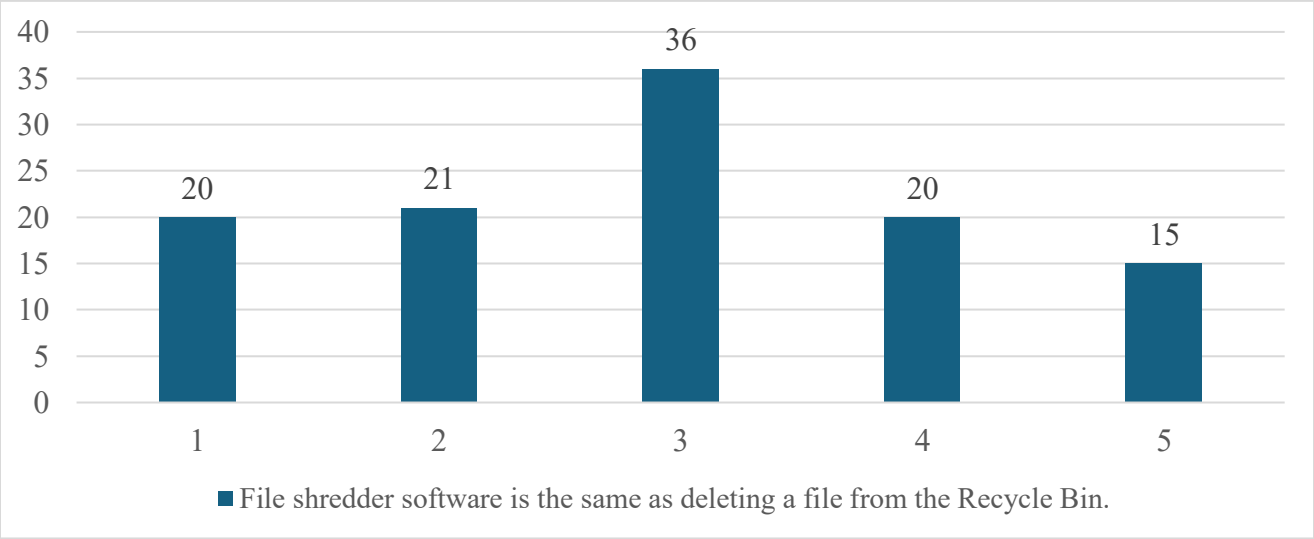
Belief that Reformatting a Drive Permanently Removes all Data



As shown in Figure 10, 46.36% of respondents believe that reformatting a drive erases all data on the drive, though responses were split. Twenty-one respondents (19.09%) strongly agreed, and thirty respondents (27.27%) slightly agreed, while twenty-one respondents (17.27%) slightly disagreed and fourteen respondents (22.72%) strongly disagreed (for a total of 39.99% disagreeing), with twenty-six respondents (23.64%) unsure. Reformatting a drive can make data recovery more difficult. However, with appropriate forensic tools, data can still be recovered from a reformatted drive if the reformatting has not overwritten it, unless the reformatting process itself overwrites all data on the drive (Dillon, 2006; Gyening, 2022).

Figure 9

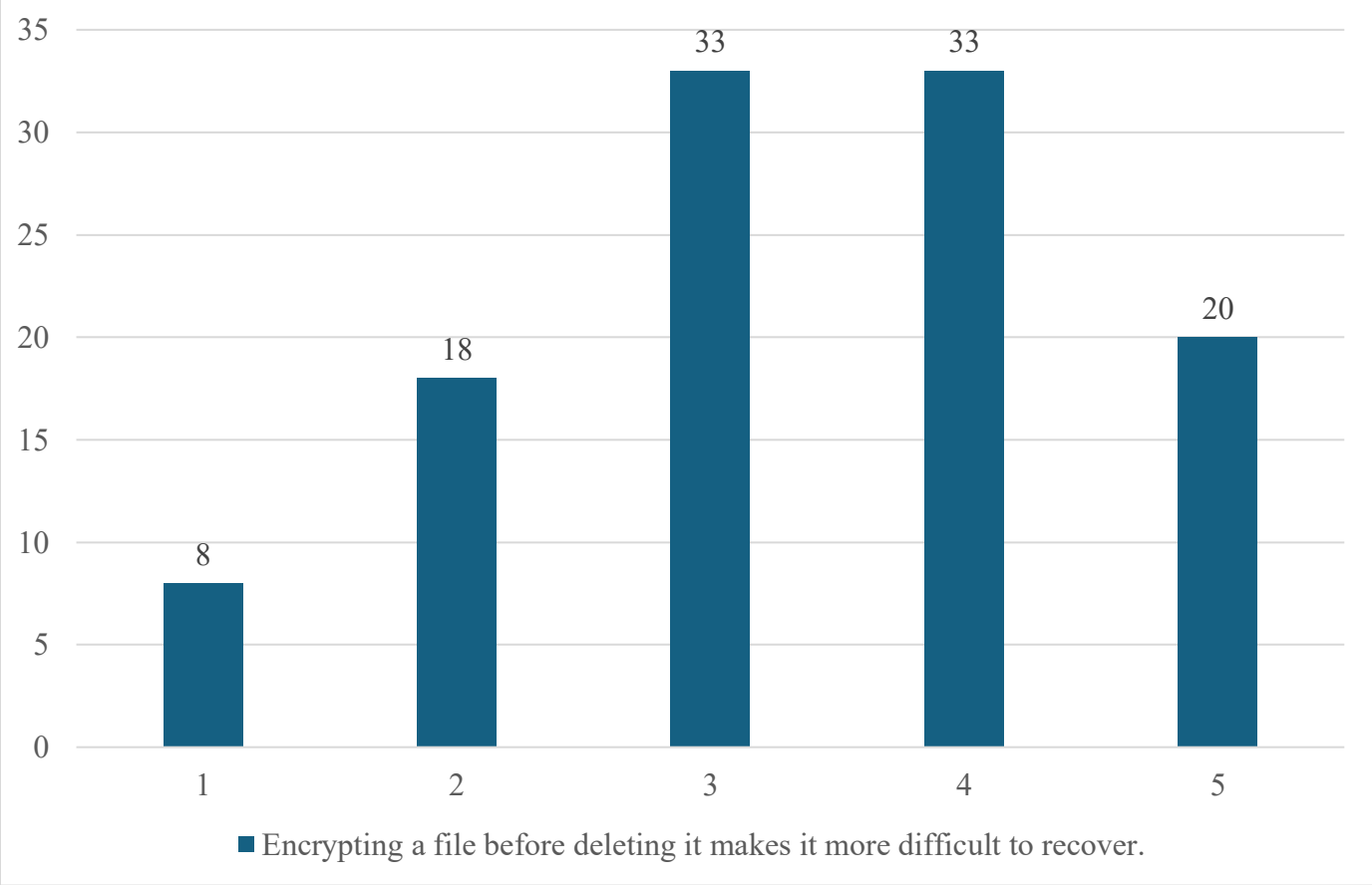
Belief that the Recycle Bin is the Same as a File Shredder



As shown in Figure 10, thirty-six respondents (32.73%) were unsure whether file shredder software was the same as emptying the Recycle Bin. For the remainder, fifteen respondents (13.64%) strongly agreed they were the same, twenty respondents (18.18%) slightly agreed (for a total of 31.82% agreeing), twenty-one respondents (19.09%) slightly disagreed, and twenty respondents (13.64%) strongly disagreed, for a total of 32.73% disagreeing. Emptying the Recycle Bin deletes the pointer to the file without touching the file itself, whereas file-shredder software overwrites the deleted file multiple times with other data to ensure it is fully erased (Al Sharif et al., 2014; Nahar et al., 2018; Weijers, 2022).

Figure 10

Belief that Encrypted Files are More Difficult to Recover.



As shown in Figure 11, 59% of respondents either slightly agreed that encrypted files are more difficult to recover or were unsure. Thirty-three respondents (29.5%) selected 4 (slightly agree), and another thirty-three respondents (29.5%) selected 3 (unsure). Twenty respondents (17.9%) strongly agreed, eighteen respondents (16.1%) slightly disagreed, and eight respondents (7.1%) strongly disagreed. Disk encryption is often used as an alternative to overwriting in Secure Erase technologies where proper file shredders are unavailable, though this is only effective if the encrypted version overwrites the unencrypted version rather than simply coexisting on the same drive (Nahar et al., 2018; Weijers, 2022).

Table 1

Correlations Between Perceived Familiarity with File Shredders and Confidence in Beliefs about File Shredders

		Survey Q1 (N 112)
I am familiar with file shredder software.	Pearson Correlation	.279
	Sig. (2-tailed)	.003**

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "I am familiar with file shredder software" with their level of agreement with the survey question 1 "File shredder software is the same as deleting a file from the Recycle Bin." The purpose was to determine whether there was a correlation between a respondent's perceived level of familiarity with file shredder software and the strength of their beliefs about how the software works. There was a weak positive correlation of r =0.279 with a 2-tailed significance of p = 0.003, significant at the < 0.01 level, suggesting a relationship between users having a preexisting belief that they are familiar with file shredder utilities and users expressing confidence in a belief about whether the Recycle Bin is a form of file shredder utility.

Table 2

Correlations Between Perceived Familiarity with how Data Storage Functions and Confidence in Understanding of the Recycle Bin

		Survey Q6 (N 112)
I understand how digital data storage functions on a technical level.	Pearson Correlation	.095
	Sig. (2-tailed)	.320
	N	112

This Pearson Correlation was performed by examining the association between respondents' level of agreement with the statement "I understand how digital data storage functions on a technical level" and their level of agreement with the statement "Once a file has been deleted from the Recycle Bin, it is gone forever." The purpose was to determine whether there was a correlation between a respondent's perceived level of familiarity with how data storage functions and the strength of their beliefs about how deleting files works. There was a weak positive correlation of $r = 0.095$ with a 2-tailed significance of $p = 0.320$, which is not statistically significant, suggesting that it is unlikely to be a strong predictor of perceived knowledge about how data storage functions and whether users feel confident that deleting a file from the Recycle Bin will permanently remove the file.

Table 3

Correlations Between Perceived Familiarity with how Data Storage Functions and Understanding of the Recycle Bin

		Survey Q10 (N 112)
My job training covered how to securely erase digital data.	Pearson Correlation	.192*
	Sig. (2-tailed)	.042
	N	112

*. Correlation is significant at the 0.05 level (2-tailed).

This Pearson Correlation was performed by examining the association between respondents' level of agreement with the statement "My job training covered how to securely erase digital data" and their level of agreement with the statement "It is impossible to prevent deleted files from being recovered." The purpose was to determine whether there was a

correlation between a respondent's perceived level of familiarity with how data storage functions and the strength of their beliefs about how deleting files works. There was a weak positive correlation of $r = 0.192$ with a 2-tailed significance of $p = 0.042$, significant at the 0.05 level, suggesting a small positive relationship between users' confidence that they were adequately trained in data deletion and their confidence that it is possible to erase data permanently.

Table 4

Correlations Between Perceived Ability to Prevent Data from Being Recovered and Strength in Belief about Whether it is Possible to Prevent Data from Being Recovered

		Survey Q10 (N 112)
I understand how to ensure a deleted file cannot be recovered.	Pearson Correlation	.116
	Sig. (2-tailed)	.224
	N	112

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "I understand how to ensure a deleted file cannot be recovered," with their level of agreement with the statement "It is impossible to prevent deleted files from being recovered." The purpose was to determine whether there was a correlation between a respondent's perceived level of familiarity with secure deletion procedures and the strength of their beliefs about how deleting files works. There was a weak positive correlation of $r = 0.116$ with a 2-tailed significance of $p = 0.224$, which is not statistically significant, suggesting a that there is unlikely to be a relationship between the strength of a user's belief about whether they know how to securely erase a file and their belief about whether secure erase protocols will fully prevent the data from being recoverable.

Table 5

Correlations Between Perceived Confidence in Training to Delete Confidential Data and Belief in the Need to use File Shredder Software to Erase Confidential Data

		Survey Q12 (N 112)
My job training covered how to securely erase digital data.	Pearson Correlation	.294
	Sig. (2-tailed)	.002**
	N	112

**. Correlation is significant at the 0.01 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "My job training covered how to securely erase digital data," with their level of agreement with the statement "I should use a shredder program to erase files containing HIPAA-protected data." The purpose was to determine whether there was a correlation between a respondent's belief that they had been trained to securely erase HIPAA-protected files and the strength of their belief that they should use file shredders to erase confidential files. There was a moderate positive correlation of $r = 0.294$ with a 2-tailed significance of $p = 0.002$, significant at the < 0.01 level, suggesting that there is a positive relationship between whether users believe they were trained in secure file deletion and the strength of their beliefs that they should use file shredder utilities.

Table 6

Correlations Between Perceived Understanding of Data Storage and the Strength of Beliefs that Reformatting a Drive Erases the Data on that Drive

		Survey Q7 (N 112)
I understand how digital data storage functions on a technical level.	Pearson Correlation	.296
	Sig. (2-tailed)	.002**

	N	112

**. Correlation is significant at the 0.01 level (2-tailed).

This Pearson Correlation was performed by examining the association between respondents' level of agreement with the statement "I understand how digital data storage functions on a technical level" and their level of agreement with the statement "Reformatting a drive permanently removes data on that drive." The purpose was to determine whether a respondent's belief that they understood how data storage worked correlated with the strength of their belief that reformatting a drive would permanently remove the data. There was a moderate positive correlation of $r = 0.296$ with a 2-tailed significance of $p = 0.002$, significant at the < 0.01 level, suggesting that there is a positive relationship between whether users believe they understand how data storage functions and the strength of their belief that reformatting a drive will fully erase all data on the drive.

Table 7

Correlations Between Perceived Understanding of Data Storage and the Strength of Belief that File Shredder Utilities Should be Used Before Discarding a Drive

		Survey Q16 (N 112)
I understand how digital data storage functions on a technical level.	Pearson Correlation	.206
	Sig. (2-tailed)	.029*
	N	112

*. Correlation is significant at the 0.05 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "I understand how digital data storage functions on a technical level," with their level of agreement with the statement "Before discarding an old drive, a file shredder utility should be run on the drive." The purpose was to determine whether there was a

correlation between a respondent's belief that they understood how data storage functioned and the strength of their belief that they should use file shredders to erase confidential files. There was a weak positive correlation of $r = 0.206$ with a 2-tailed significance of $p = 0.029$, significant at the 0.05 level, suggesting that there may be a positive relationship between whether users believe they understand how data storage functions and the strength of their belief that file shredders should be used to erase a drive prior to disposal.

Research Question 2

To what degree does a user's understanding of how data storage works influence their behavior when deleting HIPAA-protected files?

Table 8

Correlations Between Beliefs About Recycle Bin Deletions and the Need for File Shredders

		Survey Q16 (N 112)
Once a file has been deleted from the Recycle Bin, it is gone forever.	Pearson Correlation	-.015
	Sig. (2-tailed)	.875
	N	112

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "once a file has been deleted from the Recycle Bin, it is gone forever" with their level of agreement with the statement "before discarding an old drive, a file shredder utility should be run on the drive." The purpose was to determine whether the belief that removing a file from the Recycle Bin would permanently erase the file would influence whether users would be likely to also use file shredder utilities before discarding drives. There was a weak negative correlation of $r = -0.015$ with a 2-tailed significance of $p = 0.875$, which is not statistically significant, suggesting that participants who believe that emptying the Recycle

Bin permanently deletes files may be slightly less likely to see the need to use a file shredder utility, though the low significance suggests that users are unlikely to alter their behavior in either direction based on their existing beliefs.

Table 9

Correlations Between beliefs about the Recycle Bin and the Need for File Shredders to Delete Specific Files

		Survey Q12 (N 112)
Once a file has been deleted from the Recycle Bin, it is gone forever.	Pearson Correlation	.259
	Sig. (2-tailed)	.006**
	N	112

**. Correlation is significant at the 0.01 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "once a file has been deleted from the Recycle Bin, it is gone forever" with their level of agreement with the statement "I should use shredder program to erase files containing HIPAA-protected data." The purpose was to determine whether a belief that removing a file from the Recycle Bin would permanently erase the file would influence whether the user would be likely to also use file shredder utilities to erase HIPAA-protected files. There was a weak positive correlation of r =.259, with a 2-tailed significance of p = 0.006, significant at the < 0.01 level, suggesting that respondents who believe that removing files from the recycle bin permanently removes the files still believe that file shredder software should be used as the method for erasing HIPAA-protected data.

Table 10

Correlations Between Beliefs that Copies of Files may Exist on a Drive and Beliefs that Drives Should be Encrypted before Discarding

		Survey Q15 (N 112)
Even if a file is erased, copies of it may exist elsewhere on the drive without my knowledge.	Pearson Correlation	.079
	Sig. (2-tailed)	.410
	N	112

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "Even if a file is erased, copies of it may exist elsewhere on the drive without my knowledge" with their level of agreement with the statement "Before discarding an old drive, the drive should be encrypted." The purpose of this was to determine whether a belief in the presence of leftover data would influence whether users would choose to encrypt a drive before discarding it. The weak positive reported correlation was $r = 0.079$, with a significance of $p = 0.410$, which is not statistically significant, suggesting that there is unlikely to be a correlation between whether users believe that copies of files may exist elsewhere on the drive without their knowledge and whether those users believe they should encrypt drives before discarding them.

Table 11

Correlations Between Beliefs that Copies of Files may exist on a Drive and Beliefs that Drives Should be Reformatted Before Discarding

		Q14 (N 112)
Even if a file is erased, copies of it may exist elsewhere on the drive without my knowledge.	Pearson Correlation	.260
	Sig. (2-tailed)	.006**
	N	112

**. Correlation is significant at the 0.01 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "Even if a file is erased, copies of it may exist elsewhere on the

drive without my knowledge" with their level of agreement with the statement "Before discarding an old drive, the drive should be reformatted." The purpose of this was to determine whether a belief in the presence of leftover data would influence whether users would choose to reformat a drive before discarding it. The reported correlation was a weak positive correlation of $r = 0.260$, with a significance of $p = 0.006$, significant at the 0.01 level, suggesting that respondents who believe that copies of a file may exist elsewhere on the drive without their awareness are more likely to reformat a drive before discarding it.

Table 12

Correlations Between Beliefs that Copies of files may Exist on a Dive and Beliefs that File Shredders Should be run on a Drive Before Discarding

		Q16 (N 112)
Even if a file is erased, copies of it may exist elsewhere on the drive without my knowledge.	Pearson Correlation	.347**
	Sig. (2-tailed)	<.001
	N	112

**. Correlation is significant at the 0.01 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "Even if a file is erased, copies of it may exist elsewhere on the drive without my knowledge" with their level of agreement with the statement "Before discarding an old drive, a file shredder utility should be run on the drive." The purpose of this was to determine whether the belief that leftover data remains would influence whether users choose to run a file shredder on the entire drive before discarding it. The reported correlation was a moderate positive correlation of $r = 0.347$, with a significance of $p < 0.001$, significant at the 0.01 level, suggesting that respondents who believe that copies of a file may exist elsewhere on

the drive without their awareness are more likely to use file shredder software to erase a drive before discarding it.

Table 13

Correlations Between Beliefs that Encrypting a File Makes it More Difficult to Recover, and Beliefs that Drives Should be Encrypted Before Discarding

		Q15 (N 112)
Encrypting a file before deleting it makes it more difficult to recover.	Pearson Correlation	.214
	Sig. (2-tailed)	.023*
	N	112

*. Correlation is significant at the 0.05 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "Encrypting a file before deleting it makes it more difficult to recover" with their level of agreement with the statement "Before discarding an old drive, the drive should be encrypted." The purpose of this was to determine whether a belief that encrypting files makes them harder to recover correlates with a propensity to encrypt drives before discarding them. The reported correlation was a weak positive correlation of $r = 0.214$, with a significance of $p = 0.023$, significant at the 0.05 level, suggesting that respondents who believe that encrypting files makes them harder to recover are more likely to encrypt the drive prior to discarding it.

Table 14

Correlations Between Beliefs that Encrypting a File Makes It More Difficult to Recover, and Beliefs that Files Should be Encrypted before Deleting

		Q13 (N 112)

Encrypting a file before deleting it makes it more difficult to recover.	Pearson Correlation	.250
	Sig. (2-tailed)	.008**
	N	112

**. Correlation is significant at the 0.01 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "Encrypting a file before deleting it makes it more difficult to recover" with their level of agreement with the statement "Prior to deleting HIPAA-protected data, the files containing that data should be encrypted." The purpose of this was to determine whether a belief that encrypting files makes them harder to recover correlates with a propensity to encrypt HIPAA-protected files prior to deleting them. The reported correlation was a weak positive correlation of r =0.250, with a significance of p = 0.008, significant at the 0.01 level, suggesting that respondents who believe that encrypting files makes them harder to recover are more likely to encrypt HIPAA-protected files before deleting them.

Table 15

Correlations Between Beliefs that it is Impossible to Prevent Files from Being Recovered and that File Shredder Utilities Should be Used Prior to Discarding Drives

		Q16 (N 112)
It is impossible to prevent deleted files from being recovered.	Pearson Correlation	.014
	Sig. (2-tailed)	.882
	N	112

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "It is impossible to prevent deleted files from being recovered" with their level of agreement with the statement "Before discarding an old drive, a file shredder

utility should be run on the drive." The purpose of this was to determine whether a belief that there is no way to stop files from being recovered correlates with a belief that file shredders should be used on drives before discarding them. The reported correlation was a weak positive correlation, $r = 0.014$, with a significance of $p = 0.882$, which is not statistically significant, suggesting that respondents who believe it is impossible to prevent files from being recovered are unlikely to alter their behavior to use file shredder utilities.

Table 16

Correlations Between Beliefs that it is Impossible to Prevent Files from Being Recovered and that Drives Should be Encrypted Prior to Being Discarded

		Q15 (N 112)
It is impossible to prevent deleted files from being recovered.	Pearson Correlation	.152
	Sig. (2-tailed)	.109
	N	112

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "It is impossible to prevent deleted files from being recovered" with their level of agreement with the statement "Before discarding an old drive, the drive should be encrypted." The purpose of this was to determine whether a belief that there is no way to stop files from being recovered correlates with a belief that old drives should be encrypted before being discarded. The reported correlation was a weak positive correlation, $r = 0.152$, with a significance of $p = 0.109$, which is not statistically significant, suggesting that respondents who believe it is impossible to prevent files from being recovered are unlikely to alter their behavior to encrypt drives prior to discarding them.

Table 17

Correlations Between Beliefs that it is Impossible to Prevent Files from Being Recovered and that HIPAA-Protected Files Should be Encrypted Before Deletion

		Q13 (N 112)
It is impossible to prevent deleted files from being recovered.	Pearson Correlation	.230
	Sig. (2-tailed)	.015*
	N	112

*. Correlation is significant at the 0.05 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "It is impossible to prevent deleted files from being recovered" with their level of agreement with the statement "Prior to deleting HIPAA-protected data, the files containing that data should be encrypted." The purpose of this was to determine whether a belief that there is no way to stop files from being recovered correlates with a propensity to encrypt HIPAA-protected files prior to deleting them. The reported correlation was a weak positive correlation of r =0.230, with a significance of p = 0.015, significant at the 0.05 level, suggesting that users who believe that it is impossible to prevent deleted files from being recovered may be slightly more likely to encrypt HIPAA-protected files prior to deleting them.

Table 18

Correlations Between Beliefs that it is Impossible to Prevent Files from Being Recovered and that Drives Should be Reformatted Before Discarding

		Q14 (N 112)
It is impossible to prevent deleted files from being recovered.	Pearson Correlation	.185
	Sig. (2-tailed)	.051
	N	112

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "It is impossible to prevent deleted files from being recovered" with their level of agreement with the statement "Before discarding an old drive, the drive should be reformatted." The purpose of this was to determine whether a belief that there is no way to stop files from being recovered correlates with a belief that drives should be reformatted prior to disposal. The reported correlation was a weak positive correlation of $r = 0.185$, with a significance of $p = 0.051$, which is not statistically significant at the 0.05 or 0.01 levels, suggesting that respondents who believe it is impossible to prevent files from being recovered are unlikely to alter their behavior to reformat drives prior to discarding them.

Table 19

Correlations Between Beliefs that Reformatting Drives permanently Erases Data and Beliefs that Drives Should be Reformatted Before Discarding

		Q14 (N 112)
Reformatting a drive permanently removes data on that drive.	Pearson Correlation	.488
	Sig. (2-tailed)	<.001 **
	N	112

**. Correlation is significant at the 0.01 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "Reformatting a drive permanently removes data on that drive" with their level of agreement with the statement "Before discarding an old drive, the drive should be reformatted." The purpose of this was to determine whether a belief that reformatting a drive entirely erases all data on the drive correlates with a belief that drives should be reformatted prior to disposal. The reported correlation was a moderate positive correlation of $r = 0.488$, with a significance of $p < 0.001$, significant at the 0.01 level, suggesting that respondents who believe

reformatting permanently removes all data from the drive are more likely to reformat old drives before discarding them.

Table 20

Correlations Between Beliefs that Reformatting Permanently Erases Data and Beliefs that File Shredders Should be Used on Drives Prior to Discarding Them

		Q16 (N 112)
Reformatting a drive permanently removes data on that drive.	Pearson Correlation	.301
	Sig. (2-tailed)	.001**
	N	112

**. Correlation is significant at the 0.01 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "Reformatting a drive permanently removes data on that drive" with their level of agreement with the statement "Before discarding an old drive, a file shredder utility should be run on the drive." The purpose of this was to determine whether a belief that reformatting a drive entirely erases all data on the drive correlates with a belief that drives should also have file shredders run on them prior to disposal. The reported correlation was a moderate positive correlation of $r = 0.301$, with a significance of $p = 0.001$, significant at the 0.01 level, suggesting that respondents who believe reformatting permanently removes all data from the drive are more likely to use file shredder utilities on old drives before discarding them.

Table 21

Correlations Between Beliefs that File Shredders and the Recycle Bin are the Same Thing and Beliefs that the Recycle Bin Erases Files Forever

		Q8 (N 112)
Once a file has been deleted from the Recycle Bin, it is gone forever.	Pearson Correlation	.328
	Sig. (2-tailed)	<.001**
	N	112

**. Correlation is significant at the 0.01 level (2-tailed).

This Pearson Correlation was performed by correlating whether the respondents' level of agreement with the statement "Once a file has been deleted from the Recycle Bin, it is gone forever," with their level of agreement with the statement "File shredder software is the same as deleting a file from the Recycle Bin." The purpose of this was to determine whether users who believe deleting a file from the Recycle Bin will permanently erase the file also believe that the Recycle Bin is a form of file shredder. The reported correlation was a moderate positive correlation of $r = 0.328$, with a significance of $p < 0.001$, significant at the 0.01 level, suggesting that respondents who believe that emptying the Recycle Bin is sufficient to erase the files also believe that the Recycle Bin is a form of file shredder software.

Research Question 3

Is there a statistically significant relationship between healthcare workers' understanding of how files are deleted and how they should be deleted?

The Pearson Correlation results for RQ2 also examined the level of significance of the correlations. Based on the results presented above in RQ2, there was a statistically significant relationship between many of the correlations. As shown in Table 9, the correlation was statistically significant ($p = 0.006$) between respondents' beliefs that deleting files from the Recycle Bin erases them forever and their belief that file shredder programs should be used to delete HIPAA-protected data. As seen in Table 21, there was likewise a high degree of statistical significance ($p < 0.001$) in the relationship between respondents' beliefs that emptying the Recycle Bin is sufficient to erase the files within permanently and their beliefs that the Recycle Bin is a type of file shredder software, which may be related to the correlations seen in Table 9.

As shown in Table 11, there was a statistically significant association ($p = 0.006$) between the belief that deleting a file may still leave copies elsewhere on the drive and the belief that a file-shredding utility should be run on the drive prior to disposal. Likewise, as shown in Table 12, there was a statistically significant association ($p < 0.001$) between the belief that deleting a file may still leave copies elsewhere on the drive and the belief that a file shredder should be run on the entire drive before disposing of it. These results suggest a strong statistical relationship between a belief that deletion may be insufficient to remove all copies of a file and a belief that additional erasure methods are needed prior to disposal. However, as shown in Table 9, there was no statistically significant relationship ($p = 0.410$) between the belief that deleting a file may still leave behind copies elsewhere on the drive and the belief that a drive should be encrypted prior to disposal, suggesting that respondents in this situation may not view encryption as a method for ensuring files cannot be recovered. However, as Tables 13 and 14 show, there is

a statistically significant relationship between the belief that encryption makes files more difficult to recover and the belief that encryption should be used.

In the case of whether encryption should be used on the entire drive before disposal, there was a significance of $p = 0.023$, significant at the 0.05 level, while in the case of whether specific confidential files should be encrypted before deletion, the significance level was $p = 0.008$, significant at the < 0.01 level. Likewise, as shown in Table 17, there was a statistically significant ($p = 0.015$) relationship between users who believed it was impossible to prevent files from being recovered. Users reporting that they believed files should be encrypted prior to deletion suggest that they may view encryption as a security measure to counter potential file recovery. However, Table 16 shows that there is no statistically significant ($p = 0.109$) relationship between users who believed it was impossible to prevent files from being recovered. Users report a belief that the entire drive should be encrypted prior to disposal. Additionally, as shown in Table 15 and Table 18, there was no statistically significant relationship between users reporting that they believed it was impossible to prevent files from being recovered and user opinions on whether drives should be reformatted or run through file shredders prior to disposal, suggesting that users who believe it is impossible to prevent files from being recovered are unlikely to alter their behavior when it involves using secure deletion tools.

Evaluation of the Findings

Research Question 1: “Do a statistical majority of healthcare workers understand how to erase HIPAA-protected files?”

Based on the results presented above, the majority of respondents reported that they believed removing files from the Recycle Bin was sufficient to erase them permanently, and that they were unsure whether there was any difference between emptying the Recycle Bin and using

a file-shredding application. This indicates that most respondents were unsure how to ensure that a HIPAA-protected file could be erased in a way that prevents malicious actors from recovering it, supporting earlier research by Jones et al. and Shamlawi in their respective studies. However, the majority of respondents indicated awareness that there may be copies of the files they were not aware of and that they might need to locate, and reported somewhat confident agreement that reformatting a drive would securely erase the data on the drive. Additionally, examination of correlations between confidence in existing knowledge or training and confidence in beliefs about how data storage and tools function indicated a mild correlation, aligning with the UTAUT expectation that the social factor of preexisting beliefs and training influences the likelihood of adopting technology such as secure erase tools.

Research Question 2: “To what degree does a user’s understanding of how data storage works influence their behavior when deleting HIPAA-protected files?”

The results demonstrate a correlation between a user’s existing beliefs and their choice of interaction with data. Users who believe that reformatting drives will remove all data are more likely to make the additional effort to reformat drives before discarding them. Conversely, users who believe it is impossible to prevent files from being recovered will take the time to encrypt files before deleting them. Meanwhile, users who believe there is no way to prevent files from being recovered are less likely to alter their behavior by using file shredders to erase the data or reformat the drive before discarding it. Users who believe that removing a file from the Recycle Bin is sufficient to erase the data fully are unlikely to alter their behavior and begin using file-shredding applications. This suggests that, as proposed by the UTAUT model, social factors such as training, prior beliefs, and the perceived usefulness of shredder applications may influence

whether users take the time to use secure deletion methods, such as reformatting or file-shredder utilities.

Research Question 3: “Is there a statistically significant relationship between healthcare workers’ understanding of how files are deleted and how they should be deleted?”

The results indicate a statistically significant relationship between users' beliefs about how file deletion functions and their beliefs about how they should delete files. Users who believed encryption made files more difficult to recover were more likely to report that files should be encrypted prior to deletion and that drives should be encrypted before disposal, as were users who believed it was impossible to prevent files from being recovered entirely. Users who believed files may have copies elsewhere on the drive that they were unaware of were more likely to report that drives should be reformatted or run through file-shredding software before disposal. Users who reported a belief that the Recycle Bin was a type of file shredder were also more likely to believe that the Recycle Bin was a sufficient method for deleting confidential files. These results suggest a statistically significant relationship between users' beliefs about how specific tools and technologies function and their likelihood of engaging with those technologies.

Summary

An examination of the results revealed that a correlation exists between social factors, such as training and preexisting beliefs, and users' likelihood of behaving appropriately when handling confidential files or disposing of old drives. Users who believe they have been trained in specific techniques or are already familiar with those tools and techniques are more likely to be confident in their understanding, even if their actual knowledge may be limited. There appears to be a mild correlation between a user’s existing understanding of the technology and how they

behave when deleting files, and there is likewise a statistically significant relationship between how users believe data storage functions and how likely they are to engage with secure deletion tools.

The next section discusses the implications of these findings, recommends future research, and draws conclusions from these results.

Chapter 5: Implications, Recommendations, and Conclusions

The problem addressed by this study was the challenge of user handling of data in healthcare storage systems (Carlton, 2005; Shamlawi, 2018). Users of digital devices often fail to properly erase data when donating or discarding the device, resulting in the second-hand market containing a large number of supposedly empty devices that may contain personal or confidential data that can be quickly recovered. This is because users generally assume that deleting data on a drive permanently erases it, unaware that it is merely hidden rather than removed. Existing studies indicate that many second-hand devices still contain personally identifiable or confidential data that can be recovered using standard forensic tools (Jones et al., 2016; Osawaru, 2024; Sutherland et al., 2010).

In many of the cases studied, even though the drives had been formatted, researchers were still able to recover sufficient data to identify the former owners personally. This indicates that the original owners of the devices believed they had taken the necessary steps to delete their data before selling or donating the devices and, as a result, were unaware that their personally identifiable data could be recovered. This suggests that users lack a proper understanding of how data storage works, or that existing data management tools fail to communicate how to truly delete their data, instead implying that deletion or formatting is sufficient. This misinterpretation may influence how users behave when attempting to delete confidential data. In the healthcare industry, when laws often dictate that certain information must be deleted thoroughly, leaving the data in this recoverable form exposes healthcare professionals and their patients to threats, from the loss of private data to fines or legal penalties (Heath et al., 2022; Jones et al., 2016; Shamlawi, 2018).

The purpose of this quantitative, non-experimental, survey-based study was to determine what healthcare workers prioritize in file management systems when deleting data for HIPAA

compliance and to address an existing research gap regarding the relationship between users' expectations for file systems and their perceptions of how the systems work. Users' needs for both privacy and convenience, as well as their desire to reverse accidental deletions, pose a significant dilemma for operating system designers, and no comprehensive study of what average users require has been conducted. This issue affects users across the board, including average home users and corporations, who face threats from improperly deleted data. This is the dilemma developers face when seeking to appease contradictory user goals (Hadi, 2016; Jones et al., 2016). Further research is needed into users' needs and broader societal views on data privacy and retention to better understand how users want operating systems to handle their data and how these factors influence user behavior when deleting HIPAA-protected files. Based on the results of such research, designers seeking to enhance their data storage code would have a firmer baseline to work from.

The study used a quantitative Likert-scale survey with questions designed to yield data relevant to the research questions. The responses to the study were examined using a Pearson's correlation coefficient to determine potential relationships between responses and the Cronbach's Alpha to verify the consistency of responses. However, while the sample size was determined using Yamane's formula targeting a 10% margin of error, a larger sample targeting a 5% or smaller margin of error could yield more accurate results. Additionally, because the study was non-experimental, the results relied on respondents' self-reports rather than direct observation of user behavior. Future larger-scale studies or experimental observations could address these limitations.

This chapter discusses how the study's results relate to the research questions and whether the null hypothesis for each question can be rejected. It then discusses recommendations for how

affected organizations can improve their practices based on the study's results. This chapter additionally discusses potential future research to provide further insights into the factors that lead to improper data disposal and to explore more effective methods to ensure that confidential data cannot be recovered after disposal.

Implications

The purpose of this quantitative, non-experimental, survey-based study was to determine what healthcare workers prioritize in file management systems when deleting data for HIPAA compliance and to address an existing research gap regarding the relationship between users' expectations for file systems and their perceptions of how the systems work. To determine how user education and expectation affect behavior, the study was constructed around three central research questions.

R1

Do a statistical majority of healthcare workers understand how to erase HIPAA-protected files?

Hypotheses

$H1_0$

A majority of healthcare workers do not understand what happens when data is deleted.

$H1_a$

A majority of healthcare workers understand what happens to the data after deletion.

Based on the survey results, the Null Hypothesis cannot be rejected.

The majority of respondents (50%) believed that deleting a file from the Recycle Bin permanently erased it. In comparison, 31.82% believed that the Recycle Bin was equivalent to file-shredding software, and 32.73% were unsure whether there was a difference. Additionally, respondents appeared unsure when answering questions about encryption and reformatting: 23.64% were unsure whether reformatting was sufficient to erase files, 27.27% were only somewhat confident, and 29.9% were unsure whether disk encryption would make it more difficult to recover files. As such, these results appear to support the null hypothesis that most healthcare workers lack an understanding of how data deletion works or what happens to data after it is deleted.

R2

To what degree does a user's understanding of how data storage works influence their behavior when deleting HIPAA-protected files?

Hypotheses

$H2_0$

A user's understanding of how data storage functions does not significantly affect their behavior when deleting HIPAA-protected files.

$H2_a$

A user's understanding of how data storage functions significantly affects their behavior when deleting HIPAA-protected files.

Responses to questions such as "Reformatting a drive permanently removes data on that drive," "Once a file has been deleted from the Recycle Bin, it is gone forever," and "It is impossible to prevent deleted files from being recovered," suggested that a statistically

significant relationship exists between users' understanding of how data storage functions and their behavior. Respondents who believed that reformatting drives will remove all data were more likely to make the additional effort to reformat drives before discarding them, and users who believe it is impossible to prevent files from being recovered will take the time to encrypt files before deleting them. Meanwhile, respondents who reported believing that there is no way to prevent files from being recovered were less likely to alter their behavior to use file shredders to erase the data or reformat the drive before discarding it. Users who believed that removing a file from the Recycle Bin was sufficient to fully erase the data were unlikely to change their behavior and start using file-shredding applications. This suggests that, as proposed by the UTAUT model, social factors such as training, prior beliefs, and the perceived usefulness of shredder applications may influence whether users take the time to use secure deletion methods, such as reformatting or file-shredder utilities. As such, the null hypothesis can likely be rejected.

R3

Is there a statistically significant relationship between healthcare workers' understanding of how files are deleted and how they should be deleted?

Hypotheses

$H3_0$

There is no statistically significant relationship between healthcare workers' understanding of how files are deleted and how they should be deleted.

H3a

There is a statistically significant relationship between healthcare workers' understanding of how files are deleted and how they should be deleted.

The results suggest a statistically significant relationship between users' perceptions of how file deletion works and their approach to deleting files. Respondents who believed encryption made files more challenging to recover were more likely to report that they believed files should be encrypted before deletion and drives should be encrypted prior to disposal, as were respondents who believed it was impossible to prevent files from being recovered entirely. Respondents who believed that files may have copies elsewhere on the drive that they may be unaware of were more likely to report that they believed drives should be reformatted or run through file shredder software before disposal. Respondents who reported a belief that the Recycle Bin was a type of file shredder were also more likely to believe that the Recycle Bin was a sufficient method for deleting confidential files. These results suggest a statistically significant relationship between users' perceptions of specific tools and technologies and their likelihood of engaging with those technologies. As such, the null hypothesis can likely be rejected.

Recommendations for Practice

Based on the findings, the following recommendations are made for organizations working with HIPAA-protected data:

Improve training on how to delete files

By providing better training on securely erasing files, organizations can address one of the root causes of users' failure to erase data. The survey results suggest that the majority of respondents mistakenly believe that file shredders are simply another term for the Recycle Bin (Fig. 10) and that emptying the Recycle Bin fully purges the files (Fig. 6). While in-depth technical education into how file storage functions may be excessive, teaching employees that emptying the Recycle Bin does not actually delete any files and that a specific Secure Erase method such as a file shredder (with examples of what types of software may qualify) is needed

actually to erase a file, may help to ensure that users properly erase HIPAA protected files rather than merely deleting the metadata and leaving the file recoverable (Weijers, 2022).

Implement automated secure-erase and anti-theft technologies

Rather than relying on users to properly dispose of data in every case, automated tools could handle most actions regardless of user behavior. Secure erase technologies do not require manual intervention to run; they can be scheduled to execute at set intervals, often with additional configuration options specifying which files or data types should be automatically shredded. As such, installing and configuring Secure Erase software to run automatically, at least on non-SSD drives where the repeated operations will not harm the longevity of the drive, would allow organizations to ensure that confidential data would be regularly purged without needing to rely on employees to follow specific deletion procedures each time (Weijers, 2022).

Additionally, implementing standard anti-theft measures, such as full-disk encryption, can help ensure that even if a malicious actor obtains a drive containing confidential data, that actor would struggle to recover any usable information from the drive. Without access to the security hardware and the decryption key, any data on the encrypted drive would be unusable. While recovering or reverse-engineering the decryption key may be theoretically possible, it would be extremely difficult in practice. As a result, attackers could not obtain any information from the drive, even if the data were not properly deleted, as long as the drive remained fully encrypted. This method would also allow organizations to defend deleted data on SSD drives where proper secure erasure is not possible (Weijers, 2022).

Implement standardized practices for the disposal of old drives

The results suggest that users are likely to improperly erase data before disposing of old drives, potentially leaving confidential data recoverable. While better training can help to address this issue concerning individual users and files, a standardized approach, such as dictating a specific shredder utility that should be run with predefined settings on all drives before the disposal of a drive, can help to ensure that all drives are erased adequately before disposal without needing to rely on the knowledge of individual employees. Having a standardized, predefined process that employees can follow without additional knowledge on their part can reduce the likelihood of mishaps and better protect the organization (Rodrigues et al., 2024).

Recommendations for Future Research

This study identifies several issues in user education that may lead to unsafe handling of HIPAA-protected data. Future studies could therefore focus on developing more effective methods for communicating with non-expert users to address these issues. Examples of studies that could build on this research include the following.

Studies on user interface improvements and their potential contributions to user understanding could help develop better interfaces, reducing the disconnect between users' beliefs and reality, and ensuring that users can fully erase data when desired. Existing user-facing elements, such as the “Delete” and “Recycle Bin” buttons, may lead users to form incorrect assumptions about an element's function, resulting in mistaken beliefs that they are deleting data when they are not. Studies could focus on improving how interfaces convey accurate information to users without overwhelming them. For instance, this could involve reminding users that data deleted using certain methods may still be recoverable with standard data-recovery tools, or providing easier access to secure-erase tools.

Quantitative experimental studies can be conducted using various interface designs, monitoring how users interact with them to determine which interfaces are most and least effective at enabling users to perform their desired data management actions correctly and efficiently. A mixed-methods study could be conducted by interviewing healthcare workers about the issues they most frequently encounter when attempting to securely delete data and asking them to rank the severity of these issues on a Likert scale, to better inform future research projects on which aspects of the user experience are most in need of improvement. A qualitative study could be conducted by interviewing healthcare workers to determine what they consider essential for data deletion and which features they would like to see in user interfaces to better support their goals.

Studies of educational programs could also prove valuable. Users do not currently demonstrate an accurate understanding of secure data removal, even among respondents who report having received proper training. As such, future studies could focus on identifying the most effective forms of training in conveying the necessary information to users, ensuring they gain a proper understanding of how their software works and avoid developing an incorrect view of their tools.

Experimental studies could be conducted by exposing healthcare workers to educational materials on secure data deletion and then observing their behavior when asked to erase files securely, the time required, and whether they successfully secure the files. Mixed-methods studies could involve surveying healthcare workers about their past training experiences and asking them to rate the perceived usefulness of each training method on a Likert scale, while also examining trends across responses. Broader studies could also examine training in data protection and deletion, as well as the resulting user behavior, in other fields with similar data

protection regulations, to determine whether these fields face similar issues or whether improved methods already exist within them.

Conclusions

Despite the high value placed on personal data and the laws that regulate it, breaches often occur due to human error. By failing to properly erase data on old drives before disposing of them, users unknowingly expose themselves and others to the risk of data theft. To determine why users fail to delete data, even when legally required to do so, a survey was administered to 112 healthcare professionals in the United States. These users were asked several questions to assess their beliefs regarding data storage and deletion, including their understanding of the data deletion process and their behavior when deleting data.

An examination of the survey results suggested three primary pieces of information: that the majority of users do not appear to understand how data storage functions, that a user's understanding of how data storage functions significantly affects their behavior when deleting HIPAA-protected files, and that there appears to be a statistically significant relationship between what a user believes about how data storage functions and what actions that user believes should be taken to ensure the data is secure. These results suggest that users unknowingly put themselves at risk by improperly deleting confidential data because they misunderstand how to erase information securely, and that they are generally unaware that their understanding of data deletion does not align with best practices. Based on these discoveries, an overall conclusion can be reached that the majority of users are unlikely to properly dispose of confidential digital data due to a lack of proper understanding, which aligns with prior studies indicating that users often fail to delete confidential data and the overall UTAUT model's proposition that users will not

engage with technologies such as secure erase tools if they do not understand the value of the technologies and are not in an environment where adoption of the technologies is encouraged.

Building on these results, future studies could focus on investigations of user interface design or education. Studies examining education could identify shortcomings in current methods for educating users about data storage and disposal and develop improved educational methods that better ensure users' mental models align with reality. Likewise, studies of user interface design could investigate better ways for computers to communicate with users, ensuring that the operating system or storage tool does not mislead users into believing that data has been deleted when it remains on the drive, or that users have more options for disposing of their data. Such studies could help ensure that future users are better protected against data theft resulting from the recovery of confidential data from old storage media.

References

Al Sharif, S., Al Ali, M., Salem, N., Iqbal, F., El Barachi, M., & Alfandi, O. (2014, March). An approach for the validation of file recovery functions in digital forensics software tools. In *2014 6th International Conference on New Technologies, Mobility and Security* (NTMS) (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/6814005/

Ali, R. R., Mohamad, K. M., Jamel, S. A. P. I. E. E., & Khalid, S. K. A. (2018). A review of digital forensics methods for JPEG file carving. *J. Theor. Appl. Inf. Technol, 96(*17), 5841-5856. https://www.jatit.org/volumes/Vol96No17/17Vol96No17.pdf

Alghatrifi, I., & Khalid, H. (2019, December). A systematic review of UTAUT and UTAUT2 as a baseline framework of information system research in adopting new technology: a case study of IPV6 adoption. In *2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS)* (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/9073292/

AlHarbi, R., AlZahrani, A., & Bhat, W. A. (2022). Forensic analysis of anti-forensic file-wiping tools on Windows. *Journal of forensic sciences, 67*(2), 562-587. https://onlinelibrary.wiley.com/doi/abs/10.1111/1556-4029.14907

Almulihi, A. H., Alassery, F., Khan, A. I., Shukla, S., Gupta, B. K., & Kumar, R. (2022). Analyzing the Implications of Healthcare Data Breaches through Computational Technique. *Intelligent Automation & Soft Computing, 32*(3). https://cdn.techscience.cn/ueditor/files/iasc/TSP_IASC-32-3/TSP_IASC_23460/TSP_IASC_23460.pdf

Alqahtani, M., & Braun, R. (2021). Reviewing influence of UTAUT2 factors on cyber security compliance: a literature review. *Journal of Information Assurance & Cyber Security.* https://opus.lib.uts.edu.au/handle/10453/157514

Azeem, E. A. (2022). The Data Carving-The Art of Retrieving Deleted Data as Evidence. *International Journal for Electronic Crime Investigation, 6*(2), 8-8. http://ijeci.lgu.edu.pk/index.php/ijeci/article/view/101

Basile, J. L., Gaia, J., & Sanders, G. L. (2020). Who Has My Data? Factors Contributing to HIPAA (Non) Compliant Behaviors. *Journal of Strategic Innovation and Sustainability*, 15(2), 83-108. http://www.na-businesspress.com/JSIS/JSIS15-2/6_BasileFinal.pdf

Carlton, G. H. (2005). A critical evaluation of the treatment of deleted files in Microsoft Windows operation systems. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (pp. 310c-310c). IEEE. https://doi.org/10.1109/HICSS.2005.8

Certilman, S. A., & Wiechmann, E. W. (2020). ADR in the Age of Cybersecurity. *NEW JERSEY LAWYER*, 55. https://www.researchgate.net/profile/Steven_Certilman2/publication/330170077_ADR_IN_THE_AGE_OF_CYBERSECURITY/links/5f9048fe458515b7cf911035/ADR-IN-THE-AGE-OF-CYBERSECURITY.pdf

Coker, D. C. (2022). A Thematic Analysis of the Structure of Delimitations in the Dissertation. *International Journal of Doctoral Studies, 17*. https://research.ebsco.com/c/yi2or4/search/details/5gakiyphmb?db=ehh

Cresswell, J. (2013). Qualitative inquiry & research design: Choosing among five approaches. https://repositorio.ciem.ucr.ac.cr/bitstream/123456789/501/1/Qualitative%20inquiry%20%26%20research%20design.%20design%20_%20Choosing%20among%20five%20approaches.%20%281%29.pdf

Daggupati, A. (2020). Analysis of Best Practices for Data Leak Management and Prevention of Data Harvesting. https://era.library.ualberta.ca/items/8f7b5cda-c885-443d-a40c-be9ffcc27eaf

Dillon, S. (2006). Hide and seek: concealing and recovering hard disk data. *James Madison University Infosec Techreport, 35*, 17. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=af13f05749e129fb47ab8abbc9970d5cf15898c4

Ewan, P. M. (2023*). The Impact of Budgeting on the Risk of Cybersecurity Insider Threat Actions: From the Perspective of IT Engineers* (Doctoral dissertation, Northcentral University).

Flair, I. (2023). Personally identifiable information (PII). *Salem Press Encyclopedia*. https://research.ebsco.com/c/udgvh3/viewer/html/yj64gxaw7z

García-Pérez, M. A. (2012). Statistical conclusion validity: Some common threats and simple remedies. *Frontiers in psychology, 3*, 325. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00325/full

Garfinkel, S. L. (2007). Carving contiguous and fragmented files with fast object validation. *digital investigation, 4*, 2-12. https://www.sciencedirect.com/science/article/pii/S1742287607000369

Garner, J. C. (2003). Final HIPAA security regulations: a review. *Managed Care Quarterly, 11*(3), 15-27. https://europepmc.org/article/med/14983648

Gyening, P. (2022). *Cybersecurity Professionals' Behavioral Predispositions Affecting Cybersecurity Compliance Intentions: A Quantitative Study* (Doctoral dissertation, Capella University). Cybersecurity Professionals' Behavioral Predispositions Affecting Cybersecurity Compliance Intentions: A Quantitative Study - ProQuest

Hadi, A. (2016). Reviewing and evaluating existing file carving techniques for JPEG files. In *2016 Cybersecurity and Cyberforensics Conference (CCC)* (pp. 55-59). IEEE. https://doi.org/10.1109/CCC.2016.21

Heath, M., Porter, T. H., & Silvera, G. (2022). Hospital characteristics associated with HIPAA breaches. *International Journal of Healthcare Management, 15(2),* 171–180. https://doi.org/10.1080/20479700.2020.1870349

HIPAA Data Breach. (2024). April 2024 healthcare data breach report. *HIPAA Journal.com.* Retrieved June 2, 2024, from https://www.hipaajournal.com/april-2024-healthcare-data-breach-report/

Israel, G. D. (1992). Determining sample size. Retrieved June 23, 2024 https://www.researchgate.net/profile/Subhash-Basu-3/post/how_could_i_determine_sample_size_for_my_study/attachment/5ebaa4924f9a520001e613b6/AS:890361492811785@1589290130539/download/samplesize1.pdf

Izah, S. C., Sylva, L., & Hait, M. (2023). Cronbach's Alpha: A Cornerstone in Ensuring Reliability and Validity in Environmental Health Assessment. *ES Energy & Environment, 23*, 1057. https://www.espublisher.com/journals/articledetails/1057

Jadil, Y., Rana, N. P., & Dwivedi, Y. K. (2021). A meta-analysis of the UTAUT model in the mobile banking literature: The moderating role of sample size and culture. *Journal of Business Research*, 132, 354-372. https://www.sciencedirect.com/science/article/pii/S0148296321002903

Jan, J., Alshare, K. A., & Lane, P. L. (2024). Hofstede's cultural dimensions in technology acceptance models: a meta-analysis. *Universal Access in the Information Society, 23*(2), 717-741. https://link.springer.com/article/10.1007/s10209-022-00930-7

Jones, A. (2009). Lessons not learned on data disposal. *Digital Investigation, 6*(1-2), 3-7. https://www.sciencedirect.com/science/article/pii/S1742287609000498

Jones, A., Martin, T., & Alzaabi, M. (2016). The 2016 analysis of information remaining on computer hard disks offered for sale on the second-hand market in the UAE. *Journal of Digital Forensics, Security and Law.* https://doi.org/10.15394/jdfsl.2016.1428

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British journal of applied science & technology, 7*(4), 396. https://eclass.aspete.gr/modules/document/file.php/EPPAIK269/5a7cc366dd963113c6923ac4a73c3286ab22.pdf

Kalayou, M. H., Endehabtu, B. F., & Tilahun, B. (2020). The applicability of the modified technology acceptance model (TAM) on the sustainable adoption of eHealth systems in resource-limited settings. *Journal of multidisciplinary healthcare*, 1827-1837. https://www.tandfonline.com/doi/abs/10.2147/JMDH.S284973

Khanijahani, A., Iezadi, S., Agoglia, S., Barber, S., Cox, C., & Olivo, N. (2022). Factors associated with information breach in healthcare facilities: a systematic literature review. *Journal of Medical Systems,* 46(12), 90. https://link.springer.com/article/10.1007/s10916-022-01877-1

Kolil, V. K., & Achuthan, K. (2023). Longitudinal study of teacher acceptance of mobile virtual labs. *Education and Information Technologies*, *28*(7), 7763-7796. https://link.springer.com/article/10.1007/s10639-022-11499-2

Kundu, A. (2022). TAM3+, a new approach to attract teachers towards technology. *Journal of Social Sciences, 18*(1), 57-68. https://www.researchgate.net/profile/Arnab-Kundu-3/publication/359518581_TAM3_a_New_Approach_to_Attract_Teachers_Towards_Technol

ogy/links/62427e2757084c718b72baef/TAM3-a-New-Approach-to-Attract-Teachers-Towards-Technology.pdf

Lai, P. C. (2017). The literature review of technology adoption models and theories for the novelty technology. *JISTEM-Journal of Information Systems and Technology Management, 14*(1), 21-38. https://www.scielo.br/j/jistm/a/D3NXPz5WF4gQX9cSdLKQv6D

Lee, B. T. (2009). Computer and Data Disposal in Plastic Surgery: Guidelines for Health Insurance Portability and Accountability Act Compliance. *Plastic and Reconstructive Surgery, 124*(1), 186e-187e. https://journals.lww.com/plasreconsurg/fulltext/2009/07000/computer_and_data_disposal_in_plastic_surgery_.99.aspx

Loes, M. C. (2024). Examining Outcomes of Privacy Risk and Brand Trust on the Adoption of Consumer Smart Devices. https://jagworks.southalabama.edu/theses_diss/194/

Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society, 14*, 81-95. https://link.springer.com/article/10.1007/s10209-014-0348-1

McCusker, K., & Gunaydin, S. (2015). Research using qualitative, quantitative, or mixed methods and choice based on the research. Perfusion, 30(7), 537-542. https://journals.sagepub.com/doi/abs/10.1177/0267659114559116

McLaughlin, K. (2023). *A Quantitative Study of Learner Choice in Cybersecurity Training: Do They Even Want Gamification?* (Doctoral dissertation, Colorado Technical University). https://www.proquest.com/openview/ea3614ee9c8b14c4e56839985d148e1f/1

Nahar, N. F. A. M., Ab Rahman, N. H., & Mohammad, K. M. (2018). E-raser: File shredder application with content replacement by using random words function. *JOIV: International*

*Journal on Informatics Visualization, 2(*4-2), 313-317. https://www.joiv.org/index.php/joiv/article/view/175

Osawaru, G. (2024). *Electronic Health Record Data Breaches in US Healthcare Industry: A Quantitative Study Using the Protection Motivation Theory (PMT) to Mitigate Data Breaches* (Doctoral dissertation, University of the Cumberlands). https://www.proquest.com/openview/9a02a80bb550063e488f77715a6b1957/1

Pal, A., & Memon, N. (2009). The evolution of file carving. *IEEE Signal Processing Magazine, 26*(2), 59-71. https://ieeexplore.ieee.org/abstract/document/4806206/

Park, E. S., & Park, M. S. (2020). Factors of the technology acceptance model for construction IT. Applied Sciences, 10(22), 8299. https://www.mdpi.com/2076-3417/10/22/8299

Poisel, R., & Tjoa, S. (2013). A comprehensive literature review of file carving. In *2013 International conference on availability, reliability and security* (pp. 475-484). IEEE. https://doi.org/10.1109/ARES.2013.62

Robles-Gomez, A., Tobarra, L., Pastor-Vargas, R., Hernández, R., & Haut, J. M. (2021). Analyzing the users' acceptance of an IoT cloud platform using the UTAUT/TAM model. *IEEE Access*, *9*, 150004-150020. https://ieeexplore.ieee.org/abstract/document/9600837/

Randolph, S. J. (2024). *Identifying the Role of Healthcare Leaders in Protecting Sensitive Information Within Their Sector* (Order No. 31482219). Available from Dissertations & Theses @ National University; ProQuest One Academic. (3087040222). https://www.proquest.com/dissertations-theses/identifying-role-healthcare-leaders-protecting/docview/3087040222/se-2

Rodrigues, G. A. P., Serrano, A. L. M., Vergara, G. F., Albuquerque, R. D. O., & Nze, G. D. A. (2024). Impact, Compliance, and Countermeasures in Relation to Data Breaches in Publicly

Traded US Companies. *Future Internet, 16(*6), 201. https://www.mdpi.com/1999-5903/16/6/201

Rowe, N. C. (2020). Current Privacy Concerns with Digital Forensics. http://faculty.nps.edu/ncrowe/forensicpriv.pdf

Sadri, M. (2024). HIPAA: A Demand to Modernize Health Legislation. *The Undergraduate Law Review at UC San Diego, 2*(1). https://escholarship.org/uc/item/9gp2n52k

Saleh, T. (2024). *Factors Affecting Cybersecurity Awareness: A Qualitative Study in Saudi Arabia* (Doctoral dissertation, Westcliff University). https://www.proquest.com/openview/53296ce58597ad76f0192f9007bd266a/1

Sangurima, O. (2021). *Medical Practitioners' Intention to Use Secure Electronic Medical Records in Healthcare Organizations* (Doctoral dissertation, Walden University). https://www.proquest.com/openview/d7029fa14011e008c03c054db6b91ea2/1

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. https://www.mdpi.com/2227-9032/8/2/133

Shamlawi, A. (2018). *Remnant Data Analysis on Disks Offered for Donation: A Jordan Case Study 2018* (Doctoral dissertation, Princess Sumaya University for Technology). https://www.proquest.com/openview/59e96d9d7ea1f6cf1922fa4274cd9230/

Sharma, S., & India, U. The Effectiveness of Different Cyber Security Measures in Protecting Organization for Cyber Threats. https://jcdronline.org/admin/Uploads/Files/64909a20694482.82653230.pdf

Sloan, P., & Juhnke, D. H. (2016). Secure disposal of medical practice records. *Missouri medicine, 113*(4), 264. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6139920/

Slocum, T. A., Joslyn, P. R., Nichols, B., & Pinkelman, S. E. (2022). Revisiting an analysis of threats to internal validity in multiple baseline designs. *Perspectives on Behavior Science, 45*(3), 681-694. https://link.springer.com/article/10.1007/s40614-022-00351-0

Sutherland, I., Davies, G., Jones, A., & Blyth, A. J(2010). Zombie hard disks-data from the living dead. https://ro.ecu.edu.au/adf/86/

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in science education, 48,* 1273-1296. https://link.springer.com/article/10.1007/S11165-016-9602-2

Tamilmani, K., Rana, N. P., Wamba, S. F., & Dwivedi, R. (2021). The extended Unified Theory of Acceptance and Use of Technology (UTAUT2): A systematic literature review and theory evaluation. *International Journal of Information Management, 57*, 102269. https://www.sciencedirect.com/science/article/pii/S0268401220314687

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education, 2,* 53. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4205511/

U.S. Bureau of Labor Statistics. (2023). Healthcare occupations: Characteristics of the employed. U.S. Bureau of Labor Statistics. Retrieved May 19, 2024, from https://www.bls.gov/spotlight/2023/healthcare-occupations-in-2022/home.htm

U.S. Department of Health and Human Services. (2013) Frequently Asked Questions About the Disposal of Protected Health Information. Retrieved August 4, 2024, from https://www.hhs.gov/sites/default/files/disposalfaqs.pdf

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences, 39*(2), 273-315. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.2008.00192.x

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425-478. https://doi.org/10.1109/ACCESS.2021.3125497

Weijers, F. (2022). Presentation and evaluation of common methods of deleting user data in common computer file systems. https://www.researchgate.net/publication/372840010_Presentation_and_evaluation_of_common_methods_of_deleting_user_data_in_common_computer_file_systems

Appendix A: Survey Questions

Please rank how much you agree with a given statement on a scale of 1-5, with 1 being "fully disagree" and 5 being "fully agree."

Page 1

1. I understand how digital data storage functions on a technical level.
2. I understand HIPAA requirements for ensuring sensitive information is deleted.
3. I understand how to ensure a deleted file cannot be recovered.
4. I am familiar with file shredder software.
5. My job training covered how and when to securely erase digital data.

Page 2

Please rank how accurate you consider a statement to be on a scale of 1-5, with 1 being "fully inaccurate" and 5 being "fully accurate."

6. Once a file has been deleted from the Recycle Bin, it is gone forever.
7. Reformatting a drive permanently removes data on that drive.
8. File shredder software is the same as deleting a file from the Recycle Bin.
9. Encrypting a file before deleting it makes it more difficult to recover.
10. It is impossible to prevent deleted files from being recovered.
11. Even if a file is erased, copies of it may exist elsewhere on the drive without my knowledge.

Page 3

Please rank how accurate you consider a statement to be on a scale of 1-5, with 1 being "fully inaccurate" and 5 being "fully accurate."

12. I should use shredder program to erase files containing HIPAA-protected data.
13. Prior to deleting HIPAA-protected data, the files containing that data should be encrypted.
14. Before discarding an old drive, the drive should be reformatted.
15. Before discarding an old drive, the drive should be encrypted.
16. Before discarding an old drive, a file shredder utility should be run on the drive.

Appendix B: Survey Statistics

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.811	.814	16

Item Statistics

	Mean	Std. Deviation	N
I understand how digital data storage functions on a technical level.	3.51	1.040	112
I understand HIPAA requirements for ensuring sensitive information is deleted.	3.75	1.174	112
I understand how to ensure a deleted file cannot be recovered.	3.08	1.483	112
I am familiar with file shredder software.	3.40	1.277	112
My job training covered how to securely erase digital data.	3.22	1.320	112
Once a file has been deleted from the Recycle Bin, it is gone forever.	3.17	1.401	112
Even if a file is erased, copies of it may exist elsewhere on the drive without my knowledge.	3.58	1.213	112
It is impossible to prevent deleted files from being recovered.	2.75	1.430	112
Encrypting a file before deleting it makes it more difficult to recover.	3.35	1.160	112
File shredder software is the same as deleting a file from the Recycle Bin.	2.90	1.273	112
Reformatting a drive permanently removes data on that drive.	3.21	1.295	112

I should use shredder program to erase files containing HIPAA-protected data.	3.50	1.057	112
Before discarding an old drive, a file shredder utility should be run on the drive.	3.32	1.330	112
Before discarding an old drive, the drive should be encrypted.	3.47	1.252	112
Before discarding an old drive, the drive should be reformatted.	3.53	1.252	112
Prior to deleting HIPAA-protected data, the files containing that data should be encrypted.	3.55	1.192	112

Appendix C: IRB Approval

National University

9388 Lightwave Ave.
San Diego, CA 92123
irb@nu.edu

Notice of Exemption

July 7, 2025

To: Jesse Schulman

Project Title: A Survey of the Relationship Between User Understanding of Data Storage and Behavior when Deleting HIPAA-Protected Information
NU IRB Number: IRB-FY24-25-862
Determination: Exempt from further review 45 CFR 46.101 Category 2.(i). Research that only includes interactions involving educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) if at least one of the following criteria is met:
The information obtained is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects;

Status: Active - Research activities may begin as of July 7, 2025

Dear Jesse Schulman:

The study referenced above has been reviewed by the National University IRB. The IRB has determined your research is exempt from further review under 45 CFR 46.104, which means you will not need to renew your study and may begin your study effective immediately. However, if you find the need to change your study in any way, you will need to submit a modification to the IRB prior to implementing the changes. This will allow the IRB to determine whether or not the study still meets exemption criteria.

Please review your Post Approval Responsibilities here: Approved Documents Guidelines

For any questions regarding your protocol, please reach out to the IRB at irb@nu.edu.

Sincerely,

Dr. Joseph Marron, IRB Chair

Dr. Brianne Mongeon, Director, HRPP & IRB

Jenessa Eberhardt, Associate Director, HRPP & IRB

Appendix D: Notice of Study Closure

National University

9388 Lightwave Ave.
San Diego, CA 92123
irb@nu.edu

Notice of Protocol Closure

October 9, 2025

To: Jesse Schulman

Project Title: A Survey of the Relationship Between User Understanding of Data Storage and Behavior when Deleting HIPAA-Protected Information
NU IRB Number: IRB-FY24-25-862

Status: Closed as of October 9, 2025

Dear Jesse Schulman:

Thank you for your submission of materials for this research study. The National University Institutional Review Board has CLOSED your project. **You must adhere to the following conditions:**

1. Once a study has been officially closed via a Request to Close Study, it cannot be re-opened.
2. If a later use for the research data is identified, you must submit a new research proposal for the use of the previously collected data.
3. The later use of the data may qualify for an exemption, if the existing data is recorded without identifiers; however, you must submit a new research proposal prior to using the data.
4. You will maintain the confidentiality of all data collected and will adhere to the federal policy of storing all data and consent documents in a secured environment for a minimum of 3 years.

If you have any questions, you may contact the IRB at irb@nu.edu. Please include your study title and reference number in all correspondence with this office.

Sincerely,

Dr. Joseph Marron, IRB Chair

Dr. Brianne Mongeon, Director, HRPP & IRB

Jenessa Eberhardt, Associate Director, HRPP & IRB