2022

An Empirical Investigation of the Evidence Recovery Process in Digital Forensics

Kevin Parviz
Nova Southeastern University, kevinparviz@msn.com

An Empirical Investigation of the Evidence Recovery Process in Digital Forensics

by
Kevin Parviz

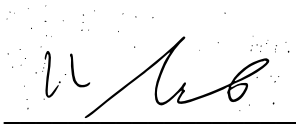
A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Cybersecurity Management

College of Computing and Engineering
Nova Southeastern University

2022

We hereby certify that this dissertation, submitted by Kevin Parviz conforms To acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

Sumitra Mukherjee, Ph.D. — **7/12/22**
Chairperson of Dissertation Committee — **Date**



Michael J. Laszlo, Ph.D. — **7/12/22**
Dissertation Committee Member — **Date**

Francisco J. Mitropoulos, Ph.D. — **7/12/22**
Dissertation Committee Member — **Date**

Approved:

Meline Kevorkian, Ed.D. — **7/12/22**
Dean, College of Computing and Engineering — **Date**

College of Computing and Engineering
Nova Southeastern University

2022

An Abstract of a Dissertation Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

An Empirical Investigation of the Evidence Recovery Process in Digital Forensics

by

Kevin Parviz

June 2022

The widespread use of the digital media in committing crimes, and the steady increase of their storage capacity has created backlogs at digital forensic labs. The problem is exacerbated especially in high profile crimes. In many such cases the judicial proceedings mandate full analysis of the digital media, when doing so is rarely accomplished or practical. Prior studies have proposed different phases for forensic analysis, to lessen the backlog issues. However, these phases are not distinctly differentiated, and some proposed solutions may not be practical. This study utilized several past police forensic analyses. Each case was chosen for having five distinct forensic phases, complete with documented amount of time spent in each phase, along with the number and type of recovered evidence. Data from these cases were empirically analyzed using common descriptive statistical analyses along with linear regression. By using linear regression, we tested the factors that determine the number of recovered evidentiary artifacts.

This study provides models by which future forensic analyses could be assessed. It presents distinctive boundaries for each forensics phase, thus eliminating ambiguity in the examination results, while assisting forensic examiners in determining the necessary depth of analysis.

Keywords: Digital forensics, backlogs, category of crimes, digital media size, digital forensic phases, triage, preview, key evidentiary artifact, linear regression.

Acknowledgments

I would like to thank my wife and daughter for their patience and support. I would like to also thank my mom for her inspirations and my dad for his encouragements. May they rest in peace.

My special thanks to my professor, Dr. Steven Horbal. He patiently taught me all I know about statistics. I would also like to thank committee chair Dr. Sumitra Mukherjee and my committee members, Dr. Laszlo and Dr. Mitropoulos for their assistance and support.

Table of Contents

List of Tables

Tables

List of Figures

Figures

Chapter 1

Introduction

Background

Digital forensics is the process of identifying potential evidence, preserving said evidence, analyzing the evidence, and presenting the evidence proficiently in a judicial proceeding (McKemmish, 1999). McKemmish (1999) and the National Institute of Standards and Technology (NIST) special publication (SP) 800-86 by Kent, et al. (2006) provided two of the most widely accepted definition and frameworks for digital forensics (Martini & Choo, 2012). However, the steady increase in the size of digital media has created a backlog for today's digital forensic examiners (Quick & Choo, 2014; Yang et al., 2016). In addition to the increase in the size of digital media (Quick & Choo, 2014; Yang et al., 2016), the Federal Bureau of Investigation (FBI) reported that computer-related crimes are also steadily increasing, from approximately 288,000 complaints filed in 2015, to over 847,000 in 2021 (Internet Crime Complaint Center, 2021). Notwithstanding such increase, many courts use the term *full analysis* of digital media in their rulings (*State v. Newman*, 2013; *US v. Stabile*, 2011; *US v. Tello*, 2017) despite the infeasibility of the analysis of every byte on a digital medium in most cases (Casey et al., 2009). Full analyses inherently entail longer judicial proceedings, which can pose problems for forensic examiners who want to abide by the courts' requests but have to adhere to the United States (US) Sixth Amendment rights of its citizens to have an expedient trial (Casey et al., 2009). If by full forensic analysis the courts mean accurate but expeditious *depth of digital forensics processing methods and analysis,* then the term has to be accurately

described. However, the term does not appear to be properly defined in the case laws or the literatures reviewed for this study.

With the exception of live digital media acquisitions such as cellular phones and Random-Access Memory (RAM), the process of digital forensics is traditionally done in the following manner: A digital medium is connected via a write-blocker, and a *forensic image* (a bit-by-bit image) of the medium is created on the examiner's computer or server (Bem & Huebner, 2007; Dancer & Skelton, 2013: Nisbet & Jacob, 2019). The forensic image is then processed with the examiner's choice of automated tools into data that are capable of being analyzed (Quick & Choo, 2018a). The imaging stage of forensics was created in order to ensure the integrity of the evidence in judicial proceedings (Dancer & Skelton, 2013). In order to provide a balance between the amount of time spent on forensic analysis of a digital medium, and the depth of such analysis, many forensic software tools such as Forensic ToolKit (FTK) provide the option for a phased analysis (Carbone, 2014), though these phases may not have a particular naming convention. As an example, FTK Imager provides an option for a brief overview and imaging of the digital media, while FTK provides everything from limited data examination (in Field Mode), to providing the choice of including additional data processing for more in-depth phases (Carbone, 2014). In FTK, once the automated acquisition and parsing of data is completed, the data are placed within their corresponding containers (e.g., pictures, videos, etc. are placed in the *Multimedia* (MM) category) (Carbone, 2014). It is this automated categorization of data that assists an examiner in expediting the forensics process (Du & Scanlon, 2019). In an example of a crime that occurred at a specific timespan, Du and Scanlon (2019) suggested that artifacts with similar file types, similar directories, and related date and time are likely to have evidentiary values. Similarly, Scrivens and Lin (2017)

introduced a three-phase examination in mobile digital forensics, one of which was the distinct task of finding the primary (key) evidence artifacts. Scrivens and Lin (2017) explained that in the example of a white-collar crime, there is a high probability that the evidence artifacts would be in a non-Multimedia (nMM) file such as documents, rather than in Multimedia (MM).

The overwhelming demand for digital forensic analyses has prompted some researchers to limit forensic examinations to specific phases. Shaw and Browne (2013) proposed conducting digital forensic examination in phases such as *field triage*, *triage*, *previews*, or *enhanced previews*, and to avoid performing an impractical *full analysis* in most cases. The depth of examination and the time spent for examining the data increase progressively in Shaw and Browne (2013)'s proposed phases. Casey et al. (2009) and Casey (2011) proposed the terms *triage*, *preliminary*, and *in-depth* as phases of forensics. Cantrell and Dampier (2013) further divided *triage* into sub-phases such as the *computer profile*, the *crime profile*, and the *presentation* process. Cantrell & Dampier (2013) created a computer profile process to analyze data according to factors such as "crime class" (hereafter referred to as the *category of crime*), and specifically indicated that the "crime potential" is one of the determining factors in the prioritization and the analysis of evidence (p. 86).

Problem Statement

This dissertation aims to empirically investigate the problem of delays and backlogs in processing digital forensic cases due to the upsurge in computer-related crimes and the increase in size of the digital media used to commit such crimes (Hitchcock et al., 2016; Nouh et al., 2019; Quick & Choo, 2014; Yang et al., 2016). The culmination of such upsurge has resulted in reduced sentences for many criminals, and lengthy waiting time for innocent people (Shaw & Browne, 2013).

At one extreme, full forensic analysis of every byte on a digital medium is infeasible, has hardly ever been conducted, and can cause harmful delays (Shaw & Browne, 2013; Casey et al., 2009). At the other extreme is to perform a *triage* with the concerns that potential evidence may be overlooked, which could result in legal challenges (Shaw & Browne, 2013). However, Shaw and Browne (2013) questioned the latter argument by stating that the backlogs created due to full forensic analyses have resulted in reduced sentencing for criminals, and lengthy waiting period for innocent people. Shaw and Browne (2013) put forth examples of people wrongly accused of child abuse, being separated from their children for absurdly long periods of time, while awaiting the results of full digital analysis of their devices. Years after Shaw and Browne (2013) warned the academic community about the negative consequences of digital forensics backlogs, Hamilton (2020) stated that 32 police departments in England and Wales have reported a total of 12,122 backlogged digital devices that included crimes such as sexual offences and terrorism. Thompson (2019) provided an alarming outlook on the current state of digital forensic labs in England. In one example, Thompson (2019) reported that a suspect who had pleaded guilty to the possession of over 4,000 Child Sexually Abusive Materials (CSAM), while distributing and sharing such images, could not be sentenced. Thompson (2019) explained that after two and a half years of waiting, the police's digital forensics lab had not examined the suspect's digital media due to backlogs; and that the presiding judge at Plymouth Crown Court called the long delay in digital forensic analysis *unacceptable*. In reflecting upon the backlog issues, Horsman (2017) declared that without major overhauls it may be difficult to see how digital forensics could be sustainable in providing support for the criminal justice system.

Research Goals

The main goal of this research was to empirically investigate the usefulness of retrieving different types of evidence during various phases of examination in digital forensics. For the purpose of this study, two evidence types — Multimedia (MM) and non-Multimedia (nMM) — were considered over five examination phases: triage, preview, imaging, examination, and legal examination. To support this goal, this research identified, dichotomized, discretized, and categorized sets of evidence types (MM and nMM) that were collected from over 100 past police cases. After processing each evidence type, the number of collected evidentiary artifacts were counted based on the past police forensic reports. These reports also revealed the Key Evidence (KE) that led the investigation to the recovery of other evidentiary artifacts, which Scrivens and Lin (2017) referred to as key recovered contents. In the next step, based on a combination of Shaw and Browne (2013) and FTK's phased analysis in Carbone (2014), this research focused on five phases of forensic examination. These five phases (triage, preview, imaging, examination, and legal examination) are the distinct stages in which forensic examinations of the past police cases were done. Police Activity Log Sheets (PALS) are documents created by police officers on daily basis and reflect the use of time during the normal tour of duty (Thornton & Harper, 1991). By using the documented past PALS, the time spent on each forensic examination phase were collected as $\mathbf{t_i}$ (in hours): triage (t_1), preview (t_2), imaging (t_{3a}), examination (t_{3b}), and legal examination (t_4). Using a combination of police forensic reports and court documents, the phase in which KE was discovered were determined. Using a combination of past PALS and police forensic reports, two categories of crime were obtained: Sex Offence (SO) and not Sex Offence (nSO). Additionally, the total size of the examined digital media was documented for each case.

Research Questions (RQ)

The goal of this dissertation was to empirically investigate the usefulness of retrieving different types of evidence during various phases of examination in digital forensics. The following questions guided this study:

RQ1

What is the relationship between hard drive capacity and evidence collected in each phase?

RQ2

What is the relationship between hard drive capacity and the number of hours spent in each phase?

RQ3

What is the relationship between hours spent in each phase and the total recovered evidence?

RQ4

Which investigative phase is most likely to produce Key Evidence for varying categories of crime?

RQ5

Which phase is most efficient in terms of evidence collected per hour?

Relevance and Significance

Over 10 years ago, Garfinkel (2010) suggested that the golden age of digital forensics was reaching an end. In reflecting upon Garfinkel (2010), Horsman (2017) stated, “As the pressures mount upon digital forensics [sic] to support criminal investigations into digital crimes, it is necessary to question whether it can continue to do so effectively.” (p. 452). Both Garfinkel (2010) and Horsman (2017), along with

many others (Hitchcock et al., 2016; Nouh et al., 2019; Quick & Choo, 2014; Yang et al., 2016) cited the growing size of storage devices as one of many factors in their prediction. The relevance and significance of this study was its attempt to provide a solution for such dire prediction.

Relevance

The relevance of this study was its practical application for judicial proceedings that involve digital forensics, where pertinent data must be accurately identified for an expeditious trial (Casey et al., 2009). In addition to evaluating the usefulness of discovered evidence in each phase of forensic examination, this research presented distinctive boundaries for each phase, thus eliminating ambiguity in the examination results at judicial proceedings.

The results of this study may have positive effects on the decision-making process for forensic examiners to determine the necessary depth of analysis by choosing the correct forensic examination phase, thus allowing appropriate time allocation for the analyses in more complex examinations. As a direct consequence, another effect of this study may be the foreseeable reduction in workloads at digital forensic labs.

Significance

The significance of this study was the application of multivariate statistical analyses in digital forensics, since studies in incorporating other sciences into the broad concept of digital forensics have been sparse and limited (Horsman et al, 2014; Taha & Yoo, 2018). As an example, linear regression were used to test the factors that determine the number of recovered evidentiary artifacts.

Issues

As previously indicated, this research focused on two independent variables (category of crime and the digital media size). However, in addition to the two, there are numerous other variables that can affect the outcomes of interest in this research. Among these are *challenges* (technical, legal, personnel, and operational challenges) brought forth by Karie and Venter (2015). By selecting the police cases that were not impacted by these challenges this research aimed to minimize their effects. However, Karie and Venter (2015)'s technical challenge – *digital media size* – was the only factor included in this research. Additionally, the police cases that were excluded from this research contained within them *complexities* (encryption and forensic tools' incompatibility with the evidence), and *volatility* (risk of data loss due to fragility, e.g. RAM acquisition) (Karie & Venter, 2015).

The other variable excluded from the sample of police cases was *urgency*. Roussev et al. (2013) described urgency as a factor in digital forensics where the results need to be produced in minutes, while (Shaw & Browne, 2013) described it as analyses that are done in a timely manner. Karie and Venter (2015) referred to urgency as the "limited window of opportunity to the [sic] collection of potential digital evidence" (p. 887). Losavio et al. (2015) justified direct examination of a digital medium without the use of any forensic software in these urgent and exigent circumstances. Such exigent forensic examinations have been performed in the past police cases but were excluded in this research. Finally, all police cases in the sample population were examined using the same forensic machine.

Assumptions and Limitations

In a few cases, during a police investigation, and prior to the start of forensic examination, the location of evidentiary artifact was revealed (e.g., due to the

suspect's confession, witness statement, etc.). In these cases, an assumption was made as to which phase of forensic examination would have recovered such evidence depending on the depth of forensic examination that was needed to recover it.

Cellular (cell) phone forensic tools have only recently started to provide the ability to perform limited triage on selected phones (Cellebrite, 2022; MSAB, 2022). In addition to having no triage database within the past police cases, these cases lack imaging information. As Dancer and Skelton (2013) indicated, the advent of cell phone forensics changed the practice of imaging, since a cell phone cannot be forensically imaged. This is due to many changes that occur when a phone is connected to a computer (Dancer & Skelton, 2013). These changes could be automatic and unintentional (e.g., written log files), or purposefully done by the cell phone forensic tools or the forensic examiner (e.g., manipulating the system's kernel or cell phone's defensive systems) (Dancer & Skelton, 2013). Such changes create a hash value that is different than the created image, therefore cannot be considered as *forensic images*. Due to the lack of cell phone database for the triage phase (t_1) and the imaging phase (t_{3a}), this study excluded cell phones as a part of its research.

Approach

Data Used for the Study

Over 100 criminal cases from police archives between 2012 and 2020 were used as the sample data for this quantitative study. These criminal cases were all resolved pursuant to digital forensic examinations; and each criminal case resulted in the suspect's conviction based on the digital forensic findings. FTK Imager and FTK were the two primary forensic tools that were used to analyze the digital media in these cases. Five distinct phases of examination were established, and the time spent on each phase were documented. The resulting evidentiary artifacts were also used as

the data for this study. Each forensic artifact on the suspect's digital media that played a role in their conviction was retrospectively evaluated along with the criminal activity that the suspect was convicted of -- Sex Offence or not Sex Offence. The total size of all the digital media for each case were also documented. The evaluation of the forensic artifacts entailed categorizing each artifact according to their type (Multimedia, non-Multimedia), the number of collected evidentiary artifacts, and the recovery of the primary key evidentiary artifact (Key Evidence).

Overview of Approach to Research Questions

To answer the research questions, descriptive statistics included Student's *t* Test and Fisher's exact test. The inferential statistics included linear regression (for answers to RQ1, 2, and 3), followed by a descriptive Chi-squared (RQ4), and a descriptive evidence collection rate (RQ5).

Definition of Terms

Digital Artifacts

In this research *digital artifacts* refers to the digital data that are not always obviously present but occur as a result of preparative and investigative procedures during a digital forensic examination.

Digital Media

A term used in this research for any electronic, magnetic, optical, or electrochemical device that is capable of storing digital data.

Evidence

This term is defined as, "Something legally submitted to a tribunal to ascertain the truth of a matter" (Merriam-Webster, 2020, Definition 1). In this research the term evidence also includes *relevant evidence*, as defined by Federal Rules of Evidence (2010), "Relevant evidence means evidence having any tendency to make the

existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence" (Rule 401(a)).

Evidentiary Artifacts

Evidence in the form of a *digital artifact*.

Forensic Image

A bit-by-bit image of a digital medium that is created on the examiner's computer or server (Bem & Huebner, 2007; Dancer & Skelton, 2013)

Key Evidence

As it is used in this research, Key Evidence means the primary evidence that has led the forensic examiner to the discovery of other evidentiary artifacts.

Multimedia Artifacts

Denotes the integration of multiple types of media in the form of graphics, videos, and audio clips.

Non-Multimedia Artifacts

All digital artifacts that are not multiple types of media.

Preview

Shaw and Browne (2013) describe *preview* as an examination after triage and prior to a full forensic analysis. As it is used in this report, preview entails limited parsing and carving of data, while using an automated process to organize the data into particular categories – as seen in FTK's Field Mode (Carbone, 2014).

Triage

Triage is the cursory evaluation of the digital media proposed by Rogers et al. (2006). As it is used in this research, triage is referred to the examination of allocated files and directories without an attempt to parse, or carve any data, and without the use of an automated process to organize the data into any particular categories. An

example of this type of triage is the use of FTK Imager for examination (Carbone, 2014).

Chapter 2

Literature Review

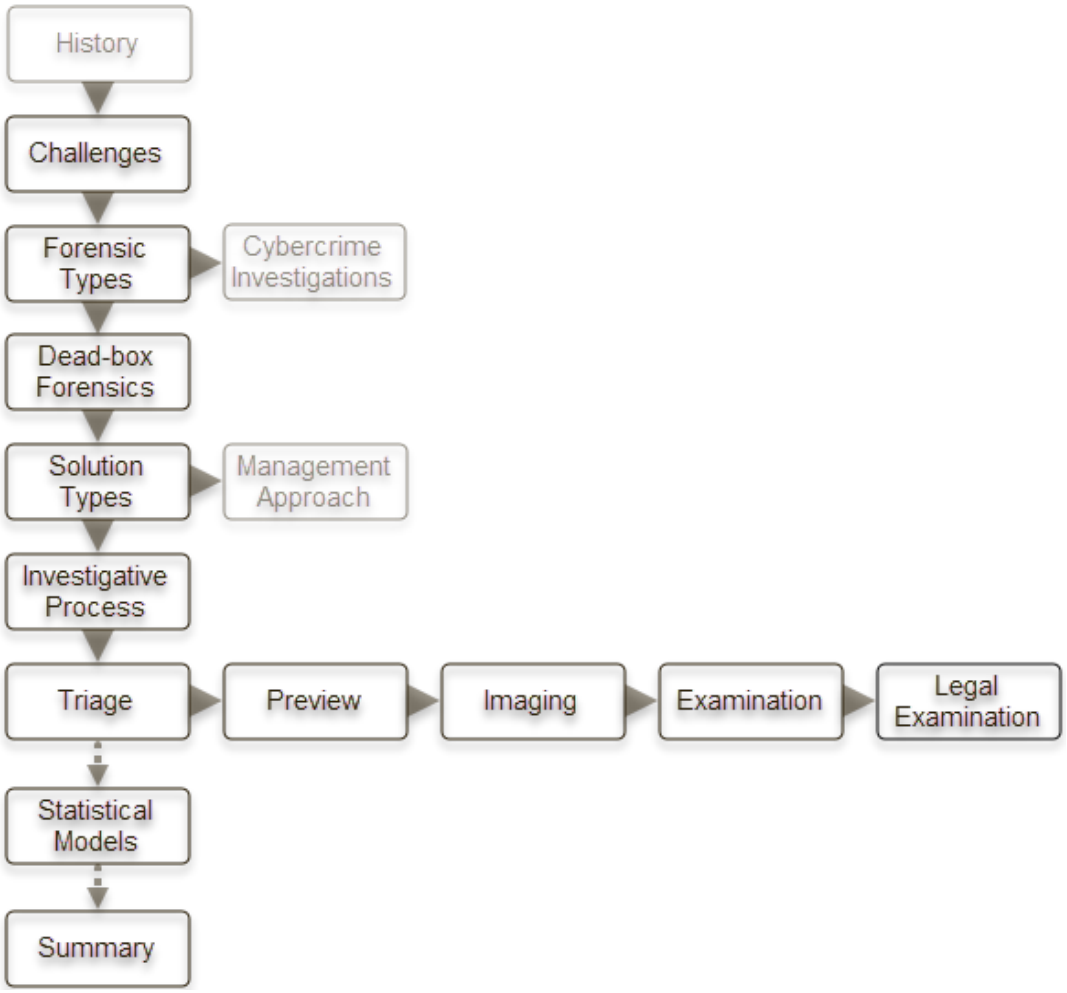
Introduction

The main goal of this research is to investigate the usefulness of various phases in digital forensics. By eliminating or reducing the time spent on impractical phases, forensic examiners may be able to focus on the accuracy of their analyses while reducing their backlogged cases. For the purposes of this research, five distinct forensic phases were evaluated: Triage, Preview, Imaging, Examination, and Legal examination. These phases were used to examine the digital media in the police cases presented for this research, and they are common phases used in many forensic labs. Each phase's usefulness was scrutinized in related literature.

However, prior to the analysis of the five phases, a brief history of digital forensics is presented followed by a summary of today's challenges. There will then be reviews of literature in different branches of digital forensics (e.g., *cybercrime investigations* vs. *dead-box forensics*), followed by the categories of proposed solutions for forensics challenges (i.e., *management approach* and *investigative process* solutions). At the end of this chapter a discussion on the statistical models and a summary will be presented.

Figure 1 depicts a summary of this chapter. The grayed-out boxes indicate the topics that are important to briefly review but are not the focus of this research.

Figure 1
Literature Review Summary



History

Reddy (2019) provided a brief history of digital forensics, which started with a few hobbyists in 1980s accumulating information with no methodology or scientific processes. The term *computer forensics* was a substitution for the process in 2002 by the Scientific Working Group on Digital Evidence (SWGDE) with their published paper, "Best practices for computer forensics" (Reddy, 2019). According to Reddy (2019), the subsequent rise in computer-aided crimes led the FBI to bring a level of

scientific rigor to this emerging field in order to use it as an investigative tool. This rather new field within the computer science did not gain popularity until a high-profile serial killer was caught solely based on the digital forensic evidence (White et al., 2011). In that case, the Bind Torture Kill (BTK) Killer was caught based on the forensic examination of metadata from a single computer file on a floppy diskette, which ended his 30-year murder spree in 2005 (White et al., 2011).

Challenges in Forensics

The metadata in the BTK Killer's floppy diskette provided the only clue in identifying the suspect (White et al., 2011). The metadata was a new way for Windows Vista to index files, which provided new artifacts for investigators compared to earlier versions of Windows (Hayes & Qureshi, 2009). The arrest of the BTK Killer not only attracted the attention of the law enforcement agencies to digital forensics, but it also highlighted the dynamic nature of forensics: That changes to the operating and file systems occur, and when they do, the methods by which evidence could be obtained may have to be recalibrated (Hayes & Qureshi, 2009). The issue of constant changes to computer systems (Hayes & Qureshi, 2009) is not the only hurdle in digital forensics. Additionally, the large variety and new complexities in file types, directories, and other stored information on computers (McKemmish, 1999), and in mobile phones (Kim et al., 2016), have necessitated constant training for digital forensic examiners, along with additional hours spent on examining the digital media. An example of these complexities is data encryption, which according to Balogun and Zhu (2013) has rendered 60% of computer crimes non-prosecutable. Another example of complexity is from the enhancements in wiping and anti-forensics techniques (Ölvecký & Gabriška, 2018). Furthermore, the gradual increase in the number and size of digital media (Quick & Choo, 2014; Yang et al., 2016), and the steady increase

in computer-related crimes each year (Internet Crime Complaint Center, 2019) have contributed to the backlog problems in today's digital forensic labs.

Challenges in Providing Solutions

Providing research-based solutions to solve the backlog problems suffer from three main problems:

1) Many studies suffered from lack of actual police cases for testing, as seen in Yang et al. (2016). The minimal number of published real police cases has been a well-known impairment to digital forensic studies: While collaboration among cyber criminals increase their sophistication, collaboration between police and academic research groups is minimal (Vincze, 2016; Yang et al., 2016). This could be due to a number of reasons from individual's right to privacy, to the classified nature of police work (Hong, et al., 2013; Irons & Thomas, 2014).

2) A second problem stems from digital forensics still being considered in its infancy (Casey, 2020; Du et al., 2017), with limited literature and research compared to other areas of computer science (Horsman et al, 2014).

3) Despite the existing guidelines by International Organization for Standardization (ISO) and other institutions (Veber & Klíma, 2014), digital forensics lacks any uniformly accepted techniques to examine digital media (Alshebel, 2020), or interpret the results (Casey, 2020). This may stem from the variations in operating systems, file formats, etc. (Lillis et al., 2016), or the differences in managing digital forensic investigations (Sudyana et al., 2019). Even the level of an examiner's knowledge lacks any agreed upon qualification standards (Cusack, 2019; Jiang et al., 2015).

Cybercrime Investigation vs. Dead-box Forensics

Despite the lack of standardization, soon after discovering digital forensics' usefulness in investigations, other areas of interest were discovered. Daniel and Daniel (2012) identified a few of these ever-increasing categories as *social media forensics*, *digital multimedia forensics*, *multiplayer game forensics*, etc. However, based on literature reviews, two primary digital forensics categories are:

1) Cybercrimes: Inspection of a compromised or compromising computer system, examination of cyber tools and logs, Intrusion Detection Systems logs (IDS) (Hungwe et al., 2019), examination of digital database integrity (Leigland & Krings, 2004), or inspection of network and cloud systems (Vincze, 2016). Many such applications of digital forensic have to take into consideration the volatility of the network in order to maintain the reliability and integrity of the evidence (Munkhondya et al., 2019), and include software forensic tools that proactively preserve data (Pasquale et al., 2018).

2) Criminal (and Civil) *dead box* investigations: Examination of computer systems to determine if a crime (or a civil infraction) was committed, and if the defendant committed it. This type of investigation primarily examines a *dead box*, where the digital medium is not powered, or is maintained in a *rest* state (Delija, 2017; Dolliver et al., 2017). It has to be noted that live RAM acquisition in this type of examination is the exception and requires a powered digital device with an operating system (Meyers et al., 2017).

Cybercrime investigation is for crimes performed using a computer (e.g., hacking), while dead-box examination is for crimes (or civil infractions) that pre-existed the invention of computers and do not necessarily require a digital device (e.g., homicide, identity theft, etc.) (Vincze, 2016). Both areas of digital forensic

research include the identification, acquisition, and analysis of digital evidence as their ultimate goal (Du et al., 2017). However, the type of evidence that is being sought necessitates different approaches to the examination (Dilijonaite, 2017). The focus of this research is on the second type of digital forensics where the examined data is obtained from dead boxes.

Solution Types

Several proposed solutions have been offered to ease the backlog of dead-box cases in forensic labs. These can be categorized into a *managerial approach* (Ademu et al., 2011; Englbrecht et al., 2020; Park et al., 2018), and an *investigative process* (Cantrell & Dampier, 2013; Karresand & Shahmehri, 2006; Quick & Choo, 2018b; Scanlon, 2016).

Digital Forensics Management Approach

Many researchers have emphasized preparations prior to the digital evidence collection, such as structuring a management body for forensic organizations (Grobler et al., 2010). Although the management aspect of digital forensic is not the focus of this research, related literature will be briefly reviewed due to its integral role.

Digital Forensic Readiness (DFR) is a proactive approach that may increase a forensic lab's capabilities (Englbrecht et al., 2020; Kebande et al., 2018; Kerrigan, 2013; Quick & Choo, 2018b). Rowlingson (2004) suggested that the digital forensics management can be broken down into four elements: Planning, policing, training, and monitoring. Ciardhuáin (2004), Rogers et al. (2006), Perumal (2009), and Agarwal et al. (2011) all included models that started with *planning* or *preparation*, entailing managerial involvement, which is consistent with the ISO/IEC 27000 guidelines.

Kerrigan (2013) presented a five-level evaluation model for DFR ranging from level one where readiness is nonexistent, to level five where the highest level of DFR

is achieved. Kerrigan (2013)'s levels, which later appeared in Englbrecht et al. (2020), included criteria such as processes, people, and technology. As an example, at level five, *processes* are optimized, *people* are highly skilled, and *technology* within the organization is sophisticated enough to develop novel tools. According to Kerrigan (2013)'s DFR levels, the database produced for the methodology portion of this research is from a level four police lab, where processes are quantitatively managed, the people are well-informed of new advancements related to forensics, and the examiners actively adopt new technologies to develop a response to the new changes.

Digital Forensics Investigative Processes

The next category of digital forensic literature deals with the technical aspects of processing data in the form of digital artifacts. This category in ISO/IEC 27042 is described by Veber and Klíma (2014) as *investigative processes* and includes *digital evidence analysis*. The remaining focus of this research will be on digital evidence analysis, while recognizing that despite attempts by ISO and many researchers, there are no standardized techniques to analyze digital media (Garfinkel, 2010; Alshebel, 2020; Casey, 2020).

The newer trend in digital forensics research focuses more on solving specific technical problems (Du et al., 2017). Based on the literature review, it appears that this tendency is also towards the recovery of less (but primary) evidence, instead of a large number of general evidentiary artifacts. The reason for this trend may be the fact that despite the increase in digital contents, files of evidentiary value constitute a small portion of the massive volume of data (Hong et al., 2013).

Cantrell and Dampier (2013) introduced a model in several phases: the *computer profile*, the *crime profile*, and the *presentation* process. It must be noted that one categorization (crime profile) is emphasized in this paper as the *category of*

crime. In the computer profile process, the computer is classified according to its volume and its user directory, while during the crime profile process, the components specific to the crime profile are examined (Cantrell & Dampier, 2013). In the presentation process, the created profiles are presented to a custom-made tool written in Perl. This tool monitors information for keywords to determine a potential crime. As an example, the tool may find five keywords commonly used in child pornography cases. Therefore, it "predicts" that at least five pieces of evidence of such crime will be found on the digital media. Although promising, the study by Cantrell and Dampier (2013) has a heavy reliance on keywords, which could produce false positive results. As an example, due to having many crime-related documents, a lawyer's computer could point to several crimes committed by its user if keywords were the primary source of forensic examination. Such false positive mistakes are documented in research papers including Garfinkel (2012). Vincze (2016) pointed to a broad automated keyword search as one of several processes that may lead to increased forensic time for investigators.

Scanlon (2016) proposed a different approach using a centralized shared database, where metadata from all police-examined digital media could be stored. Scanlon (2016) then hypothesized that this expansive database would assist examiners in identifying similar future metadata from previous cases. A similar idea has been in use through "known hash library", which contains hash-sets of known child pornography images maintained at a central location (Vrubel, 2011). Despite its theoretical potentials, one problem with Scanlon's (2016) research stems from the fact that, excluding child pornography cases, evidence in one case may not have any evidentiary value in another case (Horsman et al, 2014). Another problem with Scanlon's (2016) study is that it places substantial reliance on metadata, which is

subject to change, and vulnerable to manipulation and deletion (Ölvecký & Gabriška, 2018). Scanlon (2016)'s research is similar to Kalker et al. (2001), where "perceptual hashing" was introduced, which is used to circumvent the common hashing problems. It uses the similarity of files, as they are seen by humans, instead of strictly comparing their binary values (Olvecky et al., 2001). However, running perceptual hashing will require significant amount of time (Horsman et al., 2014). A digital forensic examination that is heavily dependent on hash values could be impractical, leading to unreliable and even erroneous results; or it may be too time-consuming.

Investigative Process Sub-categories

Further research into the digital forensics investigative process reveals several specific sub-categories of Triage, Preview, Imaging, Examination, and Legal phases.

Triage

The word *triage* stems from the field of medicine and entails ranking and caring for patients according to the severity of their injuries when faced with limited time and resources (Moser & Cohen, 2013). Triage in digital forensics was first introduced by Rogers et al. (2006), and shortly after declared by Casey et al. (2009) as an important part of a digital investigation. Since there are no standardized techniques to examine digital media (Garfinkel, 2010; Alshebel, 2020; Casey, 2020), the term triage has had different meanings according to its different attributes (Jusas et al., 2017). Roussev and Quates (2012) described triage as a speedy initial examination of data to find artifacts most pertinent to the case, or to build an understanding of the case prior to a deeper examination. Shaw and Browne (2013) described triage as the examination of selective data as opposed to the entire disk. Alrumaithi (2018) described triage as the process of "ranking various aspects and elements in the digital forensics investigation according to their importance" (p. 41). Shaw and Browne

(2013) emphasized that triage is poorly defined. As stated by Yang et al. (2016), "There exists no concrete mechanism to implement a digital triage process model on general purpose" (p. 712). Therefore, when researchers such as Cantrell et al. (2012) referred to triage as "not a forensic process by definition" (p. 30), and the evidence recovered from it inadmissible in court, it is unclear which definition of the triage process they referred to. Also, when Casey (2013) emphasized that triage may not substitute a more exhaustive analysis, it is unknown what depth of triage the paper referred to, or what it meant by "a more exhaustive analysis". As an example, when an examiner reviews the pictures on a digital medium, and then examines the metadata for any geolocation information, have they exceeded the limits set forth in triage description?

Even the suitability of forensic tools to perform triage is in dispute in scholarly articles. As an example, Roussev and Quates (2012) stated that forensic programs such as AccessData's FTK or EnCase could not be used as triage tools. The explanation provided was that FTK or EnCase should be considered "deep examination tools" with a slow throughput even on fast computers (Roussev & Quates, 2012, p.S61). However, Roussev and Quates (2012) did not take into consideration another AccessData tool called FTK Imager, or FTK's option for a phased analysis such as Field Mode. Using such options could bypass lengthy processes such as data carving, indexing, etc. (Carbone, 2014). Contrary to Roussev and Quates (2012), Montasari (2016) recommended FTK and a write-blocker for an on-site triage. Regarding EnCase, the portable version of EnCase was used in Horsman et al. (2014) as a filetype-based forensic triage tool. Additionally, Ghazinour et al. (2017) included triage as one of many tasks EnCase is able to perform. One reason for the disagreements may be due the fact that triage is a concept that is

dependent on several factors such as, the examiner's knowledge and experience, the proficiencies of the forensic software, and the time allocated to perform a triage (Shaw & Browne, 2013).

Keeping in mind the disagreements in what triage is, or what forensic tool can perform it, the following methodologies have been proposed. Jiang et al. (2015) included six stages in an investigative triage. Stages one through three were comprised of background information, arrangement of examination sequence, and collection of evidence, while the sixth stage was regarding evidence presentation in judicial proceedings (Jiang et al., 2015). Stages four and five were named *triage examination* and *deep examination* (Jiang et al., 2015). In the triage stage, Jiang et al. (2015) recommend that the examiner's progressively accumulative experience from similar prior cases can mitigate the danger of overlooking evidence. In order to achieve the maximum experience, Jiang et al. (2015) recommended studying prior cases to find attributes that increase or decrease the likelihood of data being included as evidence.

Shaw and Browne (2013) started with the notion of triage pre-cursors, which included the review of a digital medium contents (mostly via a write-blocker) or using a modified (light) version of a forensic software program to conduct an onsite examination. Shaw and Browne (2013) then introduced two categories of *administrative* and *technical triage*, while hinting at a third possible category of content triage. During the administrative triage, the forensic lab may present two approaches: 1) Accept all digital media, but prioritize them, or 2) Have the submitting agency articulate that there are grounds for finding evidence on the submitted digital devices (Shaw & Browne, 2013). A third option may be to implement a policy somewhere between the two extremes (Shaw & Browne, 2013). In the technical

triage, Shaw and Browne (2013) included the examination of allocated files, which was referred to as the "low-hanging fruit" (p. 117). The occasional examination of the Registry keys by adequately trained examiners was also deemed important in the technical triage (Shaw & Browne, 2013). As it is used in this research, the triage phase includes Shaw and Browne (2013)'s examination of allocated files and Registry keys. However, the scope of examination in the triage phase of this research is far narrower than Shaw and Browne (2013)'s triage and includes only *targeted* analyses of files and a minimal number of Registry keys (in Windows) or plist files (in Mac) according to the category of crimes. This method adheres to the original intent of Rogers et al. (2006)'s triage, where it is defined as, "processes that are conducted within the first few hours of an investigation" (p. 29).

Preview

Shaw and Browne (2013)'s enhanced preview method was different from triage in that the entire disk was searched. Using a bootable Linux CD, a *Live Set* (database) of allocated files were created, which could later be manually refined to retrieve SQLite databases, or export Registry keys in plain text (Shaw & Browne, 2013). The process was then continued when the entire disk (RAM dump, allocated, unallocated, swap files, and shadow volumes) was acquired (Shaw & Browne, 2013). Additional options were provided to the examiner, two of which included Internet messages, and virus scan (Shaw & Browne, 2013). Typically, the first stage of forensic analysis involves a search of the device for artifacts, which is then followed by the second stage of investigating how the artifact came to be on the digital medium (Shaw & Browne, 2013). Casey et al. (2009) called the first stage triage and stated that there needs to be a second stage between triage and an in-depth examination. Shaw and Browne (2013) stated that their enhanced preview was focused on the first

stage. Since the publication of Shaw and Browne (2013), similar Linux-based forensic tools have been available such as Autopsy-Sleuth Kit (Raychaudhuri, 2019). Autopsy can be used to investigate the second stage of how the artifact came to be on the digital medium. Shaw and Browne (2013)'s forensic software and Autopsy may exceed the amount of time that can be allocated to triage. This is especially the case if, as Shaw and Browne (2013) explained, manual refinement of SQLite database is needed, and the Registry keys are outputted in plain text format with no structure. Therefore, for the purpose of this research, Shaw and Browne (2013)'s method will be considered as preview.

Quick and Choo (2017) proposed collecting a *subset* of data (particular digital evidence in particular locations) rather than creating a bit-by-bit forensic image of the digital media. In subsequent phases the data was parsed through several forensic programs and examined within pertinent directories for a wide variety of pertinent artifacts. This process was shown to be saving time both in Quick and Choo (2017) practical tests, and in Quick and Choo (2018b). However, the saved time was in comparison to a full forensic examination, complete with creating a forensic image. Quick and Choo (2017)'s collection of a subset of data was not only more comprehensive than a digital forensic triage, but also provided clear explanation as to why certain digital artifacts should be searched prior to others. Quick and Choo (2017)'s methodology depended on the previous knowledge about the types of data to be examined, as seen in Jiang et al. (2015). Such information in some police cases may not be available at the onset, and a triage (cursory search of the device for artifacts) may be needed prior to Quick and Choo (2017)'s methodology. Despite a few issues, Quick and Choo (2017) presented an excellent solution similar to the *preview* phase of evidence collection in the sample population collected for this

research. It must be noted that neither Shaw and Browne (2013), nor Quick and Choo (2017), include a bit-by-bit image of the digital media in performing their suggested process.

Imaging-Legal Topics

Since the published article by Carlton (2007), making a forensic copy of the digital media has been an essential part of its process (Brown, 2010). Casey (2011) described the process as creating an identical copy of the digital medium and named it a bit-by-bit forensic copy. Once the process is concluded, a hash value of the created image is compared to the imaged disk (Dancer & Skelton, 2013). This allows the courts to apply the *Daubert* test (Daubert v. Merrell Dow Pharmaceuticals, Inc., 1993) for the reliability of the evidence. Casey (2011) summarized the Daubert criteria for evidence evaluation as follows:

1) Whether the technique can be tested.
2) Whether the technique is prone to error.
3) Whether the technique has been peer reviewed.
4) Whether the technique has general acceptance in the scientific community.

Brown (2010) stated that based on the Daubert criteria forensic (bit-by-bit) imaging is the only acceptable process in digital forensics.

In addition to the Daubert caselaw, amendment to the Federal Rules of Evidence (Federal Rules of Evidence, 2017, Rule 902(14)) went into effect on December 1, 2017 (Ries & Hill, 2017). The amendment directly affects the acceptance of imaged media as evidence:

"Certified Data Copied from an Electronic Device, Storage Medium, or File. Data copied from an electronic device, storage medium, or file, if authenticated by a process of digital identification, as shown by a certification of a qualified person that

complies with the certification requirements of Rule 902(11) or (12). The proponent also must meet the notice requirements of Rule 902 (11)" (Federal Rules of Evidence, 2017, Rule 902(14)).

Rule 902(14) can be broken down into three elements: Digital identification process, qualified person, and reasonable notice (Robins-Kaplan, 2019). The last two elements are regarding the qualification of the person collecting the evidence and providing a written notice (Robins-Kaplan, 2019). The first element directly pertains to forensic imaging. In commenting on this rule, the Advisory Committee described the acceptable imaging process to include collecting hash values and comparing them to be a match, but also noted the rule's flexibility to allow the authenticity of the copied media to include processes other than a hash value verification (Mueller et al., 2020).

Imaging-Technical Issues

As indicated, the Advisory Committee recognized that collecting a hash value is not possible in every instance. This may be due to many challenges in today's technology that did not exist at the time when bit-by-bit imaging was recommended. In his prediction for the future of forensic imaging, Garfinkel (2010) brought forth four problems, one of which was that the digital media size increase will result in having inadequate time to create an image. Other foreseen problems in forensic imaging were cloud imaging, embedded digital storage, encryption, and malware (Garfinkel, 2010). The accuracy of Garfinkel (2010)'s predictions has been demonstrated in research articles many years later. Hemdan and Manjaiah (2021) stated that in a cloud environment the old-fashioned definition of forensic imaging would mean that the entire cloud server has to be confiscated, which would inconvenience other users. Hemdan and Manjaiah (2021) suggested acquiring Virtual

Machine (VM) files with snapshots instead. Makura et al. (2021) acknowledged that the traditional forensic imaging has become a challenge in a cloud environment with no clear guidelines to conduct the task. Regarding encryption, Apple's new Big Sur with T2 chipset will prevent a forensic tool from creating an image when a password is unavailable (Nguyen, 2020). Malware programs, especially the ones affecting Application Peripheral Interface (API) can make it impossible to obtain live forensic image of a digital medium (Jansen et al., 2008; Mothi et al., 2020). As early as 2014, Baier and Knauer (2014) warned the forensic community about AFAUC (anti-forensics of data storage by alternative use of communication channels) as a tool used to obfuscate data on a digital medium. Mothi et al. (2020) discussed the new malware programs and anti-forensic techniques (primarily AFAUC) that causes problem in creating a forensic image of dead boxes. Mothi et al. (2020) explained that in AFAUC, the digital medium is accessed through its diagnostic interface to hide or even obfuscate data. The hidden data will not be in hidden areas (host-protected area and device configuration overlay), thus defeating the forensic imaging process (Mothi et al., 2020). Although cell phone forensic is not covered in this research, it has to be noted that the traditional forensic imaging with matching hash values isn't applicable to cell phones (Dancer & Skelton, 2013; Jansen et al., 2008). This is due to many changes that occur when a phone is connected to a computer (Dancer & Skelton, 2013). These changes could be automatic and unintentional (e.g. written log files), or purposefully done by the cell phone forensic tools or the forensic examiner (e.g. manipulating the system's kernel or curtailing cell phone's defensive systems) (Dancer & Skelton, 2013). The digital storage in cell phones, called NAND, have many commonalities with the solid-state drives (SSD) in computers: SSDs are multiple NAND chips on dies that are packaged into what is known as Multi-Chip

Module (MCM) (Veendrick, 2018). Unlike older electromechanical hard drives, there are three disk management technologies (the TRIM command, wear levelling and garbage collection) that extend the life of SSDs (Nisbet & Jacob, 2019). The wear levelling and garbage collection do not require an SSD to be connected to a computer to start working (Nisbet & Jacob, 2019). Let us suppose that an SSD was forensically imaged and at the time the hash value for the drive and its image matched. And let us suppose that at some point in the criminal process the defense requested that a hash value be taken on the SSD to confirm that the evidence was not tampered. If during the time when the SSD was powered for hashing, it went through the wear levelling and garbage collection process, then the drive is technically altered, and the hash values will no longer match (Nisbet & Jacob, 2019). Therefore, a bit-by-bit forensic imaging to prevent "tampering of original exhibit" as advocated by Raychaudhuri (2019) (p. 195), may not be a feasible solution.

A scientific approach to forensic imaging requires a framework to ensure the tests are *repeatable* and *reproducible* (Kessler & Carlton, 2014; NIST 2001). Kessler and Carlton (2014) experimented on two different digital media (a SATA hard drive with NTFS file systems, and a USB flash drive with FAT32) while testing the usability of a write-blocker. The results revealed that even without a write-blocker, the USB flash drive created a copy with matching hash value (Kessler & Carlton, 2014). In the case of the hard drive, the copy without the write-blocker had a different hash value (Kessler & Carlton, 2014). However, only two files caused the mismatch in the hard drive hash value, neither one of which were related to user data where evidence is likely found (Kessler & Carlton, 2014). Notably, Kessler and Carlton (2014) concluded that even the mismatched hash value did not preclude the hard drive

copy from scientific repeatability. Kessler and Carlton (2014) explained that, "repeatable findings remain intact for the content of stored files, file slack, the overwhelming majority of unallocated space, and unused space. In fact, the individual hash values of stored files remain identical when the images are compared" (p. 57). It is due to this scientific repeatability that the evidence from the hard drive copy with unmatched hash value can be admissible since another examiner can duplicate the finding. "The important lesson to learn is that differences in media hash values do not, by themselves, imply contamination of data" (Kessler & Carlton, 2014, p. 57).

The data for this research would revealed that many hours have been spent on creating bit-by-bit forensic images, when utilizing targeted analysis in the preview phase may have already been scientifically repeatable and reproducible, and the evidence legally admissible. Digital forensic cases vary from situation to situation. The intent of this research is for forensic practitioners to have a scientifically based dialogue on the best practices, and to avoid simply following a tradition. Garfinkel (2010) predicted that analysis models have to be re-imagined. Many years since its acceptance in the forensic community, the time may have come to reevaluate the role of bit-by-bit imaging as a necessary forensic process.

Examination

There has been no clear definition for *full examination* in the literature. Shaw and Browne (2013), Casey et al. (2009), and Casey (2011) mentioned full examination, but advised that the analysis of every byte on a digital medium is impractical or may not be possible. Accordingly, the word *full* was eliminated since none of examinations in this research ever included the analysis of every byte. In order to describe the examination phase, as it is referred to in this research, defining its beginning and end is necessary. In this research, the examination phase starts after

targeted *Preview* is completed and a bit-by-bit forensic image of the digital media has been prepared (when possible). For this stage of analysis, the forensic image is run through a variety of automated forensic software programs, with most options for data retrieval selected. The examination is then stopped once *enough* evidence has been collected. Deciding when to stop the examination of a digital medium has been one of many challenges in digital forensics (Presley et al., 2018). Presley et al. (2018) questioned that once enough evidence has been recovered for a conviction, should the examination continue for more incriminating or exculpatory evidence in every scenario? Presley et al. (2018) stated that preventing exceedingly long analysis is a management decision based on several factors including legal ramifications, resource availability, backlogging issues, etc. Considering that in most cases within this research the evidence was already recovered in the Preview phase, the question to answer is if the cost to benefit ratio justified the next phases.

Legal Examination

The last phase of this study was named Legal examination. At this stage, the examination is concluded, evidence gathered, and the results are reviewed. However, prior to filing for an arrest warrant, the prosecution may request additional evidence. In some cases, such requests may be legally necessary. However, many prosecutors may habitually request such analyses due to lack of forensic knowledge (Goodison et al., 2015; Liles et al., 2009; McNicholas, 2020). It has to be noted that none of the legal examinations in this study were requested by the defense counsel.

Interpretation

Horsman (2020) noted that interpreting the results is arguably different among examiners and that there is no guarantee that two examiners would evaluate the same piece of evidence equally. This may not be surprising, since the decision is ultimately

made by the judges or jurors. However, Sunde and Dror (2019) brought forth the possibility of an examiner's partiality during forensic analysis. For this study, all selected cases were successfully presented in an impartial judicial proceeding, where the evidence items were found to be *objectively* pointing to the defendant's conviction.

Statistical Models

Linear regression will be the primary regression analysis that will be used in this study. In statistical modeling, regression analysis is a method by which the relationship between two or more variables is identified or estimated (Wheelan, 2013). An example of such relationship is smoking and cancer (Wheelan, 2013). The two variables in regression analysis are called dependent and independent variables (Hosmer et al., 2013). One common example of regression modeling is linear regression. In linear regression the analysis goal is to find, "the *best fit* for a linear relationship between two variables" (Wheelan, 2013, p. 139), such as the relationship between weight and height.

Compared to other fields of computer science, there has been limited number of research incorporating statical models into digital forensics. Many such models focused on the criminology and profiling suspects (Antolos et al., 2013; Dzemydiene & Rudzkiene, 2002). Although the study by Taha and Yoo (2018) was also done in the field of criminology, its methodology could be applicable to digital forensics. Taha and Yoo (2018) created a system to identify suspects of a crime. Taha and Yoo (2018) first classified the dataset into categories based on their attributes such as the suspect's method of operation and the category of crimes. Then, the categorization attributes were ranked based on their Information Gains, which subsequently constructed the hierarchy of a decision tree. Utilizing logistic regression, the non-

linear decision boundary of categorization attributes was estimated. In their final phase, for each path on the decision tree, a Chi-square analysis was done to short-list the suspects (nodes with the highest Chi-square value) (Taha & Yoo, 2018). A similar method will be used to construct a model to address one of the research questions (RQ4).

Summary

Digital forensics was invented to use science and technology to produce evidence in judicial proceedings. As the number of digital media, their sizes, and the number of computer-related crimes increased, forensic examiners experienced backlogs. As a solution, the concept of triage was added either to evaluate a digital medium, or in some cases to be a replacement for additional examination. Researchers then found the need for a preview to be an intermediary phase between triage and examination. Others emphasized that a forensic image had to be created prior to the examination. Despite several added phases, legal counsels unfamiliar with digital forensics commonly requested post examination analysis. This trend has had the opposite effect of what was originally intended, and in contrast to the current trend towards the recovery of less (but primary) evidence. The statistical models presented in the next chapter will be used to evaluate the usefulness of each investigative phase.

Chapter 3

Methodology

Introduction

This chapter presents the research methodology. Initially, the source of data is described. This is followed by a description of the statistical models and procedures that were used to test the research questions. In subsequent subsections, research questions are presented as hypotheses, followed by the necessary formulations to empirically test them.

Data Used

The data in this research were primarily collected from digital forensic cases investigated by Detective Kevin Parviz, a taskforce officer (TFO) assigned to the southeastern region of the State of Michigan. Considered cases were criminal cases that were resolved pursuant to digital forensic examination. Each criminal case resulted in the suspect's conviction based on digital forensic findings. All cases occurred between 2012 and 2020. Overview of each case came from Detective Parviz's professional records with each case retrospectively evaluated before data entry. Some criminal cases that relied upon only a few evidentiary artifacts were deemed outliers and were not included.

The variable "category of crime" (COC) was created to reflect whether the case is related to a sexual offense. Thus, category of crime status was recorded as a "Sex Offense" (SO) or "non-Sex Offense" (nSO). The total size of the digital media (in GB) for each case was recorded. Five distinct phases of examinations were established. Briefly, these included the Triage phase (t_1), Preview phase (t_2), and Legal phase (t_4). The third phase (t_3) was split into the Imaging subphase (t_{3a}) and the

Examination subphase (t_{3b}). This is due to the fact that Triage and Preview phases do not include imaging process, however the Examination phase does. Imaging time included the length of time to physically remove the digital media (when applicable) and/or the time it took to setup the digital medium for imaging.

The time accrued at each phase was recorded and rounded to the nearest hour. The type of collected evidence at each phase was recorded and classified as "Multimedia" (MM) and "non-Multimedia" (nMM). "Key Evidence" (KE) reports the phase wherein a primary evidence artifact led the examiner to the discovery of other evidentiary artifacts. Such artifacts, along with the Key Evidence, later contributed to a suspect's conviction. Table 1 reports the conceptualized and operationalized variables collected from each considered case, as well as the created and discretized variables. The table includes the name of the variable, the type of the variable (identification, nominal, numerical, ordinal) and a description of each variable. If the variable was nominal, the levels of the variable were reported.

Table 1

Variables Collected in the Criminal Cases

Name	Type	Description
Year - Nature	Broad Identification	Year Crime Committed, Nature of Crime
COC	Nominal	Category of Crime: SO: Sex Offense, nSO: non-Sex Offense
HD (in GB)	Numerical	Size of Hard Drive in GB
t_1 (Triage)	Ordinal	Time Accrued at Triage Phase
t_1 (MM)	Numerical	Multimedia Evidence Collected at Triage Phase
t_1 (nMM)	Numerical	Non-Multimedia Evidence Collected at Triage Phase
t_2 (Preview)	Ordinal	Time Accrued at Preview Phase
t_2 (MM)	Numerical	Multimedia Evidence Collected at Preview Phase
t_2 (nMM)	Numerical	Non-Multimedia Evidence Collected at Preview Phase
t_{3a} (Imaging)	Ordinal	Time Accrued at Image Phase
t_{3b} (Examination)	Ordinal	Time Accrued at Examination Phase
t_3 (MM)	Numerical	Multimedia Evidence Collected at Image and Examination Phase
t_3 (nMM)	Numerical	Non-Multimedia Evidence Collected at Image and Examination Phase
t_4 (Legal)	Ordinal	Time Accrued at Legal Phase
t_4 (MM)	Numerical	Multimedia Evidence Collected at Legal Phase
t_4 (nMM)	Numerical	Non-Multimedia Evidence Collected at Legal Phase
KE	Nominal	Phase where Key Evidence was Discovered, t_1, t_2, t_3

Selected forensic digital media artifacts in the criminal cases were retrospectively evaluated. Case number, suspects' names, and the crime location were omitted from records. Instead, the year and nature of the crimes were used as a broad primary identification measure during the analysis. Any other identifiable information was not relevant to this analysis and was not published. Due to the unidentifiable case information in this analysis, the Institutional Review Board (IRB) at Nova Southeastern University reviewed and granted approval for *non-human subjects research* on November 16, 2020, IRB # 2020-582.

Exploratory Analysis

An exploratory analysis for the data was performed. Common statistics were generated from the collected data. For each phase, the number of pieces of evidence was tabulated. Each piece of evidence was coded as Multimedia or non-Multimedia evidence. The specific evidence was further partitioned by category of crime (Sex Offense vs non-Sex Offense).

Mean and standard deviation were reported for hard drive size, time accrued at each phase, and count for Multimedia and non-Multimedia evidence at each phase. The two-sample *t*-tests were used to test the difference in means across the category of crime status for these variables. Count and percent were reported for the category of crime status and Key Evidence. Chi-squared test for homogeneity was used to test equality of distributions across category of crime. The resulting *p*-values for each test were reported.

Specific Evaluation Methods

Regression

Linear regression was used to answer Research Question 1, 2, and 3. Coefficient estimates, their respective standard errors, and 95% confidence intervals were estimated for each independent variable. To measure each model's performance, the coefficient of determination was estimated.

Evidence Collection Rate and Relative Efficiency

The Evidence Collection Rate (ECR) was calculated by dividing the count of phase-specific artifacts (Multimedia and non-Multimedia) by the number of phase-specific hours contributed. For research question 5, ECR was calculated and applied to each phase. Relative ECR ratios can be calculated by comparing two specific ECRs; that is, dividing one ECR by a reference ECR. Preliminary analyses

demonstrated that the Legal phase consistently had the most contributed time and would serve as our reference level.

Investigating the Research Questions

RQ1

What is the relationship between hard drive capacity and the evidence collected in each phase?

To answer this question, we performed a multiple linear regression where the outcome of interest was the number of collected evidentiary artifacts in each phase. The independent variables were the hard drive size and the category of crime. After the initial analysis, separate regressions were performed for the number of Multimedia and non-Multimedia evidence in each phase.

$E[y|x] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where:

y: Count of collected evidentiary artifacts

X_1: Hard drive size in GB (continuous)

X_2: Category of crime (dichotomous categorization: SO, nSO)

ε: Gaussian noise

We hypothesized, a priori, that there was a linear relationship between the hard drives size (in GB) and the number of evidentiary artifacts collected. We expected that due to the nature of the crime, the amount of Multimedia evidence would be consistently higher in Sex Offense cases than non-Sex Offense. Thus, we hypothesized different effects of hard drive size on the number of evidentiary artifacts, depending on evidence type. We assumed that the category of crime would

affect the relationship between hard drive size and the evidence recovery at each specific phase. To investigate our hypothesis, we planned to begin with over 100 police cases to estimate the coefficients in our model.

RQ2

What is the relationship between hard drive capacity and the number of hours spent in each phase?

To answer this question, we performed a multiple linear regression where the outcome of interest was the count of hours accumulated at each phase. The independent variables were the hard drive size and category of crime. After initial analysis, separate regressions were performed for each investigation phase and results were reported and compared.

$E[y|x] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, where:

y: Count of phase specific hours accumulated.

X_1: Hard drive size in GB (continuous)

X_2: Category of crime (dichotomous categorization: SO, nSO)

ε: Gaussian noise

We hypothesized, a priori, that there was a linear relationship between the hard drives size and the hours spent on each criminal case. We also hypothesized that due to the nature of the crime, the amount of Multimedia evidence was consistently higher in Sex Offense cases than in non-Sex Offense cases. Therefore, we assumed that the category of crime would affect the relationship between hard drive size and hours at each specific phase. We further hypothesized that different phases would

have different temporal demands depending on the cases, leading to different phase-specific estimates for the relationship between hard drive size and phase-specific evidence collected.

RQ3

What is the relationship between the hours spent in each phase and the total recovered evidence?

To answer this question, we performed a multiple linear regression where the outcome of interest was the count of evidence collected during the investigation. The independent variable was the number of hours accumulated at each phase. After the initial analysis, separate regressions were performed for each phase and results were reported and compared.

$E[y|x] = \beta_0 + \beta_1 X_1 + \varepsilon$, where:

y: Count of phase specific, collected evidentiary artifacts,

X_1: Number of phase specific hours accumulated.

ε: Gaussian noise

We hypothesized, a priori, that there was a linear relationship between hours accumulated per case and the number of collected evidentiary artifacts. We also hypothesized that different phases would have different temporal demands depending on the cases, leading to different phase-specific estimates for the relationship between hours accumulated and phase-specific evidence collected.

RQ4

Which investigative phase is most likely to produce Key Evidence for varying categories of crime?

To answer this question, for each category of crime we estimated $p[j]$, the proportion of cases where Key Evidence was discovered in phase j. Results were tabulated and reported. Chi-squared tests was used to test consistency of Key Evidence across each phase. In the event of small cell counts, Fisher's exact test was utilized. Chi-squared tests were also used to test the dependence of Key Evidence across category of crime.

We hypothesized, a priori, that $p[j]$ would not be equivalent across the investigated phases—phases would have distinct differences in Key Evidence discovery. Further, we hypothesized that the distribution of $p[j]$ of Key Evidence would be dependent on the category of crime.

RQ5

Which phase is most efficient in terms of evidence collected per hour?

To answer this question, we initially calculated Evidence Collection Rate (ECR). ECR was calculated by dividing the count of phase-specific artifacts (Multimedia and non-Multimedia) by the number of phase-specific hours contributed.

$$ECR = \frac{Count\ (\ Evidence\ Collected\ for\ MM\ and\ nMM\ per\ phase)}{Count(Hours\ Contributed\ per\ phase\ )}$$

ECR: Evidence Collection Rate

Evidence Collection Rate Ratios (ECRR) were then calculated to compare each individual phase. ECRR was calculated for each time period in a phase (Triage, Preview, Examination, Legal). Our preliminary analysis indicated that the Legal

phase was the longest, on average. Thus, it served as our reference level. Resulting rate ratios among each period and the Legal period was calculated and 95% confidence for the rate ratios was also reported. ECRRs were useful to compare the relative effectiveness of each phase against the Legal phase. Breakdown by Multimedia and non-Multimedia evidence was also performed.

$$ECRR_1 = \frac{ECR(Triage)}{ECR(Legal)}$$

$$ECRR_2 = \frac{ECR(Preview)}{ECR(Legal)}$$

$$ECRR_3 = \frac{ECR(Examination)}{ECR(Legal)}$$

ECR: Evidence Collection Rate

ECRR: Evidence Collection Rate Ratio

We hypothesized that the ECRR would be greater than 1 for all (Triage, Preview, Examination) phases. We also hypothesized the varying ECRR for Multimedia and non-Multimedia in each phase.

Chapter 4

Results

Introduction

This chapter provides data analysis and the results for the previous chapter's proposed methodologies. Analyses include descriptive statistics to summarize the characteristics of the dataset and inferential statistics to find any relationship between specific variables in research questions.

Due to the interwoven nature of Imaging and Examination phases, the time spent on Imaging was added to the Examination phase to create a total number of hours:

$$t_3\ (\text{Image} + \text{Examination}) = t_{3a}\ (\text{Imaging Time}) + t_{3b}\ (\text{Examination Time})$$

Descriptive Analysis

The following table exhibits the descriptive statistics for all variables within the dataset regardless of the category of crime.

Table 2

Descriptive Statistics for the Complete Dataset

n		109
Category of crime (%)		
	Sex Offense	44 (40.4%)
	non- Sex Offense	65 (59.6%)
Hard drive size (in GB) (mean (SD))		910.20 (1770.15)
t_1 Triage Time (mean (SD))		4.73 (6.70)
t_1 Multimedia Count (mean (SD))		89.68 (274.37)
t_1 non-Multimedia Count (mean (SD))		10.18 (22.08)
t_2 Preview Time (mean (SD))		19.17 (34.61)
t_2 Multimedia Count (mean (SD))		245.94 (636.68)
t_2 non-Multimedia Count (mean (SD))		76.26 (97.07)
t_{3a} Imaging Time (mean (SD))		8.02 (7.84)
t_{3b} Examination Time (mean (SD))		24.01 (17.83)
t_3 Imaging + Examination (mean (SD))		32.03 (25.36)
t_3 Multimedia Count (mean (SD))		22.93 (37.23)
t_3 non-Multimedia Count (mean (SD))		25.88 (41.17)
t_4 Legal Time (mean (SD))		7.09 (9.87)
t_4 Multimedia Count (mean (SD))		2.28 (4.94)
t_4 non-Multimedia Count (mean (SD))		3.72 (7.49)
Key Evidence (%)		
	Triage	41 (37.6%)
	Preview	65 (59.6%)
	Examination	3 (2.8%)

Table 2 reports the descriptive statistics for all reviewed criminal cases. Among all reviewed cases, 65 were non-Sex Offenses (~60%) and 44 were Sex Offenses (~40%). The average hard drive size reviewed was ~910 GB (SD 1770). Regarding time for each phase, Examination (t_{3b}) took the longest amount of time [~24 hours (SD 17.83)], followed by Preview (t_2) [19.17 hours (SD 34.61)], Imaging (t_{3a}) [8.02 hours (SD 7.84)], Legal (t_4) [7.09 hours (SD 9.87)] and Triage (t_1) [4.74

hours (SD 6.70)]. Key Evidence was found at the Preview phase ~60% of the time, Examination ~3% of the time, and Triage ~37% of the time. Key Evidence was not found during the Legal Phase. (Further results for Key Evidence are provided under the RQ4 heading.)

Table 3

Descriptive Statistics by Category of Crime

	Sex Offense	Non- Sex Offense	*p*
n	44	65	
Hard drive size (in GB) (mean (SD))	1237.73 (1946.76)	688.49 (1617.92)	0.112
t_1 Triage Time (mean (SD))	6.73 (9.93)	3.38 (2.19)	0.01
t_1 Multimedia Count (mean (SD))	205.27 (406.99)	11.43 (18.44)	<0.001
t_1 non-Multimedia Count (mean (SD))	12.80 (29.22)	8.42 (15.53)	0.312
t_2 Preview Time (mean (SD))	24.18 (37.00)	15.77 (32.75)	0.215
t_2 Multimedia Count (mean (SD))	449.02 (716.33)	108.48 (540.14)	0.006
t_2 non-Multimedia Count (mean (SD))	81.39 (107.49)	72.78 (90.04)	0.652
t_{3a} Imaging Time (mean (SD))	11.05 (10.06)	5.97 (5.01)	0.001
t_{3b} Examination Time (mean (SD))	30.45 (22.11)	19.65 (12.65)	0.002
t_3 ($t_{3a} + t_{3b}$) Time (mean (SD))	41.50 (31.86)	25.62 (17.33)	0.001
t_3 Multimedia Count (mean (SD))	42.52 (50.46)	9.66 (13.59)	<0.001
t_3 non-Multimedia Count (mean (SD))	20.70 (27.58)	29.38 (48.15)	0.282
t_4 Legal Time (mean (SD))	7.82 (13.32)	6.60 (6.69)	0.53
t_4 Multimedia Count (mean (SD))	3.55 (6.96)	1.42 (2.59)	0.026
t_4 non-Multimedia Count (mean (SD))	3.80 (9.69)	3.68 (5.62)	0.936

Table 3 reports descriptive statistics comparing Sex Offense and non-Sex Offense cases. Regarding times, significantly more time was spent on Sex Offense cases than non-Sex Offenses at: Triage (~3.4 hours), Imaging (~5 hours), and Examination (~11 hours). This significance is also seen in the higher mean count for Multimedia evidence collected at Triage, Preview, Examination and Legal phases.

Figure 2

Hard Drive Size by Count

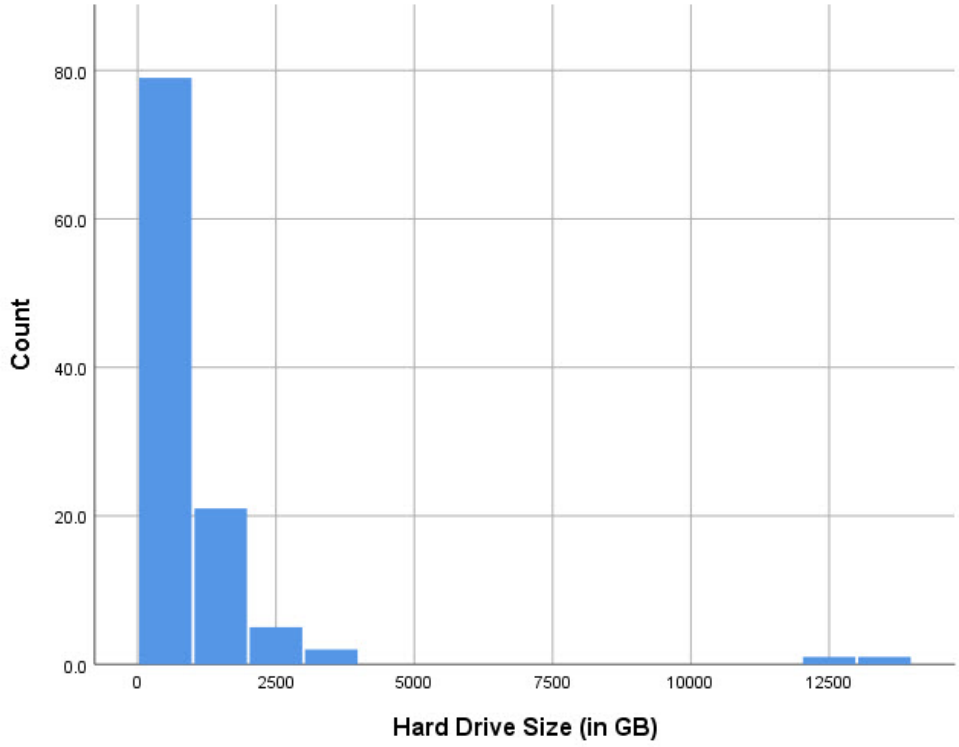


Figure 2 is a histogram that reports frequency of hard drive sizes for all participants in the study. The minimum value is 2 GB while the maximum is 13000 GB. The distribution is right-skewed, with most participants' hard drive size between 0 and 2000 GB.

Figures 3

Hard Drive Size, Categorized by Crime

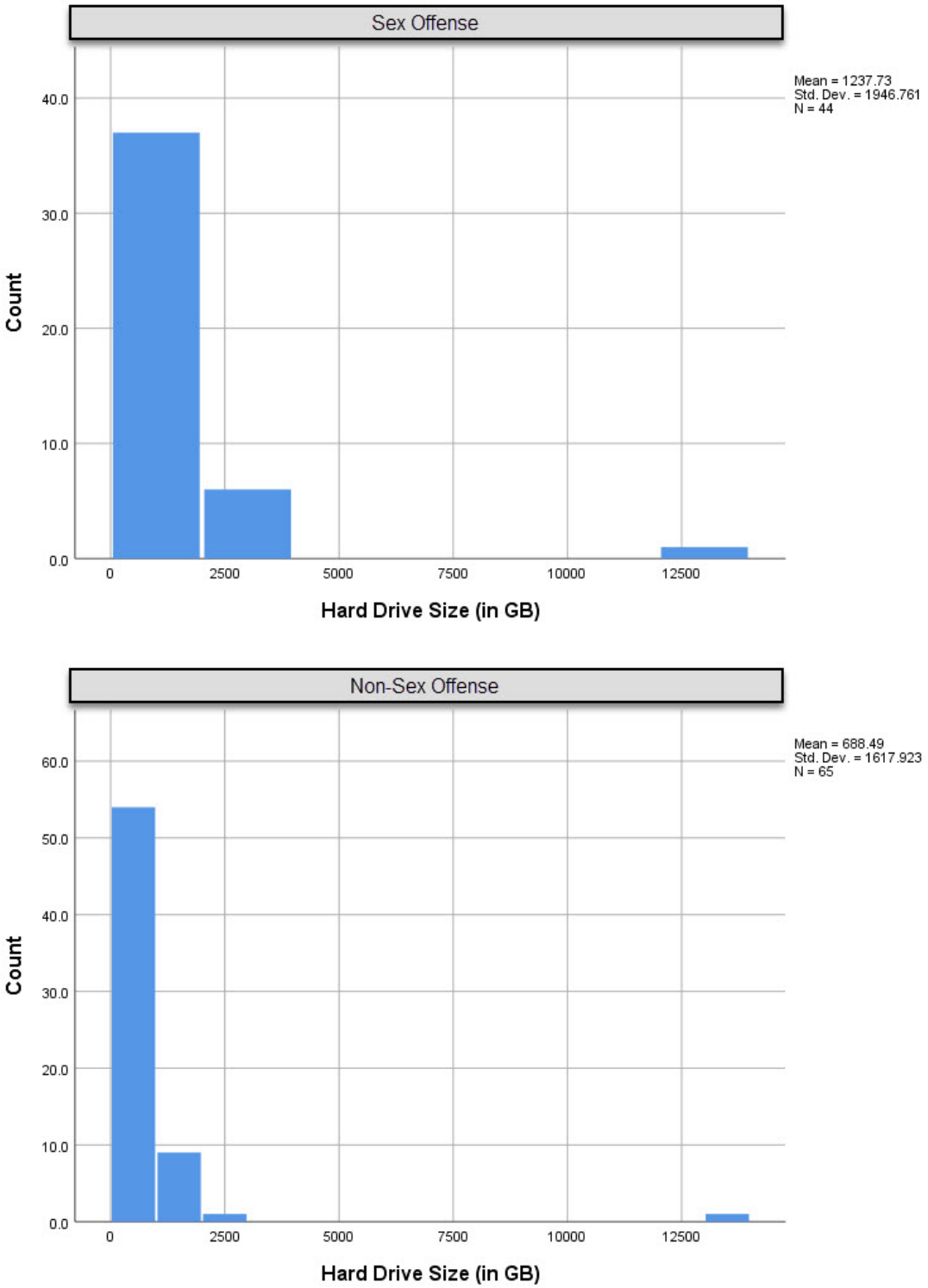


Figure 3 is a histogram that reports frequency of hard drive sizes for all participants in the study, categorized by crime. Both distributions are right-skewed.

Most participants (hard drives) for both categories of crime are between 0-2000 GB. Upon visual examination of the distributions, it was noted that non-Sex Offenses appears to have hard drive sizes lower than 1000 GB for, while there is a more even distribution of the bins from 0-2000 GB in Sex Offense cases. However, as it will be shown later, the results of the *t*-test indicate no significant differences between the means of hard drive size across categories of crime.

Figure 4

Hours Contributed for Each Phase Investigatory Phase

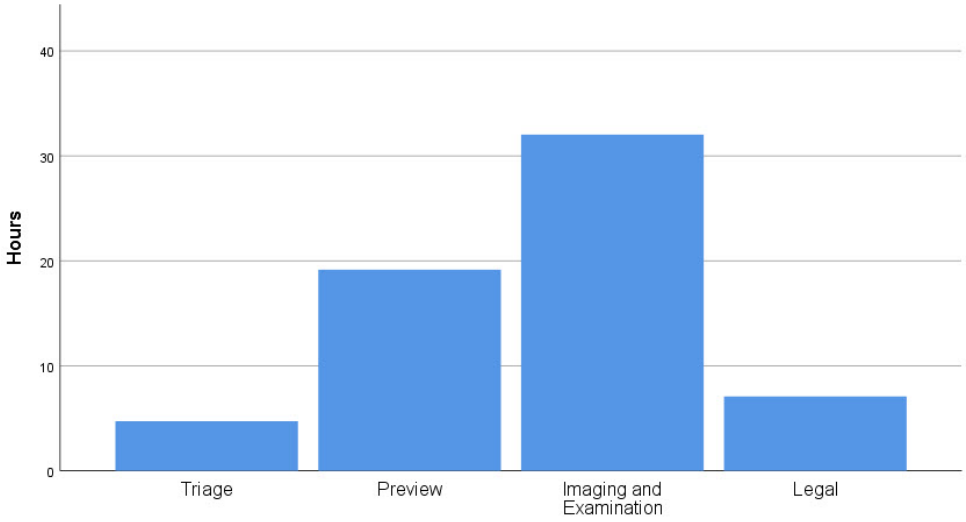


Figure 4 reports the hours contributed for each investigatory phase with the Imaging and Examination (I+E) hours combined. Time in Imaging and Examination phase was the longest (32 hours), then Preview (19 hours), Legal (7 hours) and Triage (4 hours).

Figure 5

Hours Contributed for Each Phase Investigatory Phase by Category of Crime

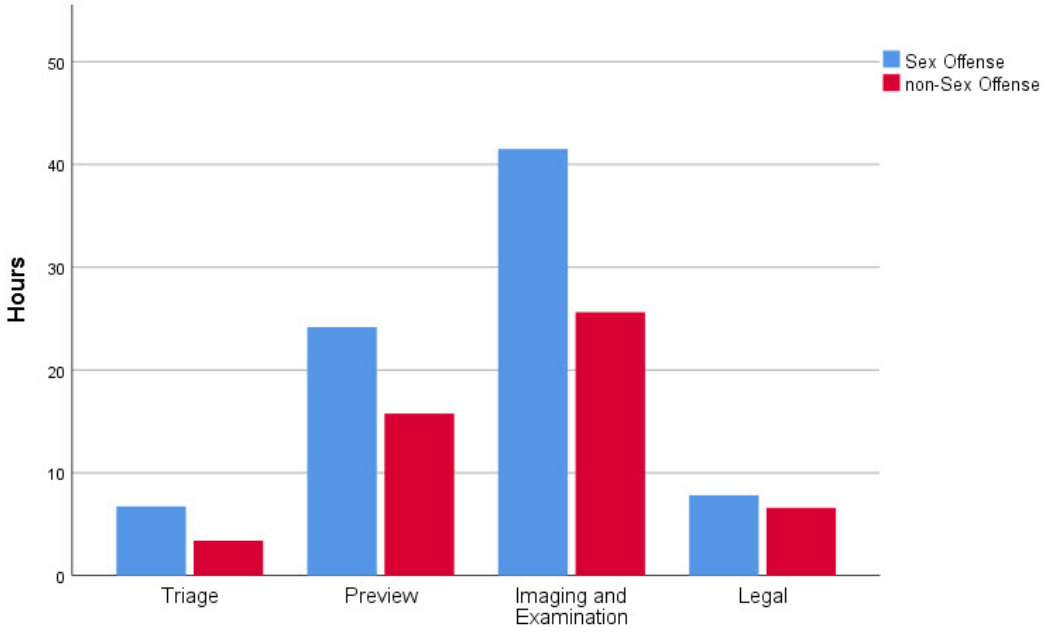


Figure 5 reports the hours contributed for each investigatory phase. Time in the Imaging and Examination phases is the longest for both categories of crime (Sex Offense 42 hours, non-Sex Offense 26 hours). The shortest time is Triage for both crime categories (Sex Offense 7 hours, non-Sex Offense 3 hours). Sex Offense time contributions were visually longer than all non-Sex Offense time contributions.

Figure 6

Evidence Count by Type of Evidence and Phase for Sex Offenses

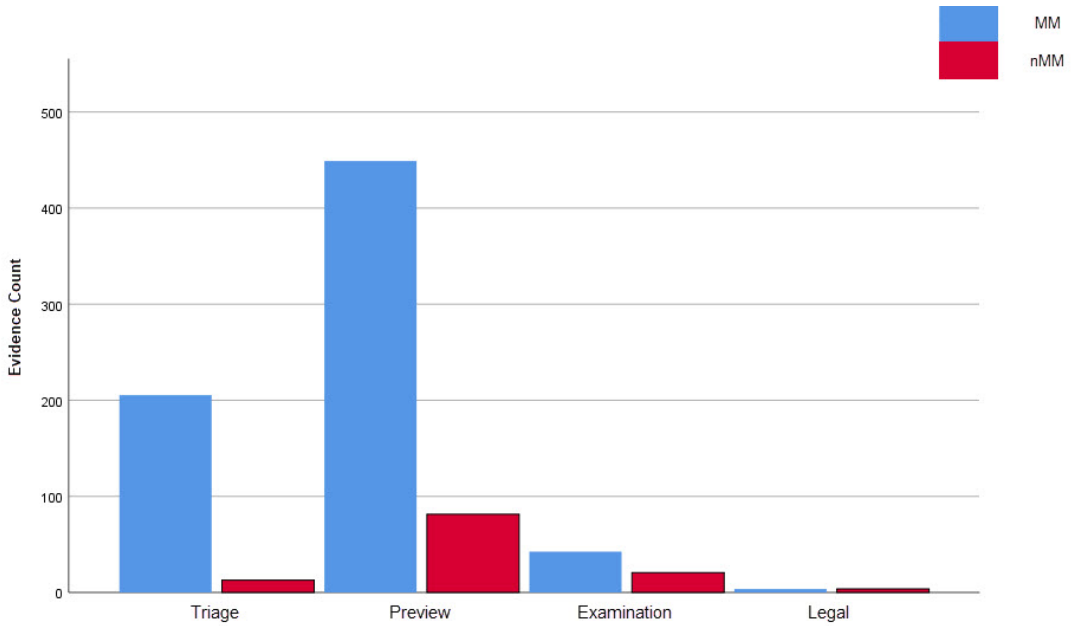


Figure 6 reports evidence counts by evidence type and phase for Sex Offense cases. Count of Multimedia evidence is highest in the Preview phase (Multimedia 449, non-Multimedia 81). Count of evidence classifications is lowest in the Legal phase (Multimedia 4, non-Multimedia 4). Most of the evidence contributed to the Triage phase is Multimedia in nature (Multimedia 205, non-multimedia 13).

Figure 7

Evidence Count by Type of Evidence and Phase for non-Sex Offenses

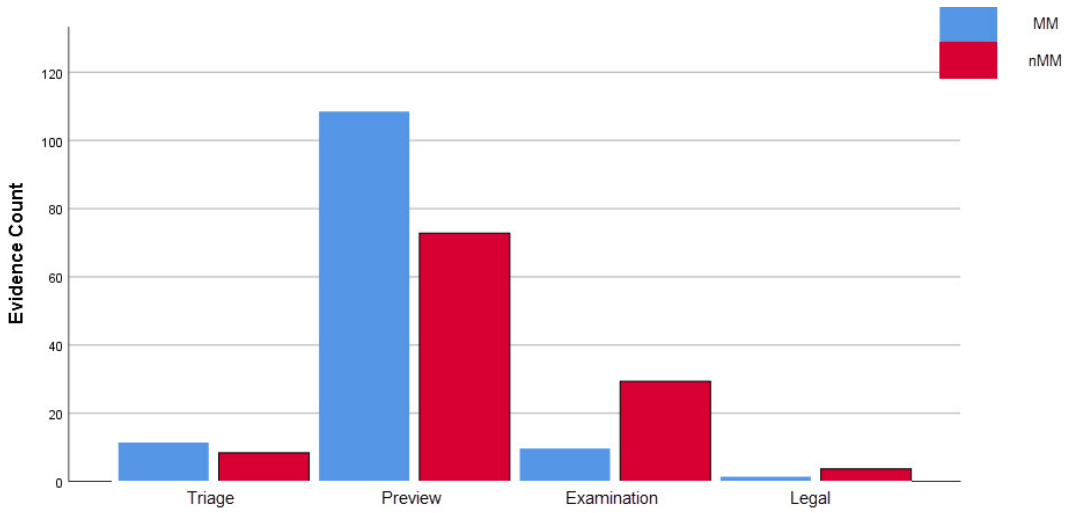


Figure 7 reports evidence count by evidence type and phase for non-Sex Offenses. Similar to Sex Offense cases, evidence count in non-Sex Offenses are most prevalent in the Preview phase and least prevalent in the Legal phase. Regarding the Triage, Multimedia and non-Multimedia evidence appear to be similar (Multimedia 11, non-Multimedia 8). The Preview phase has 181 total evidence items collected (Multimedia 108, non-Multimedia 72).

Research Questions 1, 2, and 3

The following research question results (for RQ1, 2, and 3) contain evidence types (Multimedia & non-Multimedia artifacts) as they relate to Sex Offense and non-

Sex Offense cases. For brevity, the tables presented in this chapter report the values directly pertinent to the research question. An extensive version of these tables, including the *t* and *p* values among other values, can be found in Appendix A.

RQ1

What is the relationship between hard drive capacity and the evidence collected in each phase?

In the previous chapter a multiple linear regression model was proposed for this research question. In order to examine the underlying assumptions for the proposed model, the following two output variables of interest were regressed on the two sets of input below:

Outputs of Interest:

1) Number of Multimedia evidence collected
2) Number of non-Multimedia evidence collected

Inputs:

1) Hard Drive size for Sex Offense cases
2) Hard Drive size for non-Sex Offense cases

The aforementioned regressions were repeated for all four phases of forensic examination (Triage, Preview, Image + Examination, Legal). Table 4 depicts the results of when the output variables of interest are regressed on the inputs of Hard Drive size in Sex Offense cases.

Table 4

Regression of Multimedia Evidence Count on Hard Drive Size in Sex Offense Cases

Outcome	Phase	Intercept	β	R^2
Multimedia	Triage	-31.10	0.19	0.83
	Preview	30.48	0.34	0.83
	Imaging + Exam	38.12	0.00	-0.01
	Legal	3.37	0.00	-0.02

In Table 4 the R^2 values were highest for Multimedia evidence at Triage and Preview phases (R^2 =0.83). All other phases had very poor coefficients of determination. Therefore, in the Triage and Preview phases 83% of the variance for Multimedia evidence were explained by hard drive size in Sex Offense cases. For every 100 GB increase in hard drive size, 19 Multimedia artifacts could be expected in Triage, while 34 Multimedia artifacts could be expected in the Preview phase.

Table 5

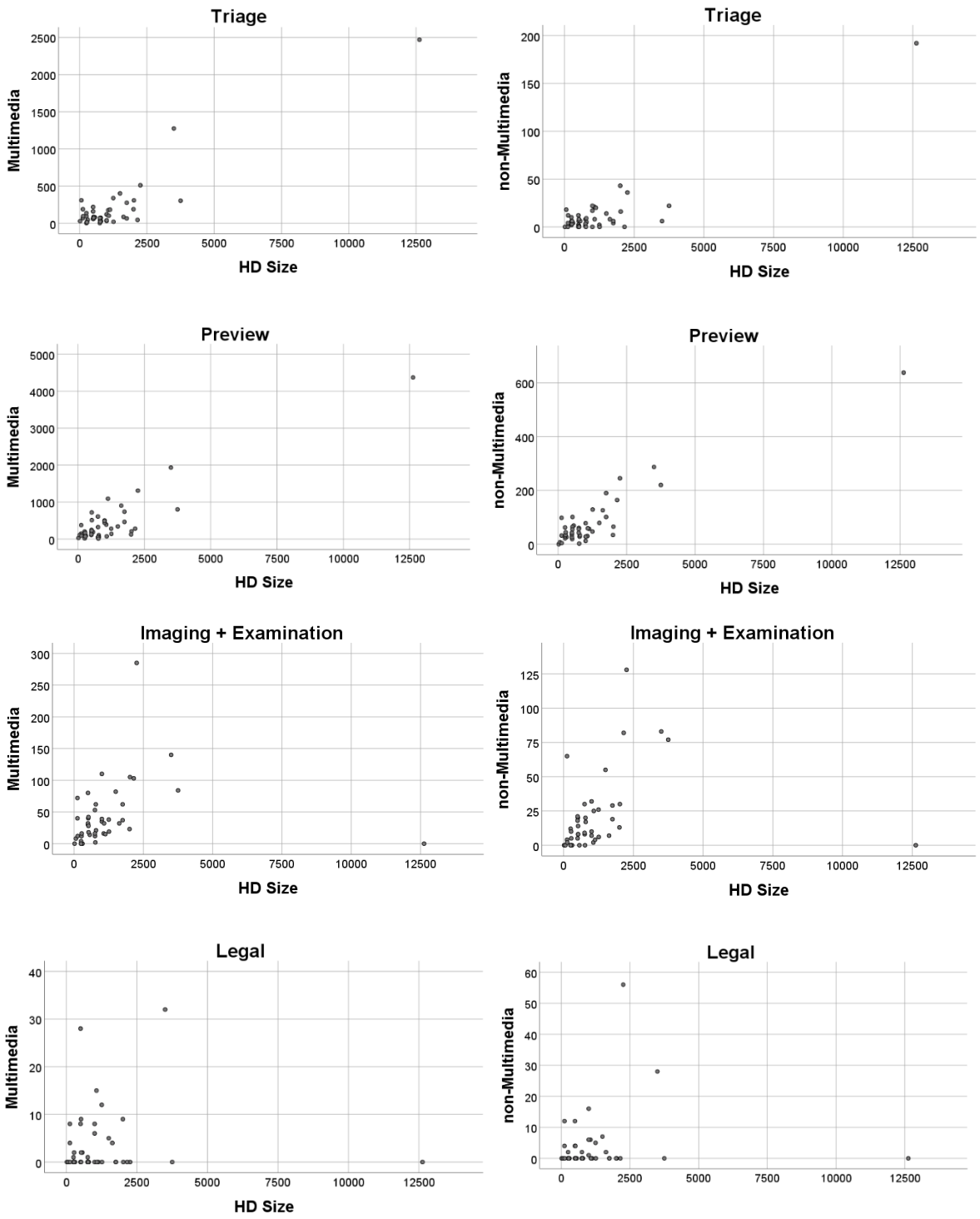
Regression of non-Multimedia Evidence Count on Hard Drive Size in Sex Offense Cases

Outcome	Phase	Intercept	β	R^2
non-Multimedia	Triage	-4.11	0.01	0.82
	Preview	18.00	0.05	0.86
	Imaging + Exam	17.43	0.00	0.01
	Legal	3.18	0.00	-0.01

As in Table 4, Table 5 reports high R^2 values for the Triage and Preview phases ($R^2 > 0.80$). The other two phases had very poor coefficients of determination. Therefore, in the Triage and Preview phases over 80% of the variance for non-Multimedia evidence types were explained by the hard drive size in Sex Offense cases. For every 100 GB increase in hard drive size, 1 non-Multimedia artifact could be expected in Triage, while 5 non-Multimedia artifacts could be expected in the Preview phase.

Figure 8

Scatterplot for the Regression of Evidence Type Count on Hard Drive Size in Sex Offense Cases



Note. The y-axis scales are different according to the data range.

Figure 8 displays scatterplots of both evidence types by Hard Drive size in Sex Offense cases -- in separate phases. In Triage and Preview phases, while the overall clustering shapes appear similar for Multimedia and non-Multimedia, the scales are different. That is, in Triage most Multimedia evidence collected was between 0-500, while non-Multimedia evidence was between 0-50. In Preview most Multimedia evidence collected was between 0-500, while non-Multimedia evidence was between 0-100. As for Imaging + Examination and Legal phases, the clustering appears scattered. For most Imaging + Examination, Multimedia evidence collected was between 0-100, while non-Multimedia evidence was between 0-50. Most of the collected evidence in the Legal phase was below 10.

Table 6

Regression of Multimedia Evidence Count on Hard Drive Size in non-Sex Offense Cases

Outcome	Phase	Intercept	β	R^2
Multimedia	Triage	9.30	0.00	0.06
	Preview	-111.91	0.32	0.92
	Imaging + Exam	9.44	0.00	-0.01
	Legal	1.34	0.00	-0.01

Table 6 reports the regression results for inputs Hard Drive size in non-Sex Offense at each individual phase by Multimedia evidence type. In this table only the Preview phase had a high coefficient of determination ($R^2 = 0.92$). All other phases had very poor coefficients of determination. Therefore, in the Preview phase over 92% of the variance for the Multimedia evidence was explained by hard drive size in non-Sex Offenses. For every 100 GB increase in hard drive size, 32 Multimedia artifacts could be expected in the Preview phase.

Table 7

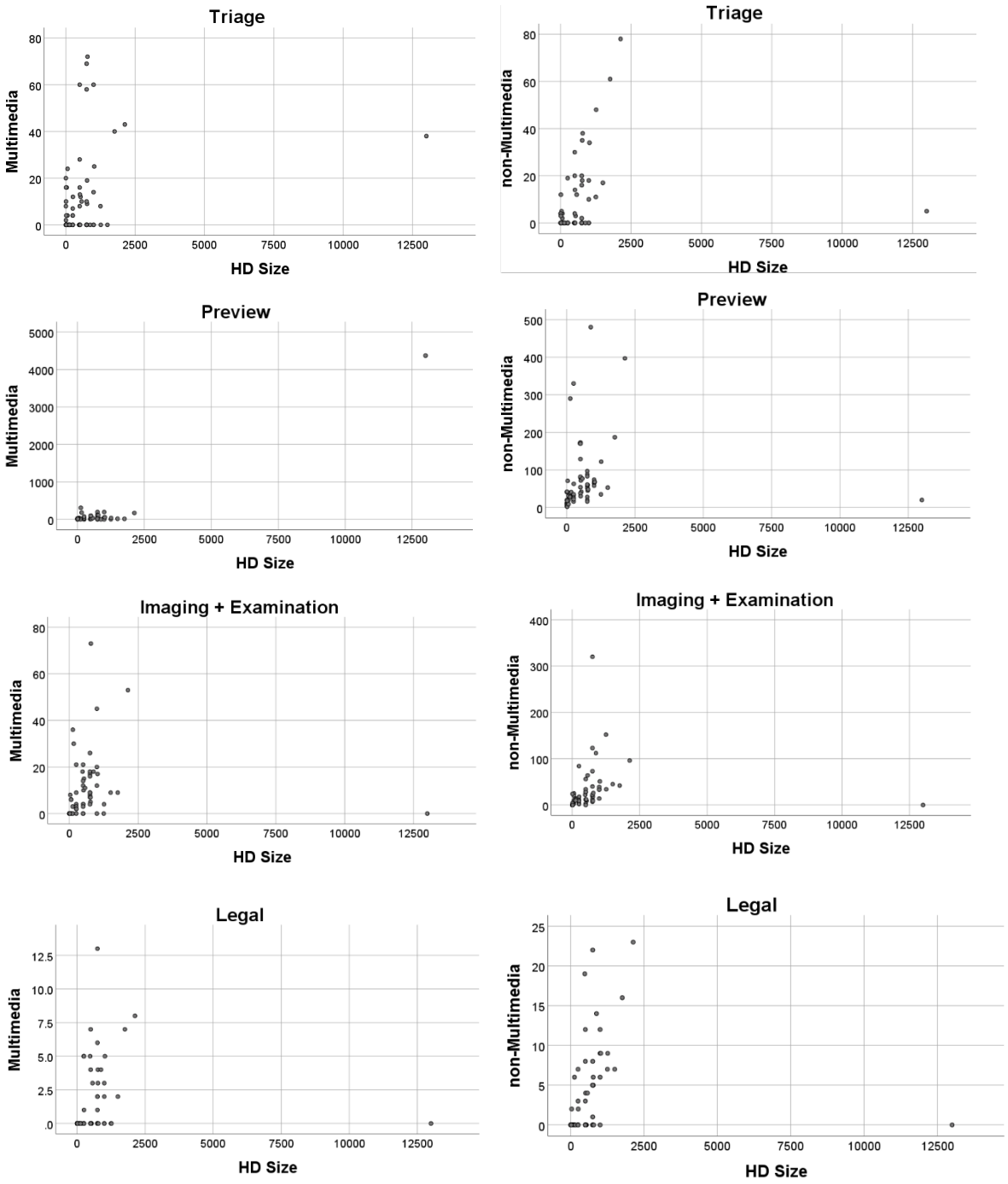
Regression of non-Multimedia Evidence Count on Hard Drive Size in non-Sex Offense Cases

Outcome	Phase	Intercept	β	R^2
non-Multimedia	Triage	7.30	0.00	0.01
	Preview	70.88	0.00	-0.01
	Imaging + Exam	28.39	0.00	-0.01
	Legal	3.42	0.00	0.00

Table 7 reports the regression results for Hard Drive size in non-Sex Offense at each individual phase -- by non-Multimedia evidence type. In this table all phases had very poor coefficients of determination, and no discernable effect was observed between the input and output.

Figure 9

Scatterplot for the Regression of Evidence Type Count on Hard Drive Size in non-Sex Offense Cases



Note. The y-axis scales are different according to the data range.

Figure 9 displays the scatterplots of both evidence types by Hard Drive size in non-Sex Offense cases in separate phases. Although Table 7 reported poor coefficients of determination for the non-Multimedia in the Triage phase, visual examination of Figure 9 shows that a line of best fit could be drawn had it not been for one outlier. Other observations include overall similar clustering shapes for Multimedia and non-Multimedia in Triage with similar scale (0-80), with most evidence being under 40.

RQ2

What is the relationship between hard drive capacity and the number of hours spent in each phase?

In order to examine the underlying assumptions for the proposed model, the output variable of interest (time) was regressed on the two sets of input below:

1) Hard Drive size for Sex Offense cases
2) Hard Drive size for non-Sex Offense cases

The regressions were repeated for all four phases of forensic examination (Triage, Preview, Image + Examination, Legal). Table 8 depicts the results of when the output variables of interest (time) is regressed on the input of Hard Drive size in Sex Offense cases.

Table 8

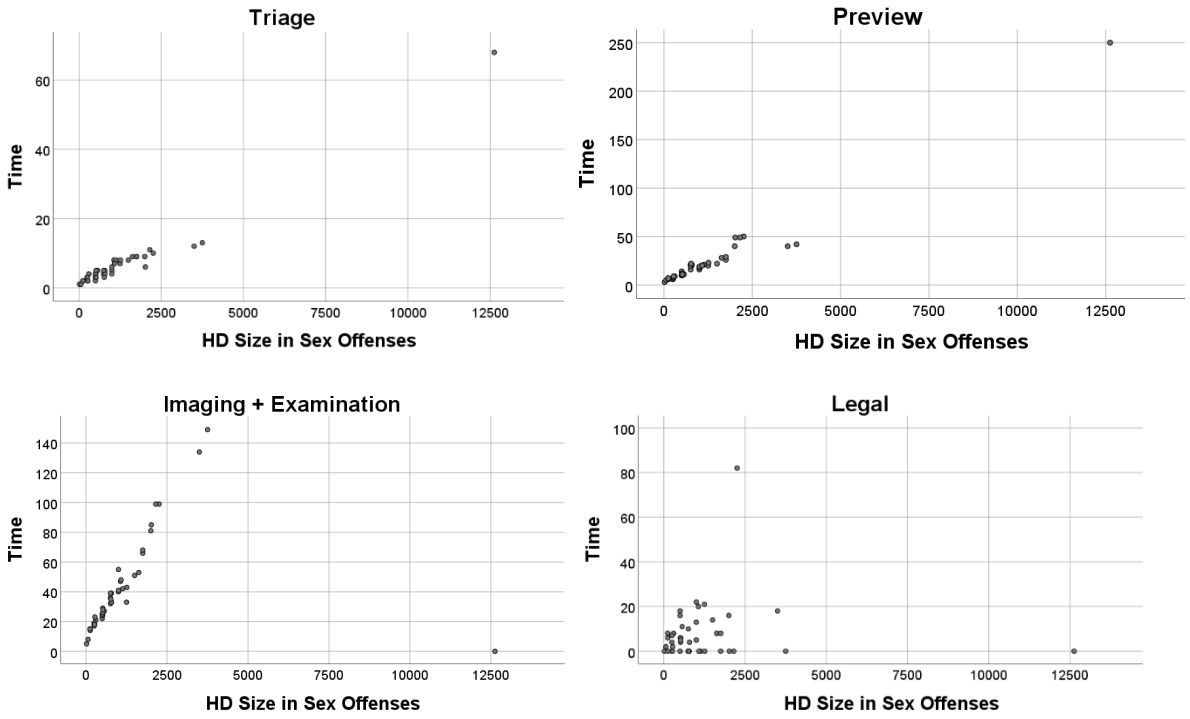
Regression of Time (hrs.) on Hard Drive Size in Sex Offense Cases

Outcome	Phase	Intercept	β	R^2
Time	Triage	0.53	0.01	0.96
	Preview	1.13	0.02	0.96
	Imaging + Exam	36.78	0.00	0.03
	Legal	7.50	0.00	-0.02

According to Table 8, Hard Drive size was predictive of hours spent for the Triage ($\beta = 0.01$, $R^2 = 0.96$), and Preview ($\beta = 0.02$, $R^2 = 0.96$) phases, but not for Imaging + Examination, or Legal phases ($\beta = 0.00$, $R^2 = 0.03$, -0.02). For every 100 GB increase in hard drive size, 1 hour of examination could be expected in Triage, while 2 hours of examination is expected to be spent in the Preview phase.

Figure 10

Scatterplot for the Regression of Time (hrs.) on Hard Drive Size in Sex Offense Cases



Note. The y-axis scales are different according to the data range.

Figure 10 displays the scatterplot for time by Hard Drive size in Sex Offense cases, at different stages of analysis. The trajectories of both Triage and Preview appear to be similar, where most clustering occurs under 20 and 50 hours respectively. As for Imaging + Examination, despite poor coefficients of determination in Table 8, visual examination of Figure 10 shows that a line of best fit

could be drawn had it not been for one outlier. The legal phase shows no apparent effect between time and the Hard Drive.

Table 9

Regression of Time (hrs.) on Hard Drive Size in non-Sex Offense Cases

Outcome	Phase	Intercept	β	R^2
Time	Triage	3.02	0.00	0.14
	Preview	1.89	0.02	0.99
	Imaging + Exam	24.97	0.00	-0.01
	Legal	6.60	0.00	-0.02

Table 9 reports the regression results for the relationship between hard drive capacity and the number of hours spent in each phase -- controlling for the non-Sex Offense cases. Only in the Preview phase Hard Drive size was a strong predictive of the hours spent ($\beta = 0.02$, $R^2 = 0.99$). Therefore, for every 100 GB increase in Hard Drive size, 2 hour of examination is expected to be spent in the Preview phase. Imaging + Examination and Legal phases showed a poor coefficient of determination.

Figure 11

Scatterplot for the Regression of Time (hrs.) on Hard Drive Size in non-Sex Offense Cases

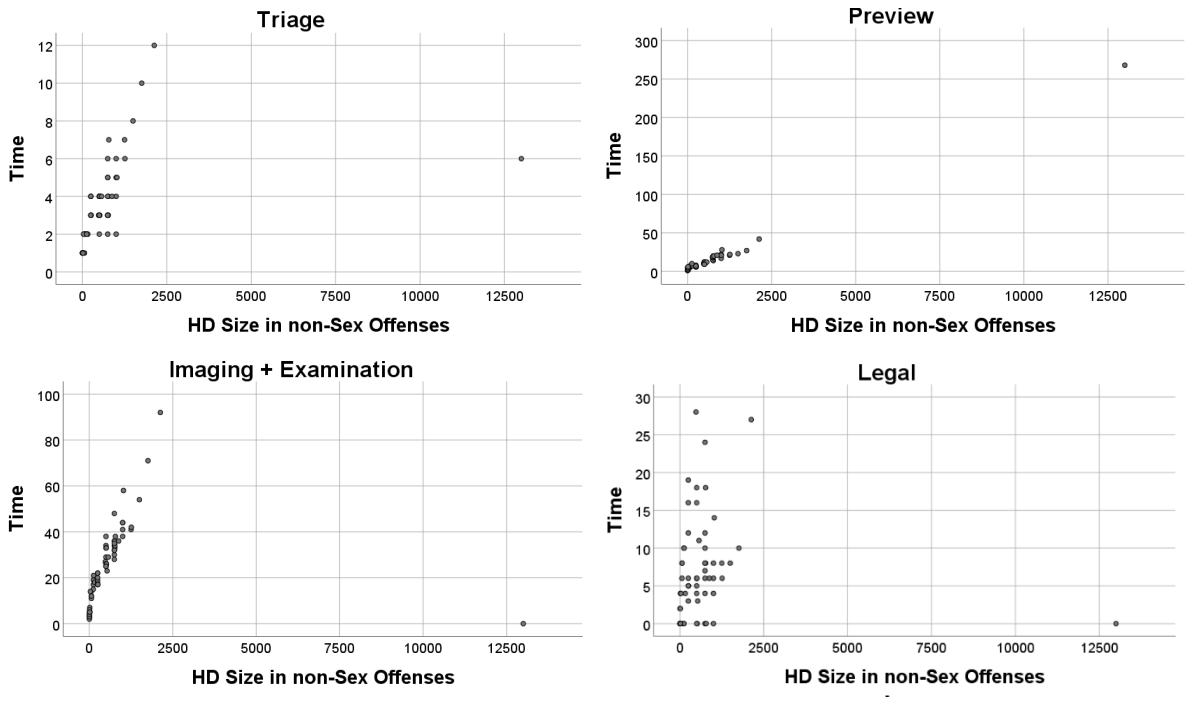


Figure 11 displays the scatterplot for time by Hard Drive size in non-Sex Offense cases, at different stages of analysis. The trajectories of both Triage and Imaging + Examination phases appear to be similar, where most clustering occurs under 8 and 60 hours respectively. For both phases, despite poor coefficients of determination in Table 9, visual examination of Figure 11 shows that a line of best fit could be drawn had it not been for one outlier. The Preview phase in non-Sex Offense cases appears to have the same trajectory as the Triage and Preview phases in Sex Offense cases. The legal phase shows no apparent effect between examination time and the Hard Drive.

RQ3

What is the relationship between the hours spent in each phase and the total recovered evidence?

To examine the underlying assumptions for the proposed model, the following two output variables of interest were regressed on the two sets of input below:

Outputs of Interest:

1) Number of Multimedia evidence collected
2) Number of non-Multimedia evidence collected

Inputs:

1) Time spent in Sex Offense cases
2) Time spent in non-Sex Offense cases

The regressions were repeated for all four phases of forensic examination (Triage, Preview, Image + Examination, Legal).

Table 10

Regression of Multimedia Evidence Count on the Time Spent in Sex Offense Cases

Outcome	Phase	Intercept	β	R^2
Multimedia	Triage	-41.92	36.74	0.80
	Preview	33.39	17.19	0.78
	Imaging + Exam	-0.49	1.04	0.41
	Legal	2.33	0.16	0.07

Table 10 reports the regression results for the Multimedia evidence collected, and the time spent on Sex Offense cases. Both Triage ($\beta = 36.74$, $R^2 = 0.80$) and Preview phases ($\beta = 17.19$, $R^2 = 0.78$) demonstrated strong coefficients of determination values, indicating good model fit. For every 1 hour of examination in Triage, ~37 Multimedia artifacts are expected to be found. In the Preview phase, for every 1 hour spent examining the digital media, the recovery of 17 Multimedia artifacts could be expected. Imaging + Examination exhibited a poor coefficient of determination value, while indicating that for every 1 hour spent in Imaging + Examination, the recovery of ~1 Multimedia artifacts could be expected. Legal phase exhibited the lowest coefficient of determination value, indicating a poor model fit.

Table 11

Regression of non-Multimedia Evidence Count on the Time Spent in Sex Offense Cases

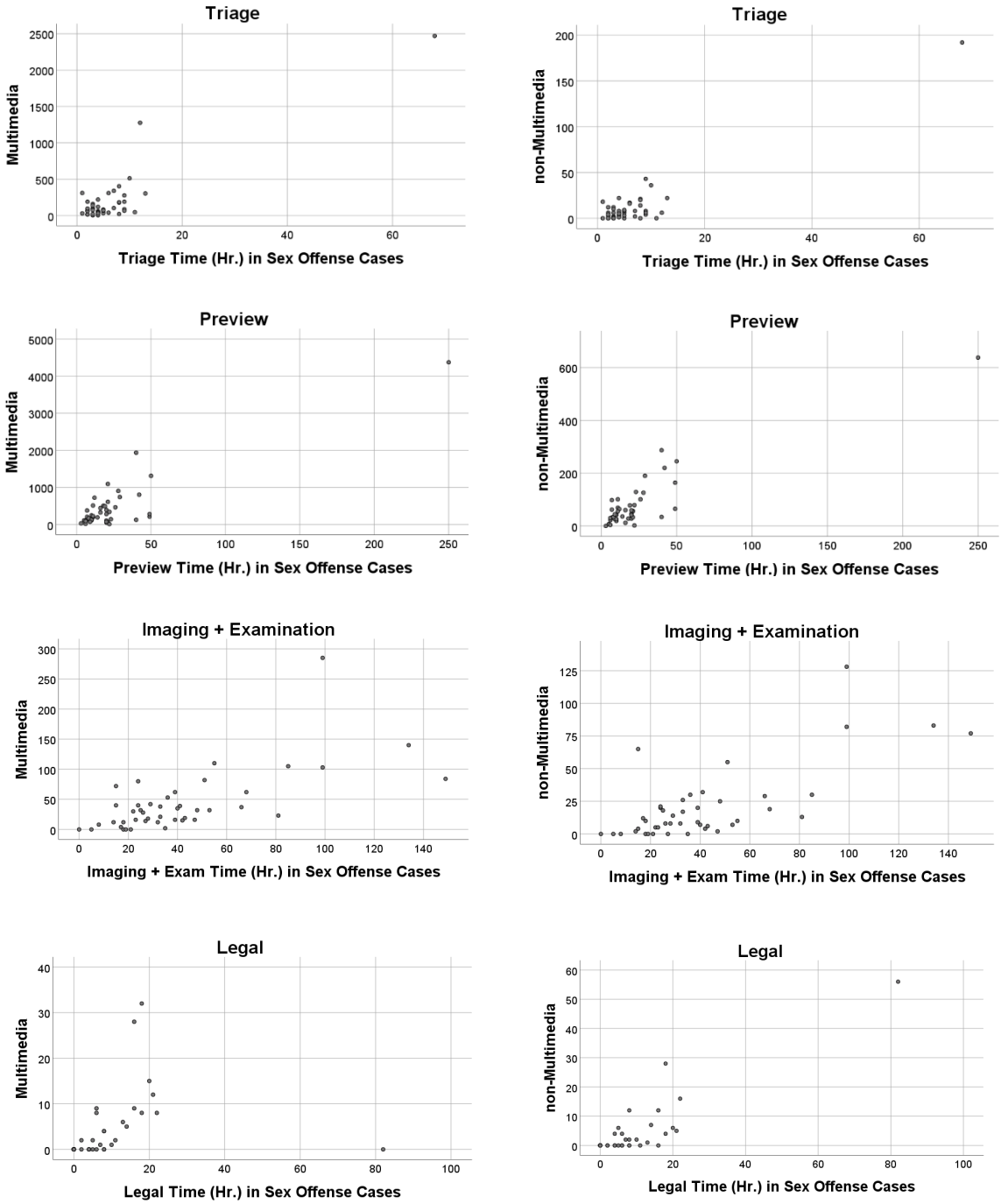
Outcome	Phase	Intercept	β	R^2
non-Multimedia	Triage	-5.80	2.76	0.88
	Preview	18.78	2.59	0.79
	Imaging + Exam	-5.21	0.62	0.51
	Legal	-1.29	0.65	0.79

Table 11 reports the regression results for the non-Multimedia evidence collected, and the time spent on Sex Offense cases. Triage ($\beta = 2.76$, $R^2 = 0.88$), Preview ($\beta = 2.59$, $R^2 = 0.79$), and Legal ($\beta = 0.65$, $R^2 = 0.79$) demonstrated strong

coefficients of determination values, indicating good model fit. Imaging + Examination ($\beta = 0.62$, $R^2 = 0.51$) exhibited a poor coefficient of determination value. For every 1 hour of examination in Triage and in Preview phases, ~3 non-Multimedia artifacts are expected to be found. In Imaging + Examination and Legal phases the numbers decline, where for every 10 hours of examination ~6 to 7 non-Multimedia artifacts is expected to be found.

Figure 12

Scatterplot for the Regression of Evidence Type Count on the Time Spent in Sex Offense Cases



Note. The y-axis scales are different according to the data range.

Figure 12 displays scatterplots of both evidence types by the time input in Sex Offense cases. The figure is in separate phases. In Triage and Preview phases, while the overall clustering shapes appear similar for Multimedia and non-Multimedia, the scales are different. That is, in Triage most Multimedia evidence collected was between 0-500, while non-Multimedia evidence was between 0-50. In Preview most Multimedia evidence collected was between 0-1,000, while non-Multimedia evidence was between 0-200. As for Imaging + Examination and Legal phases, the clustering appears scattered. For most Imaging + Examination, Multimedia evidence collected was between 0-100, while non-Multimedia evidence was between 0-50. Most of the collected evidence in the Legal phase was below 10. For both evidence types.

Table 12

Regression of Multimedia Evidence Count on the Time Spent in non-Sex Offense Cases

Outcome	Phase	Intercept	β	R^2
Multimedia	Triage	-3.32	4.36	0.26
	Preview	-144.31	16.03	0.94
	Imaging + Exam	-1.04	0.42	0.27
	Legal	0.00	0.22	0.30

Table 12 reports the regression results for the Multimedia evidence collected, and the time spent on non-Sex Offense cases. All phases with the exception of the Preview phase displayed poor coefficients of determination values. The Preview phase ($\beta = 16.03$, $R^2 = 0.94$) demonstrated a strong coefficient of determination, indicating good model fit. For every 1 hour of examination in Preview, ~16 Multimedia artifacts are expected to be found.

Table 13

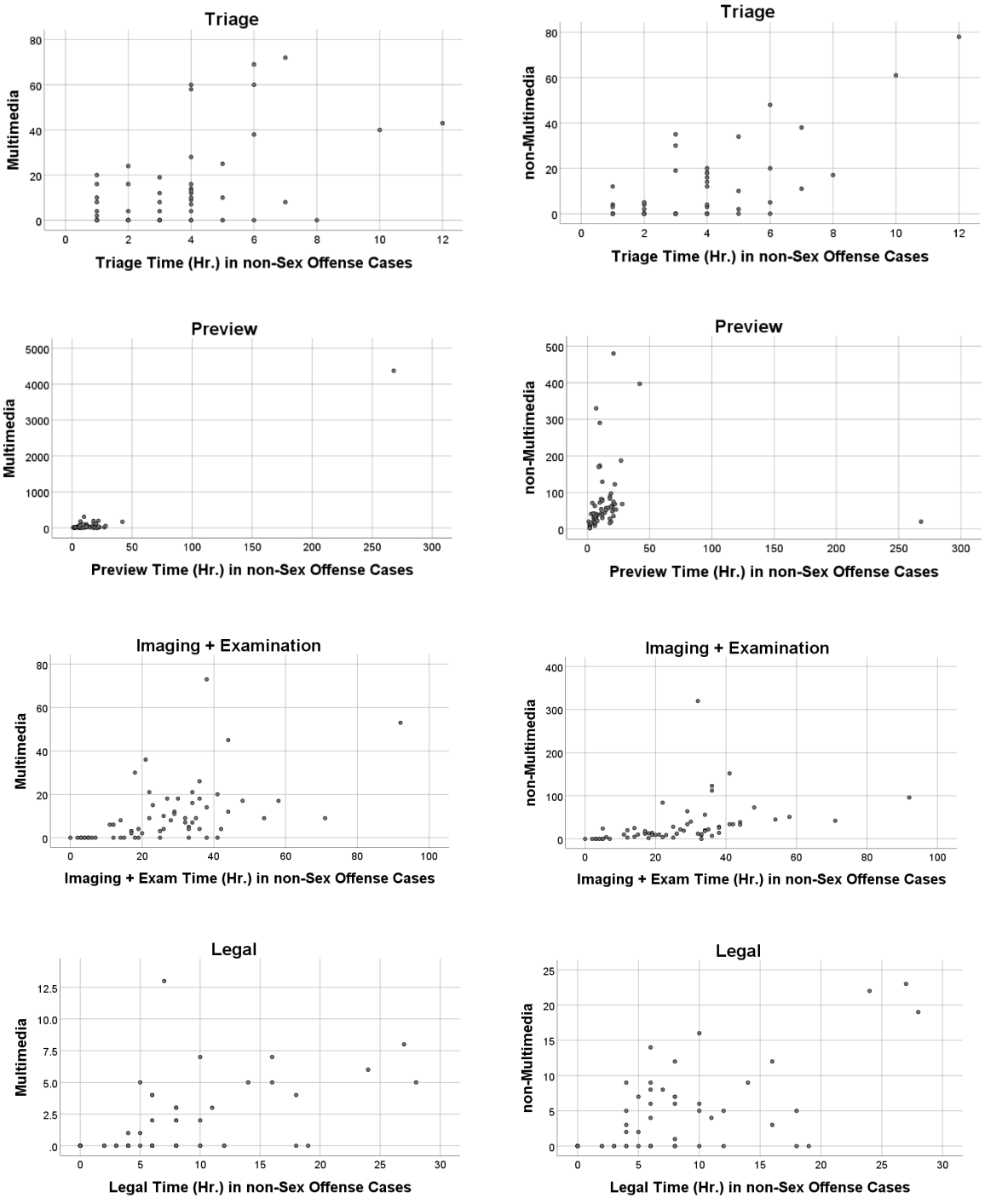
Regression of non-Multimedia Evidence Count on the Time Spent in non-Sex Offense Cases

Outcome	Phase	Intercept	β	R^2
non-Multimedia	Triage	-8.75	5.07	0.50
	Preview	71.16	0.10	-0.01
	Imaging + Exam	-0.22	1.16	0.16
	Legal	-0.06	0.57	0.45

Table 13 reports the regression results for the non-Multimedia evidence collected, and the time spent on non-Sex Offense cases. All phases displayed poor coefficients of determination values indicating poor model fit.

Figure 13

Scatterplot for the Regression of Evidence Type Count on the Time Spent in non-Sex Offense Cases



Note. The y-axis scales are different according to the data range.

Figure 13 displays scatterplots of both evidence types by the time input in non-Sex Offense cases. The figure is in separate phases. In Triage, most evidentiary artifacts were between 0-80. In Preview phases, a high concentration of Multimedia evidence is seen between 0-250, the overall clustering of non-Multimedia evidence is seen between 0-100. In Imaging + Examination, the concentration for both evidence types are seen between 0-80, while in the Legal phase, the clustering appears between 0-12. The distribution of the evidence in Triage and Legal phases appear to be scattered, showing no apparent effect between the analysis time and the number of recovered evidence.

Research Questions 4 and 5

The last two research questions are focused on the quantity of collected evidence, regardless of the type of evidence. That is, the results are geared towards – what is commonly referred to as, *return on investment*.

RQ4

Which investigative phase is most likely to produce Key Evidence for varying categories of crime?

To answer this question, for each category of crime we estimated $p[j]$, which is the proportion of cases where Key Evidence was discovered in phase j. We hypothesized, a priori, that $p[j]$ would not be equivalent across the investigated phases. That is, phases would have distinct differences in Key Evidence discovery. We further hypothesized that the distribution of $p[j]$ of Key Evidence would be dependent on the category of crime. Chi-squared tests was used to test consistency of Key Evidence across each phase. In the event of small cell counts, Fisher's exact test was used. Chi-squared tests were also used to test the dependence of Key Evidence across category of crime.

Table 14

Key Evidence by the Category of Crime Status

Phase	Sex Offense	non-Sex Offense	*p*
Triage	40	1	<<0.01
Preview	4	61	--
Imaging/Exam	0	3	--

Figure 14

Phase by Evidence Count, Stratified by the Category of Crime

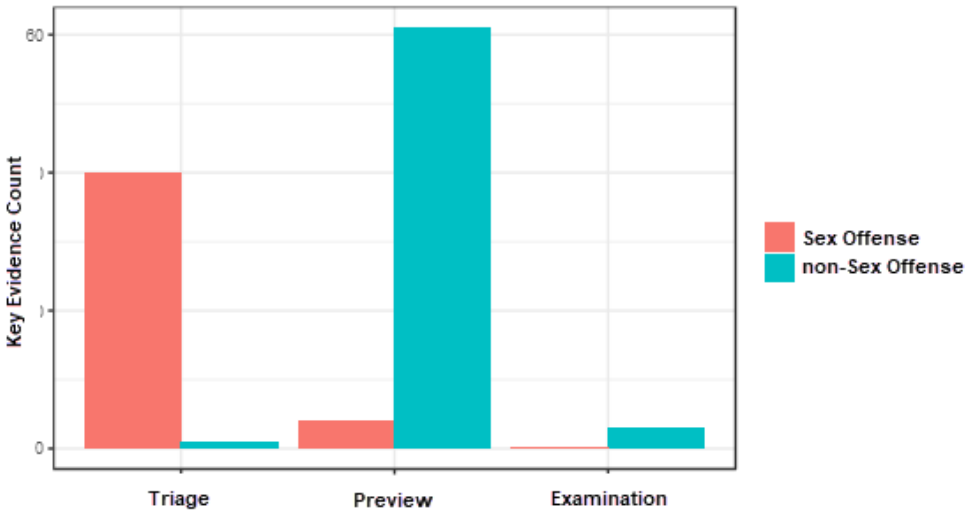


Table and Figure 14 report that Key Evidence was most prevalent in the Triage phase for Sex Offenses and the Preview phase for non-Sex Offenses. Key Evidence was not found in the Examination phase for Sex Offense cases, and not found in the Legal phase for any observed crime. The *p*-value for the chi-square test was much less than 0.01.

Due to the overwhelming differences in distribution of evidence and underlying assumptions of the data, Key Evidence for Sex Offenses is most likely to

be found in the Triage phase, while Key Evidence for non-Sex Offenses is most likely to be found in the Preview phase.

RQ5

Which phase is most efficient in terms of evidence collected per hour?

To answer this question, we initially calculated Evidence Collection Rate (ECR). ECR was calculated by dividing the count of phase-specific evidence type (Multimedia and non-Multimedia) by the number of phase-specific hours contributed.

$$ECR = \frac{Count\ (\ Evidence\ Collected\ for\ MM\ and\ nMM\ per\ phase)}{Count(Hours\ Contributed\ per\ phase\ )}$$

ECR: Evidence Collection Rate

The ECR values for each phase was then divided by the ECR value of the Legal phase, where the Evidence Collection Rate Ratios (ECRR) were calculated as shown below:

$$ECRR_1 = \frac{ECR(Triage)}{ECR(Legal)}$$

$$ECRR_2 = \frac{ECR(Preview)}{ECR(Legal)}$$

$$ECRR_3 = \frac{ECR(Examination)}{ECR(Legal)}$$

ECR: Evidence Collection Rate

ECRR: Evidence Collection Rate Ratio

Table 15

Evidence Collection Rate per Time Period

Phase (t)	Evidence Collected	Time Contributed	Evidence Collection Rate
Triage	10885	787	13.831
Preview	35120	2089	16.81187
Imaging/Exam	5320	3491	1.523919
Legal	654	773	0.846054

Table 15 reports the ECR per investigation phases. The least efficient phase was the Legal phase, finding 0.84 evidentiary artifacts per hour, while the most efficient was the Preview phase, finding 16.81 evidentiary artifacts per hour.

Table 16

Evidence Collection Rate Ratios

Phase (t)	Evidence Collection Rate	ECRR Ref 1	ECRR Ref 2	ECRR Ref 3	ECRR Ref 4
1	13.8310	**1**	0.8227	9.0766	16.348
2	16.8118	1.2155	**1**	11.032	19.871
3	1.52392	0.1102	0.0906	**1**	1.8012
4	0.84605	0.0612	0.0503	0.55523	**1**

Table 16 reports the evidence collection rate ratios with varying refence level.

Table 17

ECRRs and 95% CI

Phase (t)	ECRR	95% CI
Triage	16.35	(15.11, 17.71)
Preview	19.87	(18.39, 21.50)
Imaging/Examination	1.8	(1.67, 1.96)
Legal	**1**	NA

Table 17 reports evidence collection rate ratios and 95% confidence intervals using the Legal phase at the reference level. Relative to the Legal phase, the Preview

phase was ~20 times more efficient (ECRR = 19.87, 95% CI 18.39, 21.50), the Triage phase was ~16 times more efficient (ECRR = 16.35, 95% CI 15.11, 17.71), and the Imaging/Examination phase was ~2 times more efficient (ECRR = 1.8, 95% CI 1.67, 1.96).

Table 18

Evidence Collection Rates by the Category of Crime

Sex Offenses			
Phase (*t*)	Evidence Collected	Time Contributed	Evidence Collection Rate
Triage	9595	296	32.416
Preview	23338	1064	21.934
Imaging/Exam	2782	1826	1.5235
Legal	334	323	0.9671

Non-Sex Offenses			
Phase (*t*)	Evidence Collected	Time Contributed	Evidence Collection Rate
Triage	1290	220	5.8636
Preview	11782	1025	11.495
Imaging/Exam	2538	1665	1.5243
Legal	429	331	0.7716

Table 18 reports the ECR per investigation phases stratified by category of crime. For both categories of crime, the least efficient phase was the Legal phase, finding ~1 (Sex Offenses), and 0.77 (non-Sex Offenses) evidentiary artifacts per hour, while the most efficient was the Triage phase for Sex Offenses finding 32.42 evidentiary artifacts per hour and the Preview phase for non-Sex Offenses, finding 11.50 evidentiary artifacts per hour.

Table 19

ECRRs and 95% Confidence Intervals

Sex Offenses			Non-Sex Offenses		
Phase (t)	ECRR	95% CI	Phase (t)	ECRR	95% CI
Triage	33.52	(29.99, 37.57)	Triage	7.60	(6.73, 8.60)
Preview	22.68	(20.32, 25.39)	Preview	14.90	(13.36, 16.67)
Imaging/Exam	1.58	(1.40, 1.77)	Imaging/Exam	1.98	(1.51, 2.03)
Legal	**1**	NA	Legal	**1**	NA

Table 19 reports evidence collection rate ratios and 95% confidence intervals stratified by category of crime and using the Legal phase at the reference level. Compared to the Legal phase, all other phases remained more efficient. For Sex Offenses, the Triage phase was ~34 times more efficient (ECRR = 33.52, 95% CI 29.99, 37.57) than the Legal phase. For non-Sex Offenses the Preview phase was ~15 times more efficient than the Legal phase (ECRR = 14.90, 95% CI 13.36, 16.67).

Chapter 5

Conclusions, Recommendations, and Summary

Overview

This dissertation presented an empirical investigation into the common digital forensic recovery phases. For that purpose, several statistical models were presented. The highlights of the results in Chapter 4 will be discussed below.

Conclusions and Implications

Within the descriptive results, Table 3 reported the count of both recovered evidence types (Multimedia and non-Multimedia) broken down by the category of crime (Sex Offense and non-Sex Offense). In the Triage and Preview phases, the number of Multimedia artifacts ($n_{(Sex\ Offense)} \sim 205$, $n_{(non\text{-}Sex\ Offense)} \sim 11$) noticeably outnumbered the number of non-Multimedia artifacts ($n_{(Sex\ Offense)} \sim 13$, $n_{(non\text{-}Sex\ Offense)} \sim 8$). This could be simply explained since Multimedia files require a quick visual (or audible) examination of the files, whereas non-Multimedia files such as documents, emails, and texts, require slower process of reading the material. In Triage and Preview, the time spent on a case is important, thus the examination may focus more on Multimedia files. This is regardless of the category of crime. In contrast to that, in the Examination and Legal phases where there are no time constraints, non-Multimedia evidence mostly outnumbers Multimedia evidence.

In addition to time constraints in Triage and Preview, the two phases follow an organized methodology for analysis (certain directories and certain files). This is not the case for the Imaging + Examination or Legal phases, where for most part an examiner would search anywhere within a digital medium. As observed in the results, time spent in the Imaging + Examination phase was therefore the longest (32 hours),

even longer than all other phases combined (Preview (19 hours), Legal (7 hours), and Triage (4 hours)). Research questions 4 and 5 are put forth to investigate if the longer observed time for forensic analysis was justified.

Research Questions 4 and 5

The goal of this dissertation was to empirically investigate the usefulness of different forensic phases when retrieving evidentiary artifacts. RQ4 focused on the quality of evidence, or *Key Evidence*. As explained previously, Key Evidence is a primary evidentiary artifact that provides investigative leads to the recovery of other artifacts. An example of a Key Evidence in an identity theft case would be when the examiner finds a picture of the victim's credit card on the suspect's computer for the first time. This artifact will potentially have a date and timestamp, creating a timeline for other artifacts to be search for. It also may be found in a directory where other relevant artifacts could be found. The results of RQ4 showed that in Sex Offenses, Key Evidence was mainly found in the Triage phase. In non-Sex Offenses, Key Evidence was mainly found in the Preview phase. As a matter of fact, in 109 police cases presented in this study, only three instances of Key Evidence discovery were documented for Imaging + Examination, and none for the Legal phase. Finding such significant piece of evidence primarily in the first two phases of forensic examination emphasizes the importance of including these two phases in the forensic process. Additionally, depending on the type of criminal case and the backlog at the forensic lab, this conclusion may render the remaining two phases of Imaging + Examination and Legal ineffective. This is especially the case if the number of artifacts in the next two phases is small compared to the amount of time that is needed to conduct the examination.

RQ5 specifically aimed to compare the number of recovered evidence in each phase to the amount of time spent (return on investment). The results showed that relative to the Legal phase, the Triage and Preview phases were ~16 and ~20 times more efficient respectively. The results showed high efficiency for both Triage and Preview phases even when the cases were stratified by the category of crime. The implication of the results from RQ4 and RQ5 may be that examiners should seek justification for advancing to the next phases of Imaging + Examination and Legal, contrary to the current (and possibly outdated) practices of automatically advancing to the next phases after Triage and Preview.

Research Questions 1, 2 and 3

Another goal of this dissertation was to provide predictive models for forensic examiners. These models could provide digital forensic practitioners the ability to anticipate the number of evidentiary artifacts they can recover in addition to the amount of time they will spend in specific forensic phases – based on hard drive size.

In RQ1 the relationship between the number of evidence collected and the size of the hard drive was examined. Table 20 depicts a summary of the results showing only the high values for coefficients of determination where hard drive size was a strong predictor of the recovered evidence.

Table 20

RQ1 Summary Results with Strong Coefficients of Determination

Phase	Sex Offense				non-Sex Offense			
	Multimedia		non-Multimedia		Multimedia		non-Multimedia	
	β	R^2	β	R^2	β	R^2	β	R^2
Triage	0.19	0.83	0.01	0.82				
Preview	0.34	0.83	0.05	0.86	0.32	0.92		
Imaging + Exam								
Legal								

Based on this table, for example, an examiner who receives a hard drive for a Child Sexually Abusive Material (CSAM), could anticipate the recovery of 19 Multimedia and 1 non-Multimedia artifacts for every 100GB of hard drive space in the Triage phase.

Furthermore, RQ2 examined the relationship between the time spent examining a case and the size of the hard drive. As it can be seen in Table 21, a systemic pattern of predictability can be found for the Triage and Preview phases in Sex Offense cases, in addition to the Preview phase in non-Sex Offenses.

Table 21

RQ2 Summary Results with Strong Coefficients of Determination

Phase	Sex Offense		non-Sex Offense	
	β	R2	β	R2
Triage	0.01	0.96		
Preview	0.02	0.96	0.02	0.99
Imaging + Exam				
Legal				

Table 21 can be used by forensic examiners to gauge the amount of time they should allocate for Triage and Preview in Sex Offenses, in addition to Preview in non-Sex Offenses. These models estimate that this amount of time is between 1 and 2 hours for every 100GB of hard drive size. Combining RQ1 and 2 in the previous example of the CSAM case; an examiner could not only anticipate the recovery of 19 Multimedia and 1 non-Multimedia artifacts for every 100GB of hard drive space in the Triage phase, but also expect the process to take 1 hour (for every 100GB of hard drive).

Predictability is also seen in RQ3, when the relationship between the number of evidence collected, and the time spent on a case was examined in each category of

crime. Table 22 depicts a summary of the results showing only when time spent on a case is a strong predictor of the recovered evidence.

Table 22

RQ3 Summary Results with Strong Coefficients of Determination

Phase	Sex Offense Multimedia		non-Multimedia		non-Sex Offense Multimedia		non-Multimedia	
	β	R^2	β	R^2	β	R^2	β	R^2
Triage	36.74	0.80	2.76	0.88				
Preview	17.19	0.78	2.59	0.79	16.03	0.94		
Imaging + Exam								
Legal			0.65	0.79				

The results from RQ1 and 3 also show that the hard drive size and time spent on a case follow a systemic pattern – at least for Triage and Preview phases in Sex Offense cases. This pattern is not seen in Imaging + Examination and Legal phases in most of the research question results. The reason could be that the two phases have no time constraints and follow no structured methodology in analyzing the digital media. The processes stop once *enough evidence* has been collected; however, what constitutes enough evidence is unclear (Presley et al., 2018).

Recommendation

Several recommendations could be made to improve similar future studies. Firstly, the sample population in this dissertation was limited in size, which was the reason why *p* values were not greatly considered. A larger sample size in future studies could improve our hypothesis testing. Secondly, the selected police cases had limited variables. Many other police cases that included other factors (listed on p. 8) were not included in this study. Therefore, caution should be rendered when deriving conclusions from this dissertation on police cases that include additional variables not covered in this study. Thirdly, hard drives used in this study were mostly between 0

and 2000GB (Figure 2) from 2012 until 2020. The newer cases in the dataset show gradual increase in capacity to reach as high as 13000GB in 2020 (p. 46). Updated research involving high-capacity hard drives may be needed to test the conclusions derived in this study.

Summary

The purpose of this dissertation was to empirically investigate the problem of delays and backlogs in digital forensic processing. In order to examine this problem and research a possible solution, five common forensic examination phases were studied. These phases included: Triage, Preview, Imaging, Examination, and Legal phases. Each phase was distinctly explained and differentiated. To quantitatively test each phase, 109 criminal cases from police archives were used as sample data. Using linear regressions, chi-square, and Evidence Collection Rate, the results were produced and analyzed. These results revealed noticeably large amounts of time spent on the Imaging + Examination, and Legal phases. This was despite the fact that Key Evidence was likely to be found in Triage and Preview phases, and that the evidence collection rate for Imaging and Examination and Legal phases was extremely low.

In addition to the quantitative study of each phase's usefulness, within literature review, the controversial topic of Imaging was reviewed. The literature review clearly explained the legal and technical implications of creating a bit-by-bit image of a hard drive. At the conclusion of the literature review it was suggested that contrary to the current practices of routinely creating a forensic image of a digital medium, examiners should seek justification for doing so. This is the same suggestion that was concluded from the quantitative analysis of the Imaging + Examination and Legal phases.

Along with evaluating the usefulness of each phase, this study provided predictive models for forensic examiners. These models provided digital forensic practitioners the ability to anticipate the number of evidentiary artifacts they can recover in specific phases. Furthermore, predictive models were presented to estimate the amount of time they will spend in specific forensic phases based on hard drive size.

Appendix A

Extended Results for RQs 1, 2, and 3

RQ1: Extended Results

Table 23

Regression Results for Hard Drive Size and Categories of Crime

T_1 Triage, Total Evidence

	β	*t*	*p*
Intercept	231.71	3.28	0.00
Hard Drive Size (in GB)	0.10	9.04	<0.01
COC (non-Sex Offense)	-141.49	-3.45	<0.01
R^2 0.49			

T_2 Preview, Total Evidence

	β	*t*	*p*
Intercept	245.43	3.02	0.01
Hard Drive Size (in GB)	0.35	27.00	<0.01
COC (non-Sex Offense)	-154.25	-3.27	<0.01
R^2 0.88			

T_3 Image + Exam, Total Evidence

	β	*t*	*p*
Intercept	80.33	3.77	0.00
Hard Drive Size (in GB)	<0.01	1.15	0.25
COC (non-Sex Offense)	-22.01	-1.78	0.08
R^2 0.03			

T_4 Legal, Total Evidence

	β	*t*	*p*
Intercept	8.58	2.37	0.02
Hard Drive Size (in GB)	0.01	0.96	0.34
COC (non-Sex Offense)	-1.94	-0.92	0.36
R^2 <0.01			

Table 24

Regression Results for Hard Drive Size and Categories of Crime, by Evidence Type (Multimedia, non-Multimedia)

T_1 Triage by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	228.04	3.39	0.00	Intercept	3.67	0.61	0.54
Hard Drive Size (GB)	0.10	8.83	<0.01	Hard Drive Size (GB)	0.01	7.75	<0.01
COC (non-Sex Offense)	141.26	-3.63	<0.01	COC (non-Sex Offense)	-0.23	-0.06	0.95
R^2 0.49				R^2 0.37			

T_2 Preview by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	203.07	2.71	0.01	Intercept	42.35	1.45	0.15
Hard Drive Size (GB)	0.32	27.11	<0.01	Hard Drive Size (GB)	0.03	5.65	<0.01
COC (non-Sex Offense)	-1.60	43.48	0.00	COC (non-Sex Offense)	6.04	0.36	0.72
R^2 0.89				R^2 0.23			

T_3 Image + Exam, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	71.95	6.27	<0.01	Intercept	8.38	0.60	0.55
Hard Drive Size (GB)	0.00	1.04	0.30	Hard Drive Size (GB)	0.00	0.90	0.37
COC (non-Sex Offense)	-31.81	-4.78	<0.01	COC (non-Sex Offense)	9.80	1.21	0.23
R^2 0.20				R^2 0.02			

T_4 Legal, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	5.45	3.29	0.00	Intercept	3.13	1.24	0.22
Hard Drive Size (GB)	0.00	0.48	0.63	Hard Drive Size (GB)	0.00	1.06	0.29
COC (non-Sex Offense)	-2.06	-2.14	0.03	COC (non-Sex Offense)	0.12	0.08	0.94
R^2 0.05				R^2 0.01			

Table 25

Regression Results: Hard Drive Size for Sex Offense by Total Evidence

T_1 Triage, Total Evidence

	β	*t*	*p*
Intercept	-35.21	-1.16	0.26
Hard Drive Size (in GB)	0.21	15.36	0.00
R^2 0.85			

T_2 Preview, Total Evidence

	β	*t*	*p*
Intercept	50.48	0.89	0.38
Hard Drive Size (in GB)	0.39	15.67	0.00
R^2 0.85			

T_3 Image + Exam, Total Evidence

	β	*t*	*p*
Intercept	55.55	4.10	0.00
Hard Drive Size (in GB)	0.01	1.05	0.30
R^2 >0.00			

T_4 Legal, Total Evidence

	β	*t*	*p*
Intercept	6.55	2.61	0.01
Hard Drive Size (in GB)	0.00	0.59	0.56
R^2 . 0.02			

Table 26

Regression Results: Hard Drive Size for Sex Offense by Evidence Type (Multimedia, non-Multimedia)

T_1 Triage by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	-31.10	-1.04	0.31	Intercept	-4.11	-1.87	0.07
Hard Drive Size (GB)	0.19	14.55	0.00	Hard Drive Size (GB)	0.01	14.23	0.00
R^2 0.83				R^2 0.82			

T_2 Preview by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	30.48	0.62	0.54	Intercept	18.00	2.47	0.18
Hard Drive Size (GB)	0.34	14.66	0.00	Hard Drive Size (GB)	0.05	16.08	0.00
R^2 0.83				R^2 0.86			

T_3 Image + Exam, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	38.12	4.21	0.00	Intercept	17.43	3.55	0.00
Hard Drive Size (GB)	0.00	0.90	0.37	Hard Drive Size (GB)	0.00	1.23	0.22
R^2 -0.01				R^2 0.01			

T_4 Legal, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	3.37	2.67	0.01	Intercept	3.18	1.81	0.08
Hard Drive Size (GB)	0.00	0.26	0.79	Hard Drive Size (GB)	0.00	0.65	0.52
R^2 -0.02				R^2 -0.01			

Table 27

Regression Results: Hard Drive Size for non-Sex Offense by Total Evidence

T_1 Triage, Total Evidence

	β	*t*	*p*
Intercept	16.60	4.36	0.00
Hard Drive Size (in GB)	0.01	2.16	0.03
R^2 0.05			

T_2 Preview, Total Evidence

	β	*t*	*p*
Intercept	-41.03	-1.96	0.05
Hard Drive Size (in GB)	0.32	26.97	0.00
R^2 0.92			

T_3 Image + Exam, Total Evidence

	β	*t*	*p*
Intercept	37.82	5.33	0.00
Hard Drive Size (in GB)	<0.01	0.44	0.66
R^2 - 0.01			

T_4 Legal, Total Evidence

	β	*t*	*p*
Intercept	4.76	4.55	0.00
Hard Drive Size (in GB)	0.00	0.81	0.42
R^2 .0.01			

Table 28

Regression Results: Hard Drive Size for non-Sex Offense by Evidence Type (Multimedia, non-Multimedia)

T_1 Triage by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	9.30	3.85	0.00	Intercept	7.30	3.51	0.00
Hard Drive Size (GB)	0.00	2.24	0.03	Hard Drive Size (GB)	0.00	1.36	0.18
R^2 0.06				R^2 0.01			

T_2 Preview by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	-111.91	-5.36	0.00	Intercept	70.88	5.79	0.00
Hard Drive Size (GB)	0.32	26.80	0.00	Hard Drive Size (GB)	0.00	0.40	0.69
R^2 0.92				R^2 -0.01			

T_3 Image + Exam, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	9.44	5.11	0.00	Intercept	28.39	4.34	0.00
Hard Drive Size (GB)	0.00	0.31	0.76	Hard Drive Size (GB)	0.00	0.39	0.70
R^2 -0.01				R^2 -0.01			

T_4 Legal, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	1.34	3.81	0.00	Intercept	3.42	4.50	0.00
Hard Drive Size (GB)	0.00	0.54	0.59	Hard Drive Size (GB)	0.00	0.87	0.39
R^2 -0.01				R^2 <-0.00			

Table 29

Regression Results for Hard Drive Size by Total Evidence

T_1 Triage, Total Evidence

	β	*t*	*p*
Intercept	0.36	0.02	0.99
Hard Drive Size (in GB)	0.11	9.23	0.00
R^2 0.44			

T_2 Preview, Total Evidence

	β	*t*	*p*
Intercept	-6.77	-0.25	0.80
Hard Drive Size (in GB)	0.36	26.65	0.00
R^2 0.87			

T_3 Image + Exam, Total Evidence

	β	*t*	*p*
Intercept	44.35	6.50	0.00
Hard Drive Size (in GB)	0.01	1.42	0.16
R^2 0.01			

T_4 Legal, Total Evidence

	β	*t*	*p*
Intercept	5.41	4.72	0.00
Hard Drive Size (in GB)	>0.00	1.12	0.27
R^2 >0.00			

Table 30

Regression Results for Hard Drive Size by Evidence Type (Multimedia, non-Multimedia)

T_1 Triage by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	-2.93	-0.13	0.87	Intercept	3.30	1.73	0.09
Hard Drive Size (GB)	0.10	9.00	0.00	Hard Drive Size (GB)	0.01	7.89	0.00
R^2 0.43				R^2 0.36			

T_2 Preview by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	-59.00	-2.35	0.02	Intercept	52.23	5.67	0.00
Hard Drive Size (GB)	0.34	26.49	0.00	Hard Drive Size (GB)	0.03	5.68	0.00
R^2 0.87				R^2 0.23			

T_3 Image + Exam, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	19.95	5.01	0.00	Intercept	24.40	5.49	0.00
Hard Drive Size (GB)	>0.00	1.63	0.11	Hard Drive Size (GB)	>0.00	0.72	0.47
R^2 0.02				R^2 <-0.00			

T_4 Legal, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	2.08	3.90	0.00	Intercept	3.33	4.13	0.00
Hard Drive Size (GB)	0.00	.80	0.43	Hard Drive Size (GB)	0.00	1.06	0.29
R^2 <-0.00				R^2 >0.00			

RQ2: Extended Results

Table 31

Extended Regression Results of Time (hrs.) on Hard Drive Size According to the Category of Crime

6.1 Triage			
	β	*t*	*p*
Intercept	5.17	3.44	0.00
Hard Drive Size (in GB)	0.002	11.26	<0.01
COC (not Sex Offense)	-1.84	-2.11	0.04
R^2 0.56			
6.2 Preview			
	β	*t*	*p*
Intercept	-2.08	-1.13	0.26
Hard Drive Size (in GB)	0.02	65.23	<0.01
COC (not Sex Offense)	2.24	1.07	0.04
R^2 0.10			
6.3 Imaging			
	β	*t*	*p*
Intercept	14.54	5.83	<0.01
Hard Drive Size (in GB)	0.00	2.18	0.04
COC (not Sex Offense)	-4.6-	-3.17	0.02
R^2 0.13			
6.4 Examination			
	β	*t*	*P*
Intercept	38.62	6.68	<0.01
Hard Drive Size (in GB)	0.00	1.58	0.11
COC (not Sex Offense)	-9.99	-2.98	<0.01
R^2 0.09			
6.5 Legal			
	β	*t*	*p*
Intercept	8.81	2.60	0.01
Hard Drive Size (in GB)	0.00	0.23	0.82
COC (not Sex Offense)	-1.15	0.59	0.56
R^2 0.02			

Table 32

Regression of Time (hrs.) on Hard Drive Size in Sex Offense Cases

T_1 Triage Time

	β	*t*	*p*
Intercept	0.53	1.53	0.14
Hard Drive Size (in GB)	0.01	32.92	0.00
R^2 0.96			

T_2 Preview Time

	β	*t*	*p*
Intercept	1.13	0.84	0.40
Hard Drive Size (in GB)	0.02	31.80	0.00
R^2 0.96			

T_3 Image + Exam Time

	β	*t*	*p*
Intercept	36.78	6.55	0.00
Hard Drive Size (in GB)	0.00	1.55	0.13
R^2 0.03			

T_4 Legal, Total Evidence

	β	*t*	*p*
Intercept	7.50	3.11	0.00
Hard Drive Size (in GB)	0.00	0.24	0.81
R^2 .0.02			

Table 34

Regression of Time (hrs.) on Hard Drive Size in non-Sex Offense Cases

T_1 Triage Time

	β	*t*	*p*
Intercept	3.02	11.01	0.00
Hard Drive Size (in GB)	0.00	3.40	0.00
R^2 0.14			

T_2 Preview Time

	β	*t*	*p*
Intercept	1.89	4.71	0.00
Hard Drive Size (in GB)	0.02	87.86	0.00
R^2 0.99			

T_3 Image + Exam, Total Evidence

	β	*t*	*p*
Intercept	24.97	10.64	0.00
Hard Drive Size (in GB)	0.00	0.70	0.49
R^2 -0.01			

T_4 Legal, Total Evidence

	β	*t*	*p*
Intercept	6.60	7.26	0.00
Hard Drive Size (in GB)	0.00	-0.1	0.99
R^2 -0.02			

RQ3: Extended Results

Table 35

Extended Regression Results of Multimedia and non-Multimedia Evidence Count on the Time Spent in Sex Offense Cases

T_1 Triage by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	-41.92	-1.26	0.22	Intercept	-5.80	-3.13	0.00
Time (Hr.)	36.74	13.14	0.00	Time (Hr.)	2.76	17.76	0.00
R^2 0.80				R^2 0.88			

T_2 Preview by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	33.39	0.55	0.58	Intercept	18.78	2.11	0.04
Time (Hr.)	17.19	12.50	0.00	Time (Hr.)	2.59	12.74	0.00
R^2 0.78				R^2 0.79			

T_3 Image + Exam, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	-0.49	-0.05	0.96	Intercept	-5.21	-1.08	0.29
Time (Hr.)	1.04	5.61	0.00	Time (Hr.)	0.62	6.75	0.00
R^2 0.41				R^2 0.51			

T_4 Legal, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	2.33	1.97	0.06	Intercept	-1.29	-1.67	0.10
Time (Hr.)	0.16	2.02	0.05	Time (Hr.)	0.65	12.91	0.00
R^2 0.07				R^2 0.79			

Table 36

Extended Regression Results of Multimedia and non-Multimedia Evidence Count on the Time Spent in non-Sex Offense Cases

T_1 Triage by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	-3.32	-0.91	0.37	Intercept	-8.75	-3.49	0.00
Time (Hr.)	4.36	4.80	0.00	Time (Hr.)	5.07	8.13	0.00
R^2 0.26				R^2 0.50			

T_2 Preview by Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	-144.31	-818	0.00	Intercept	71.16	5.69	0.00
Time (Hr.)	16.03	32.83	0.00	Time (Hr.)	0.10	0.30	0.77
R^2 0.94				R^2 -0.01			

T_3 Image + Exam, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	-1.04	-0.40	0.70	Intercept	-0.22	-0.02	0.98
Time (Hr.)	0.42	5.00	0.00	Time (Hr.)	1.16	3.63	0.00
R^2 0.27				R^2 0.16			

T_4 Legal, Evidence Type

Multimedia	*β*	*t*	*p*	Non-Multimedia	*β*	*t*	*p*
Intercept	0.00	0.00	1.00	Intercept	-0.06	-0.08	0.94
Time (Hr.)	0.22	5.28	0.00	Time (Hr.)	0.57	7.23	0.00
R^2 0.30				R^2 0.45			

References

Ademu, I. O., Imafidon, C. O., & Preston, D. S. (2011). A new approach of digital forensic model for digital forensic investigation. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *2*(12), 175-178.

Agarwal, A., Gupta, M., Gupta, S., & Gupta, S. C. (2011). Systematic digital forensic investigation model. *International Journal of Computer Science and Security (IJCSS)*, *5*(1), 118-131.

Alrumaithi, A. M. (2018). *Prioritisation in digital forensics: A case study of Abu Dhabi Police* (Doctoral dissertation, Liverpool John Moores University).

Alshebel, A. K. S. (2020). *Standardization Requirements for Digital Forensic Laboratories: A Document Analysis and Guideline* (Doctoral dissertation, Auckland University of Technology).

Antolos, D., Liu, D., Ludu, A., & Vincenzi, D. (2013, July). Burglary crime analysis using logistic regression. In *International Conference on Human Interface and the Management of Information* (pp. 549-558). Springer, Berlin, Heidelberg.

Baier, H., & Knauer, J. (2014, May). AFAUC--Anti-forensics of storage devices by alternative use of communication channels. In *2014 Eighth International Conference on IT Security Incident Management & IT Forensics* (pp. 14-26). IEEE.

Bem, D., & Huebner, E. (2007). Computer forensic analysis in a virtual environment. *International journal of digital evidence*, *6*(2), 1-13.

Brown, C. L. T. (2010). *Computer evidence: Collection and preservation* (2nd ed.). Boston, MA: Course Technology.

Cantrell, G., & Dampier, D. (2013). Evaluation of the semi-automated crime-specific digital triage process model. In G. Peterson & S. Shenoi (Eds.), *International Federation for Information Processing: Vol 410. International Conference on Advances in Digital Forensics*, *IX*, *83-98*, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41148-9_6

Carbone, F. (2014). *Computer forensics with FTK*. Packt Publishing Ltd.

Carlton, G. H. (2007). A grounded theory approach to identifying and measuring forensic data acquisition tasks. *Journal of Digital Forensics, Security and Law*, *2*(1), 2.

Casey, E., Ferraro, M., & Nguyen, L. (2009). Investigation delayed is justice denied: proposals for expediting forensic examinations of digital evidence. *Journal of Forensic Sciences*, *54*(6), 1353-1364.

Casey, E. (2011). *Digital evidence and computer crime: Forensic science, computers, and the internet* (3rd ed.). Academic press.

Casey, E. (2013). Triage in digital forensics. *Digital Investigation*, *2*(10), 85-86.

Casey, E. (2020). Standardization of forming and expressing preliminary evaluative opinions on digital evidence. *Forensic Science International: Digital Investigation*, *32*, 200888.

Cellebrite. (2022). *Universal Forensic Extraction Device (UFED) User Manual*. Cellebrite. https://www.cellebrite.com/en/support/.

Ciardhuáin, S. Ó. (2004). An extended model of cybercrime investigations. *International Journal of Digital Evidence*, *3*(1), 1-22.

Cusack, B. (2019). Extracting Benefits from Standardization of Digital Forensic Practices. *Policing: A Journal of Policy and Practice*. Oxford Academic.

Dancer, F. C. T., & Skelton, G. W. (2013, November). To change or not to change: That is the question. In *2013 IEEE International Conference on Technologies for Homeland Security (HST*), 212-216, Waltham, MA. https://doi.org/10.1109/THS.2013.6698957

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 113 S. Ct. 2786, 125 L. Ed. 2d 469 (1993).

Daniel, L., & Daniel, L. (2012). Digital forensics for legal professionals. *Syngress Book Co*, *1*, 287-293.

Delija, D. (2017). Remote digital forensics practices. *International Journal of Digital Technology & Economy*, *2*(1), 27-36.

Dilijonaite, A. (2017). Digital forensic readiness. *Digital Forensics*, 117-145.

Dolliver, D. S., Collins, C., & Sams, B. (2017). Hybrid approaches to digital forensic investigations: A comparative analysis in an institutional context. *Digital Investigation*, *23*, 124-137.

Dumont, C., Johnson, A., Castro, J., Leonard, A., Palmer, J., & Craig, T. (2016). Forensic Tool Comparison. *Leahy Center for Digital Investigation (LCDI).*

Du, X., Le-Khac, N. A., & Scanlon, M. (2017). Evaluation of digital forensic process models with respect to digital forensics as a service. *arXiv preprint arXiv:1708.01730*.

Du, X., & Scanlon, M. (2019, August). Methodology for the automated metadata-based classification of incriminating digital forensic artefacts. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*, *1-8,* Canterbury, United Kingdom. https://doi.org/10.1145/3339252.3340517

Dzemydiene, D., & Rudzkiene, V. (2002). Multiple regression analysis in crime pattern warehouse for decision support. In A. Hameurlain, R. Cicchetti, & R.Traunmüller (Eds.), *International Conference on Database and Expert Systems Applications*: *Vol 2453. 13th International Conference Proceedings / DEXA 2002* (pp. 249-258). Berlin, Germany: Springer-Verlag. https://doi.org/10.1007/3-540-46146-9_25

Englbrecht, L., Meier, S., & Pernul, G. (2020). Towards a capability maturity model for digital forensic readiness. *Wireless Networks*, *26*(7), 4895-4907.

Federal Rules of Evidence (2010). 28 US Code § 401 *Et seq*. https://www.uscourts.gov/sites/default/files/evidence-rules-procedure-dec2017_0.pdf

Federal Rules of Evidence (2017). 28 US Code § 902 *Et seq*. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjk7qHPuYjvAhUWHs0KHaUdDl8QFjAAegQIAhAD&url=https%3A%2F%2Fwww.uscourts.gov%2Fsites%2Fdefault%2Ffiles%2Fevidence-rules-procedure-dec2017_0.pdf&usg=AOvVaw35i05o8TBJAp6Mlcui1mJX

Garfinkel, S. L. (2010). Digital forensics research: The next 10 years. *digital investigation*, *7*, S64-S73.

Ghazinour, K., Vakharia, D. M., Kannaji, K. C., & Satyakumar, R. (2017, September). A study on digital forensic tools. In *2017 IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI)* (pp. 3136-3142). IEEE.

Goodison, S. E., Davis, R. C., & Jackson, B. A. (2015). Digital evidence and the US criminal justice system. Identifying Technology and Other Needs to More Effectively Acquire and Utilize Digital Evidence. *Priority Criminal Justice Needs Initiative. Rand Corporation*.

Grobler, C. P., Louwrens, C. P., & von Solms, S. H. (2010, February). A framework to guide the implementation of proactive digital forensics in organisations. In *2010 International conference on availability, reliability and security* (pp. 677-682). IEEE.

Hamilton, F. (2020, April 23). *Police struggling to clear evidence backlog of 12,000 devices.* The Times. https://www.thetimes.co.uk/article/police-struggling-to-clear-evidence-backlog-of-12-000-devices-rpmhmfnpt

Hemdan, E. E. D., & Manjaiah, D. H. (2021). An efficient digital forensic model for cybercrimes investigation in cloud computing. *Multimedia Tools and Applications*, 1-28.

Hitchcock, B., Le-Khac, N. A., & Scanlon, M. (2016). Tiered forensic methodology model for digital field triage by non-digital evidence specialists. *Digital Investigation*, *16*, S75-S85. https://doi.org/10.1016/j.diin.2016.01.010

Hong, I., Yu, H., Lee, S., & Lee, K. (2013). A new triage model conforming to the needs of selective search and seizure of electronic evidence. *Digital Investigation*, *10*(2), 175-192.

Horsman, G., Laing, C., & Vickers, P. (2014). A case-based reasoning method for locating evidence during digital forensic device triage. *Decision Support Systems*, *61*, 69-78.

Horsman, G. (2017). Can we continue to effectively police digital crime?. *Science & Justice*, *57*(6), 448-454.

Horsman, G. (2020). Opinion: Does the field of digital forensics have a consistency problem?. *Forensic Science International: Digital Investigation*, 300970. https://doi.org/10.1016/j.fsidi.2020.300970

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hungwe, T., VENTER, H. S., & Kebande, V. R. (2019, May). Scenario-based digital forensic investigation of compromised MySQL database. In *2019 IST-Africa Week Conference (IST-Africa)* (pp. 1-11). IEEE.

Internet Crime Complaint Center (2021). *Internet crime report.* (2021 IC3 annual report). https://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3Report.pdf

ISO/IEC 27043. (2015). *Information technology–Security techniques–Incident investigation principles and processes.* The British Standards Institution.

Jansen, W., Delaitre, A., & Moenner, L. (2008, January). Overcoming impediments to cell phone forensics. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)* (pp. 483-483). IEEE.

Jiang, J. G., Yang, B., Lin, S., Zhang, M. X., & Liu, K. Y. (2015). A practical approach for digital forensic triage. In *Applied Mechanics and Materials* (Vol. 742, pp. 437-444). Trans Tech Publications Ltd.

Jusas, V., Birvinskas, D., & Gahramanov, E. (2017). Methods and tools of digital triage in forensic context: Survey and future directions. *Symmetry*, *9*(4), 49.

Kalker, T., Haitsma, J., & Oostveen, J. C. (2001). Issues with digital watermarking and perceptual hashing. *Multimedia Systems and Applications IV, 4518*, 189-198. https://doi.org/10.1117/12.448203

Karresand, M., & Shahmehri, N. (2006, June). File type identification of data fragments by their binary structure. In *Proceedings of the IEEE Information Assurance Workshop* (pp. 140-147).

Karie, N. M., & Venter, H. S. (2015). Taxonomy of challenges for digital forensics. *Journal of forensic sciences*, *60*(4), 885-893.

Kebande, V. R., & Venter, H. S. (2018). Novel digital forensic readiness technique in the cloud environment. *Australian Journal of Forensic Sciences*, *50*(5), 552-591.

Kent, K., Chevalier, S., Grance, T., & Dang, H. (2006). Guide to integrating forensic techniques into incident response. *NIST Special Publication*, *10*(14), 800-86.

Kessler, G. C., & Carlton, G. H. (2014). A study of forensic imaging in the absence of write-blockers. *Journal of Digital Forensics, Security and Law*, *9*(3), 51.

Leigland, R., & Krings, A. W. (2004). A formalization of digital forensics. *International Journal of Digital Evidence*, *3*(2), 1-32.

Lillis, D., Becker, B., O'Sullivan, T., & Scanlon, M. (2016). Current challenges and future research areas for digital forensic investigation. *arXiv preprint arXiv:1604.03850*.

Liles, S., Rogers, M., & Hoebich, M. (2009, January). A survey of the legal issues facing digital forensic experts. In *IFIP International Conference on Digital Forensics* (pp. 267-276). Springer, Berlin, Heidelberg.

Losavio, M., Pastukov, P., & Polyakova, S. (2015). Cyber black box/event data recorder: Legal and ethical perspectives and challenges with digital forensics. *Journal of Digital Forensics, Security and Law*, *10*(4), 4.

Makura, S., Venter, H. S., Kebande, V. R., Karie, N. M., Ikuesan, R. A., & Alawadi, S. (2021). Digital forensic readiness in operational cloud leveraging ISO/IEC 27043 guidelines on security monitoring. *Security and Privacy*, e149.

Martini, B., & Choo, K. K. R. (2012). An integrated conceptual digital forensic framework for cloud computing. *Digital Investigation*, *9*(2), 71-80.

McKemmish, R. (1999). What is forensic computing? *Australian Institute of Criminology, trends and issues in crime and criminal justice,* 118, 1-6.

McNicholas III, J. B. (2020). Digital Forensic Readiness: An Examination of Law Enforcement Agencies in the State of Maryland. *Dakota State University.* Beadle Scholar.

Merriam-Webster. (2020). Evidence. In *Merriam-Webster.com dictionary*. Retrieved October 10, 2020, from https://www.merriam-webster.com/dictionary/Evidence

Meyers, C., Ikuesan, A. R., & Venter, H. S. (2017, November). Automated RAM analysis mechanism for windows operating system for digital investigation. In *2017 IEEE Conference on Application, Information and Network Security (AINS)* (pp. 85-90). IEEE.

Montasari, R. (2016). Formal two stage triage process model (ftstpm) for digital forensic practice. *Int. J. Comput. Sci. Secur*, *10*, 69-87.

Moser, A., & Cohen, M. I. (2013). Hunting in the enterprise: Forensic triage and incident response. *Digital Investigation*, *10*(2), 89-98.

Mothi, D., Janicke, H., & Wagner, I. (2020). A novel principle to validate digital forensic models. *Forensic Science International: Digital Investigation*, *33*, 200904.

MSAB. (2022). *Getting the best digital evidence is what matters*. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi12brSy77sAhXOVs0KHUlbB44QFjACegQIBhAC&url=https%3A%2F%2Fwww.msab.com%2Fdownload%2Fproduct_sheets%2Fen%2FXRY_-_Product_Family_EN.pdf&usg=AOvVaw3UFQNAoBO-TFBUlr7IR9Wb

Mueller, C. B., Kirkpatrick, L. C., & Richter, L. L. (2020). *Federal rules of evidence: With advisory committee notes and legislative history and cases*. Wolters Kluwer. NY.

Munkhondya, H., Ikuesan, A., & Venter, H. (2019, February). Digital forensic readiness approach for potential evidence preservation in software-defined networks. In *ICCWS 2019 14th International Conference on Cyber Warfare and Security: ICCWS* (Vol. 268).

Nguyen, A. (2020, November 23). Big Sur, big changes. *Sumuri*. https://sumuri.com/big-sur-big-changes/

Nisbet, A., & Jacob, R. (2019, August). TRIM, wear levelling and garbage collection on solid state drives: a prediction model for forensic investigators. In *2019 18th IEEE International Conference on Trust, Security and Privacy In Computing and Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (pp. 419-426). IEEE.

National Institute of Standards and Technology (NIST). (2001, November 7). *General Test Methodology for Computer Forensics Tools.* v1.9. U.S. Department of Commerce. http://www.cftt.nist.gov/Test

Nouh, M., Nurse, J. R., Webb, H., & Goldsmith, M. (2019). Cybercrime investigators are users too! Understanding the socio-technical challenges faced by law enforcement. *arXiv preprint arXiv:1902.06961*.

Ölvecký, M., & Gabriška, D. (2018, September). Wiping techniques and anti-forensics methods. In *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)* (pp. 000127-000132). IEEE.

Palmer, G. (2001, August). A road map for digital forensic research. In *First digital forensic research workshop (DFRWS)*, *27-30,* Utica, New York.

Park, S., Kim, Y., Park, G., Na, O., & Chang, H. (2018). Research on digital forensic readiness design in a cloud computing-based smart work environment. *Sustainability*, *10*(4), 1203.

Pasquale, L., Alrajeh, D., Peersman, C., Tun, T., Nuseibeh, B., & Rashid, A. (2018, May). Towards forensic-ready software systems. In *2018 IEEE/ACM 40th International Conference on Software Engineering: New Ideas and Emerging Technologies Results (ICSE-NIER)* (pp. 9-12). IEEE.

Perumal, S. (2009). Digital forensic model based on Malaysian investigation process. *International Journal of Computer Science and Network Security*, *9*(8), 38-44.

Presley, S. S., Landry, J. P., & Black, M. (2018). Using project management knowledge and practice to address digital forensic investigation challenges. *2018 KSU conference on cybersecurity education, research and practice.* (PP. 1-25).

Quick, D., & Choo, K. K. R. (2014). Impacts of increasing volume of digital forensic data: A survey and future research challenges. *Digital Investigation*, *11*(4), 273-294. https://doi.org/10.1016/j.diin.2014.09.002

Quick, D., & Choo, K. K. R. (2017). Big forensic data management in heterogeneous distributed systems: Quick analysis of multimedia forensic data. *Software: Practice and Experience, 47*(8), 1095-1109. https://doi.org/10.1002/spe.2429

Quick, D., & Choo, K. K. R. (2018a). Digital forensic data reduction by selective imaging. In L. Chen, K. K. R. Choo, S. Chow, R. Deng (Eds.), *Springer briefs on cyber security systems and networks. Big Digital Forensic Data, 1* (pp. 69-92). https://doi.org/10.1007/978-981-10-7763-0_4

Quick, D., & Choo, K. K. R. (2018b). Digital forensic intelligence: Data subsets and Open-Source Intelligence (DFINT+ OSINT): A timely and cohesive mix. *Future Generation Computer Systems*, *78*, 558-567.

Raychaudhuri, K. (2019). A comparative study of analysis and extraction of digital forensic evidences from exhibits using disk forensic tools. *International Journal of Cyber-Security and Digital Forensics*, *8*(3), 194-206.

Reddy, N. (2019). Introduction to cyber forensics. *Practical Cyber Forensics* (pp. 1-28). https://doi.org/10.1007/978-1-4842-4460-9

Ries, D. G., & Hill, C. H. (2017). Digital Forensics in the Courts.

Robins-Kaplan (2019). *What's Happening? The Impact of FRE 902(14) on eDiscovery.* Real Talk. https://www.robinskaplan.com/resources/legal-updates/real-talk-the-robins-kaplan-business-law-update/2019/real-talk-the-robins-kaplan-business-law-update-winter-2019/whats-happening-the-impact-of-fre-902-14-on-ediscovery

Rogers, M. K., Goldman, J., Mislan, R., Wedge, T., & Debrota, S. (2006). Computer forensics field triage process model. *Journal of Digital Forensics, Security and Law, 1*(2), 18-38. Retrieved from https://commons.erau.edu/jdfsl/vol1/iss2/2/

Roussev, V., Quates, C., & Martell, R. (2013). Real-time digital forensics and triage. *Digital Investigation*, *10*(2), 158-167.

Rowlingson, R. (2004). A ten-step process for forensic readiness. *International Journal of Digital Evidence*, *2*(3), 1-28.

Scanlon, M. (2016, August). Battling the digital forensic backlog through data deduplication. In *2016 Sixth International Conference on Innovative Computing Technology (INTECH), 10-14,* IEEE. Dublin, Ireland. https://doi.org/10.1109/INTECH.2016.7845139

Scrivens, N., & Lin, X. (2017, May). Android digital forensics: Data, extraction and analysis. In *Proceedings of the ACM Turing 50th Celebration Conference-China* (pp. 1-10). https://doi.org/10.1145/3063955.3063981

Shaw, A., & Browne, A. (2013). A practical and robust approach to coping with large volumes of data submitted for digital forensic examination. *Digital Investigation*, *10*(2), 116-128.

State v. Newman, 838 N.W.2d 317, 21 Neb. App. 29 (Ct. App. 2013).

Sudyana, D., Prayudi, Y., & Sugiantoro, B. (2019). Analysis and evaluation digital forensic investigation framework using ISO 27037: 2012. *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, *8*(1), 1-14.

Sunde, N., & Dror, I. E. (2019). Cognitive and human factors in digital forensics: problems, challenges, and the way forward. *Digital investigation*, *29*, 101-108.

Taha, K., & Yoo, P. D. (2018). A forensic system for identifying the suspects of a crime with no solid material evidences*. Proceedings of 2018 IEEE, 16th Intl Conf. on Dependable, Autonomic and Secure Computing, 16th Intl Conf. on Pervasive Intelligence and Computing, 4th Intl Conf. on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), 576-583,* Athens, Greece. https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00107

Thompson, T. (2019, July 16). *Forensic delays 'deeply concerning' as case backlog grows*. Police Professional. https://www.policeprofessional.com/news/forensic-delays-deeply-concerning-as-case-backlog-grows/

Thornton, G., & Harper, R. H. (1991). *Detectives Or Clerks?: An Examination of the Work of Detectives*. *Cambridge Laboratory,* EPC-1991-109.

US v. Stabile, 633 F.3d 219 (3d Cir. 2011).

US v. Tello, No. 17-cr-20333 (E.D. Mich. Oct. 4, 2017).

Veber, J., & Klíma, T. (2014). Influence of standards ISO 27000 family on digital evidence analysis. *Proceedings of the 22nd Interdisciplinary Information Management Talks*, 103-114.

Veendrick, H. (2018). *Bits on Chips*. Springer.

Vincze, E. A. (2016). Challenges in digital forensics. *Police Practice and Research*, *17*(2), 183-194.

Vrubel, A. (2011). Creation and maintenance of MD5 hash libraries, and their application in cases of child pornography. *The Sixth International Conference on Forensic Computer Science (ICoFCS)*, 137-141. https://doi.org/10.5769/C2011015

Wheelan, C. (2013). *Naked statistics: Stripping the dread from the data*. WW Norton & Company.

Yang, B., Li, N., & Jiang, J. (2016). A new triage process model for digital investigations. *Paper presented at Advanced Information Management, Communicates, Electronic and Automation Control Conference,* 712-717. https://doi.org/10.1109/IMCEC.2016.7867302