UNDERSTANDING CHANGES IN INDIVIDUAL AND FIRM BEHAVIOR IN RESPONSE TO SECURITY AND PRIVACY FACTORS

by

Joseph Buckman

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MANAGEMENT INFORMATION SYSTEMS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF MANAGEMENT
WITH A MAJOR IN MANAGEMENT INFORMATION SYSTEMS

In the Graduate College

THE UNIVERSITY OF ARIZONA

2018

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by ***Joseph Buckman***, titled ***Understanding Changes in Individual and Firm Behavior in Response to Security and Privacy Factors*** and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Matthew Hashim Date: (***07/18/2018***)

Jesse Bockstedt Date: (***07/18/2018***)

Tiemen Woutersen Date: (***07/18/2018***)

Laura Brandimarte Date: (***07/18/2018***)

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: ***Matthew Hashim*** Date: (***07/18/2018***)

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Joseph Buckman

ACKNOWLEDGEMENTS

Throughout this journey I have had ups and downs but the supporting cast around me has helped me through it all. The two most important people in my world are my wife Roxanne and son Perrin. They have been with me every step of the way and on into the future. Thank you for sticking with me through thick and thin.

I want to thank my two advisors Jesse Bockstedt and Matthew Hashim. They took a chance with me as their first student even though they barely knew me, and I was at my lowest point. Over the years they have become close friends and I owe my career to them. They offered the guidance, time, and skills that have allowed me to grow as both a researcher and a person.

The final person I want to thank is Tiemen Woutersen. He has been a great mentor and friend that has pushed and helped me to excel beyond what I thought I could do.

DEDICATION

To my dad who wanted me to go after my vision and give it my all. I love and miss you more than words can express.

TABLE OF CONTENTS

List of Tables…….. 8
List of Figures……. 9
Abstract……... 10
Chapter 1: Introduction……... 12
Chapter 2: Essay 1 – Relative Privacy Valuations under Varying Disclosure Dimensions... 21
2.1 Research Model and Hypotheses…….. 23
2.2 Experimental Procedures, Analyses, and Results…… 27
2.2.1 Experimental Procedures Common to each Study……..... 27
2.2.2 Study 1: Information Disclosure Factors and WTA…….. 31
2.2.2.1 Study 1 Data Analysis and Results……... 32
2.2.3 Study 2: Increasing the Salience of Treatment Factors……. 34
2.2.3.1 Study 2 Data Analysis and Results……... 35
2.2.4 Study 3: Highlighting the Consequences of Private Information Disclosure… 36
2.2.4.1 Study 3 Data Analysis and Results……... 37
2.2.5 Post Hoc Analyses and Comparisons across Studies…… 38
2.3 Discussion……. 41
Chapter 3: Essay 2 – Fool Me Twice: An Analysis of Repeat Data Breaches within Firms. 46
3.1 Characteristics of Data Breach Notification Laws…… 48
3.2 Predicting Future Breaches……... 55
3.3 Model Development…….. 56
3.3.1 Hazard Model…… 56
3.3.2 Multinomial Logit……. 59
3.4 Data……... 60
3.4.1 Privacy Rights Clearinghouse……... 60
3.4.2 Identity Theft Laws……... 64
3.4.3 Data Breach Notification Policies……. 65
3.5 Data Analysis and Results…… 67
3.5.1 What impacts the survival of a firm that experiences repeat breaches?..... 68
3.5.1.1 Control Variables…….. 68
3.5.1.2 Hypothesized Variables…… 69
3.5.2 Are particular industries more susceptible to certain types of breaches?..... 71
3.5.3 Does the breach type affect the risk of experiencing a particular repeat breach?..... 72
3.6 Discussion and Implications……. 73
Chapter 4: Essay 3 – Impact of Data Breaches on the Benefits of Healthcare Information Technology……... 80
4.1 Research Model and Hypotheses…….. 82
4.1.1 Hospital Data Breach…...…….. 84
4.1.2 Process of Care...……... 85
4.2 Research Methodology and Analysis…… 86
4.3 Healthcare Data…...……... 88
4.4 Results……..…… 92
4.5 Discussion……. 96
Chapter 5: Conclusion…… 101

Appendix A: Essay 1 Supplementary Data……..... 106
Appendix B: Essay 3 Mediation Analysis…….. 129
References…….. 135

LIST OF TABLES

Table 1. Summary Statistics by Population, Study, and Treatment..…… 32
Table 2. Main Experimental Results Using Tobit Regressions…….. 33
Table 3. Pooled Analysis Using Tobit Regressions……... 38
Table 4. Logit Pricing Out of Market……. 40
Table 5. Variable Descriptions……... 66
Table 6. Summary Statistics……... 67
Table 7. Log-normal Hazard Model……... 70
Table 8. Multinomial Logit by Industry Type…… 72
Table 9. Multinomial Logit for Next Data Breach Type…… 73
Table 10. Index Variable Descriptions……... 89
Table 11. Relationships between HIT, Process, and Outcome...…… 93
Table 12. 2SLS for Fixed Effects Panel Data Model……. 95

List of Figures

Figure 1. Role of HIT on Processes and Outcomes……... 82
Figure 2. Hypothesized Role of HIT and Data Breach on Process of Care……... 86

ABSTRACT

We investigate consumer and firm responses to information privacy and security factors across three unique essays. The first essay uses three experiments across two different populations (college students and Amazon Mechanical Turk workers) to capture consumer valuations for an information disclosure. Each experiment manipulates characteristics of a required privacy disclosure by altering the information context, intended secondary use of the disclosed private information, and the requirement to disclose personally identifiable information. Across the three experiments, we consistently observe null effects for each of the privacy factors with the exception of two population dependent exceptions. Our participants do acknowledge the increased risk introduced by the experimental factors and the increased saliency and awareness from experiments two and three lead to higher privacy valuations on average. However, there is no consistent manifestation as signification main effects for the three privacy factors. The second essay analyzes firms experiencing multiple data breaches and determines which policies in data breach notification laws are effective deterrents. The results from estimating a parametric hazard model indicate that allowing the individual responsible for maliciously breaching a firm's data and requiring firms disclose breach information to a state attorney general deter firms from subsequent breaches. We also find that states that do not require breach notification when consumers are unlikely to be harmed see an increase in risk of future breach. Additionally, we investigate the relationship between industry type and breach type as well as prior breach type and subsequent breach type. Our results suggest that government agencies are more likely to have an internal breach, educational institutions are more likely to experience system hacking, and retail businesses are more susceptible to employee related breaches. The relationship between prior and subsequent breach types indicates that firms

are more likely to experience future data breaches of the same type. The third essay focuses on data breaches within hospitals by studying the effect of breaches on patient outcomes through changes in process of care. We merge several sources of data to create a unique panel data set containing information on hospitals' healthcare information technology characteristics, process of care measures, meaningful use attestation, and data breach experience. Estimating a 2SLS fixed effect panel data model provides that experiencing a data breach leads to improvements in process of care. Meaningful use attestation, on the other hand, reduces the process of care for common medical conditions. We also find that improving the process of care for medical conditions leads to better patient outcomes for those conditions. Thus, our findings demonstrate that through influencing the process of care a data breach improves patient outcomes while achieving meaningful use worsens patient outcomes. The combination of these three essays offers a unique perspective into how consumers and firms perceive information privacy and security. Ultimately, consumers and firms demonstrate that information privacy and security is not a priority unless proper incentive mechanisms and adequate information are present.

1. Introduction

Economists and industry leaders consider consumer data to be the primary asset that fuels the digital economy (Burke 2015). Consumer data enables firms to provide accurate product recommendations and customized content in order to personalize the consumer's online experience. Personalization can boost sales by an average of 20% (Soojian 2015), leading firms to enhance their data gathering capabilities and maximize the amount and quality of data obtained about consumers. A large portion of the data comes from consumers sharing their personal information knowing that a firm is collecting some form of it (Acquisti et al. 2015). Information sharing between consumers and firms has been successful despite growing concerns about information privacy because of the overwhelming trust placed on firms to protect the personal information they store (Olenski 2016). Unfortunately, the steady increase in the number of data breaches[1] each year since 2005 appears to be eroding consumer trust in firms' ability to secure their personal and private information (Help Net Security 2016).

In this thesis, we implement three essays to better understand consumer (1) privacy concerns and (2) firms' responses to information privacy threats. Specifically, we address the research question: *"How and to what extent do individual and firm behaviors change in response to information privacy factors?"* The situations we explore with consumers and firms include consumers' disclosing information to firms and organizational data breaches.

Essay 1 investigates relative changes to the value people place on the disclosure of private information given certain aspects of the disclosure. As consumers become aware of increasingly pervasive data collection practices, empirical and anecdotal evidence suggest that

[1]A data breach is an incident in which an individual name plus a Social Security number, driver's license number, medical record, or financial record is put at risk because of exposure (Identity Theft Resource Center 2016).

they place value on their private information[2]. Research has shown that the monetary value that consumers demand for their private information has increased significantly over time (Huberman et al. 2005; Danezis et al. 2005; Cvrcek et al. 2006; Acquisti et al. 2013; Staiano et al. 2014). In addition, consumers expect decision-making enhancements and personalization services in exchange for the disclosure of private information (Shah 2015). However, it is not clear that consumers understand the real market value of private information. It is also not clear that consumers understand the implications from disclosing personal information, that is, how and where their information is used, aggregated, packaged, and resold to other parties.

Many Internet companies trade services for access to personal information. In 2017, consumers and Internet pundits caused an uproar when Unroll.me, a website that helps users unsubscribe from email lists, revealed it had been selling user data to generate revenue for their free to consumer service (Feldman 2017). In 2018, the #DeleteFacebook social media campaign began in response to Cambridge Analytica acquiring personal data on approximately 87 million unknowing Facebook users (Bever 2018). The U.S. government even took notice of the incident and held congressional hearings with Facebook CEO Mark Zuckerberg. Despite all of this, recent data compiled by the strategic marketing firm Kepios indicate that very few people actually left Facebook and instead the number of monthly active users grew by approximately 4% (Zetlin 2018). These examples demonstrate that there is a misalignment between expectations of privacy and actual privacy, but they also indicate that consumers do not understand the value of their personal information for these service providers. Essay 1 focuses on this second aspect: understanding what dimensions of an information disclosure online affect a consumer's valuation for their private information.

[2] Throughout Essay 1, we use the term "private information" to refer to any information an online firm would not know unless disclosed by the consumer. Thus, private information may include personal information such as gender and race, which may not be considered private in a face-to-face exchange.

Next, Essay 2 studies changes in the likelihood of an organization experiencing subsequent data breaches given notification laws requiring firms disclose details regarding the breach incident. Data breaches have far-reaching consequences for consumers and firms alike. For example, consumers experience emotional and cognitive distress whereas firms often bear direct financial damages (Solove 2006). Some of the financial costs firms incur come from state breach notification laws[3] because they require firms notify potentially compromised consumers, provide those consumers with a minimum duration of identity theft protection, and settle fines charged by overseeing regulatory agencies (Romanosky and Acquisti 2009). Furthermore, over twenty percent of breached firms report losing a significant portion of their consumer base following public notification of a security incident (Burns 2017).

Prior data breach literature has primarily been concerned with the aftermath following a breach due to the presence of state-level breach notification laws. Several researchers have illustrated that a significant number of firms adjust their data security practices following a breach, thereby giving credence to the assumption that firms take precautions to prevent future breaches (Samuelson Law, Technology, and Public Policy Clinic 2007; Schneider 2009; Sen and Borle 2016). A prominent criminal and social science theory for such firm behavior is deterrence theory, which suggests that firms, while not acting perfectly rational, are reasonably aware of the punishments and benefits associated with legislative policies and abide by those policies only when punishment is greater than the benefit. For instance, recent high-profile cases include the 2013 Target and 2014 Home Depot data breaches in which the cumulative costs for both

[3]The first instance of a notification law in the United States was the California Civil Code Section 1798.29 introduced in 2003, which required all firms with business in California to report a privacy or security breach to the individuals affected by the breach. Following a data breach incident in 2008 at ChoicePoint, a data aggregation company that held billions of consumers' private information records, states across the U.S. used the California legislation as a model for creating their own data breach disclosure laws (Gatzlaff and McCullough 2010). As of July 2017, forty-eight of the fifty states have enacted some form of breach notification law.

companies have exceeded $550 million (Daly 2016). Target's 2013 breach resulted in stolen credit cards, damaged reputation, loss of firm value, among other consequences; leading the company to be among the first U.S. adopters of EMV Chip-and-PIN technology for payment processing, which is a strong countermeasure against future credit card data threats. However, according to Privacy Rights Clearinghouse (PRC), approximately 31% of disclosed data breaches are part of a series of multiple breaches at the same organization. We focus on these multiple breach events within an organization as it calls to question the effectiveness of data breach notification law as a deterrent to future breach. The main objectives for this essay are to better understand why firms experience future data breaches by studying the policies which compromise data breach notification laws across the U. S. and to provide guidance for preventing future data breaches.

Lastly, Essay 3 concerns data breaches in the healthcare industry. Scholars believe that health information technology (HIT)[4] is the critical factor driving improvements in U.S. healthcare quality and decreases in healthcare costs (Appari and Johnson 2010). However, anecdotal evidence suggests that the implementation of HIT and sharing health information with other collective healthcare organizations can also be disruptive to normal healthcare operations by introducing new threats to patients' information security and privacy (Miller and Tucker 2009; Appari and Johnson 2010; McCullough et al. 2010). For instance, the transition from paper records to electronic medical records (EMR) may increase patient records' vulnerability to information systems hacking as well as the magnitude of the number of records that can be compromised at one time. Evidence of such threats can be found in the steadily rising number of

[4] "An information system including all computer-based components which are used by healthcare professionals or the patients themselves in the context of inpatient or outpatient care to process patient-related data, information, or knowledge" (Pinsonneault et al. 2017). Health information technology includes the range of technologies used in healthcare organizations such as electronic medical records, health information exchange, computerized physician order entry systems, and clinical decision support systems.

reported data breaches[5] in the healthcare industry. With the exception of 2015, the number of data breaches in hospitals has grown each year since 2009. The tremendous threat to patients' privacy and information security is now considered one of the leading barriers to entry for healthcare providers who have yet to adopt EMR or implement HIT (Kruze et al. 2017).

The literature on data breaches and information security in healthcare is a pervasive area due to HIT's handling of individuals most sensitive information and the life-threatening consequences of system failure (Appari and Johnson 2010; Kim et al. 2015). The current infrastructure for the U.S. medical industry places hospital management's focus on (1) financial spending and (2) properly caring for patients. Hospital management is concerned with the amount a hospital spends on each patient arriving as well as overhead costs surrounding daily operations. Data breaches may be costly for hospitals as they can shift funding toward breach recovery and away from new medical equipment, staff, or expansion. The recovery costs associated with a data breach in the healthcare industry are expensive for both the healthcare organization as well as patients. A recent study from Ponemon Institute shows that healthcare organizations pay more than $350 per record stolen (Ponemon Institute 2017). Additionally, the organization may receive fines from the Department of Health and Human Services (DHHS) for not adhering to security guidelines.

Patient care often times relies on proper information and efficient processes. Following a breach, patients' medical information may become unavailable in the HIT system, or worse, the communication relationship between physician and patient may suffer for fear of future compromise. Furthermore, hospital staff endures process reengineering as a means to rectify the breach, which has been shown to impact staffs' ability to care for patients (Angst et al. 2012).

[5] The U.S. Department of Health and Human Services defines a data breach in a healthcare organization as an impermissible use or disclosure that compromises the security or privacy of patients' protected health information.

Much of the prior literature on hospital data breaches has focused on financial spending (e.g., information security investments) with less attention given to the impact of a data breach on patient care. In this essay, we study the latter with the objective of informing the effect of a hospital data breach on process of care and subsequently patient outcomes.

Taken together, these three essays paint an interesting picture of how consumers and firms perceive and approach information privacy and security. For instance, Essay 1 deepens the understanding of consumer privacy valuations by studying relative changes in private information valuation in a realistic, multidimensional disclosure decision. Using economic experiments, we study how the information context, the requirement to disclose personally identifying information, and the service provider's plans to sell personal information to third parties affect the value consumers place on their private information. Interestingly, with the exception of two sample-specific instances, we largely find null effects, which are in contrast to prior work that has typically looked at these dimensions independently (e.g., Culnan 1993; Berendt et al. 2005). We also find that the null effects persist even after increasing the saliency of the privacy factors in the disclosure decision and highlighting the consequences associated with disclosing information to the service provider. However, post hoc analysis and insights from a post-experiment survey suggest that some participants do acknowledge the increased risk introduced by these disclosure dimensions by pricing themselves out of the market altogether. Our findings suggest it might be an all or nothing type of decision as opposed to an activation of individual factors the prior literature suggests are important in a multi-dimension private information disclosure. The results also suggest that online disclosure decisions are evolving, especially in settings that incorporate multiple disclosure dimensions.

Essay 2 uses a parametric hazard model of future breach risk to analyze components of state-level data breach notification laws. We find that the presence of a notification law actually increases the risk of a future breach. Additionally, states that do not require breach notification when consumers are unlikely to be harmed also see an increase in risk of future breach. Policies successfully reducing the risk of future breach include allowing the individual responsible for a breach to face criminal charges if he or she exhibited malintent and requiring firms disclose breach information to a state attorney general. We also estimate a multinomial logit model of an industry's likelihood of experiencing a particular type of breach. The results of our estimation provide evidence supporting a relationship between industry and breach type, demonstrating that government agencies are more likely to be breached internally rather than externally. Educational institutions, retail, and non-retail businesses have a higher likelihood of falling victim to a technical breach such as a hacker infiltrating their information systems. Retail businesses are susceptible to both employee related non-technical breaches and technical breaches. Finally, we estimate a second multinomial logit model but look at future subsequent breach's likelihood of being related to a prior breach. The model we estimate indicates that subsequent breaches are in fact more likely to be of the same breach type as the prior breach.

The novel findings from Essay 2 provide a deeper understanding of the impact of data breach notification laws on firm-level responses to a data breach by identifying a potential source of why firms continue to experience future breaches. Specifically, the punishments associated with requiring firms to disclose the occurrence of a data breach are sufficient in generating organizational and information security change, but these changes may not be sufficiently addressing future threats. Our analyses of the relationships between industry type, prior breach

type, and future breach type further our contribution by offering several empirical and practical contributions to data breach prevention.

Essay 3 uses panel data to estimate a two stage least squares model. The first stage finds evidence that processes of care[6] for certain conditions are significantly affected by HIT implementation and experiencing a data breach. Specifically, achieving meaningful use with hospitals' HIT leads to a decrease in the process of care. Experiencing a data breach, however, leads to improvements in process of care. Estimation in the second stage reveals that improving process of care translates to a lower mortality and readmission rate for those same conditions (i.e., improved patient outcomes). We conclude that data breaches are having a positive impact on patient outcomes.

Our findings in Essay 3 provide a unique view into hospital operations. In particular, our results indicate that meaningful use attestation provides similar disruption to hospital staffs' patient care as other types of HIT implementation shown in the literature. However, we extend the prior research by demonstrating that the disruptions from achieving meaningful use have more severe consequences than simply reducing the efficiency of care provision. The results regarding the relationship between data breach experience and process of care improvements suggests that the audits and potential process changes within the hospital following a breach are extending to improved patient care. Thus, data breaches may be detrimental to patient privacy but a positive source of change for patient care.

We present the methodologies, analyses, results, and discussion for each essay in the following sections. Section 2 provides the completed work and findings for Essay 1. Section 3

[6] We define process of care as the combined sequence of events that hospital staff follow to care for a patient. For example, nurses and doctors follow an established set of procedures for patients experiencing a heart attack that are consistent across all hospitals.

explains the story and results for Essay 2. Section 4 illustrates Essay 3 and presents its developments. Finally, Section 5 concludes the thesis.

2. Essay 1 - Relative Privacy Valuations under Varying Disclosure Dimensions

Existing economics of privacy research shows that individuals act strategically within an information market of buyers and sellers (Posner 1981), only disclosing information after considering privacy-related consequences. Laufer and Wolfe (1977) defined this situation as a privacy calculus of costs and benefits associated with every information disclosure decision. Later, the privacy calculus was extended to show that individuals consider their general concerns, prior experiences, Internet trust, and personal Internet interest, before making a disclosure decision (Culnan and Armstrong 1999; Dinev and Hart 2006).

Regarding benefits in the privacy calculus, empirical studies have identified monetary rewards to be an effective means of obtaining information in privacy decisions (Phelps et al. 2000; Caudill and Murphy 2000; Hui et al. 2007; Xu et al. 2010; Preibusch 2015). Auction mechanisms (e.g., second-price) have frequently been used to establish point estimates of the monetary value individuals require to disclose single pieces of private information, such as age, weight, or location (Danezis et al. 2005; Huberman et al. 2005; Cvrcek et al. 2006; Staiano et al. 2014). Their observed values varied widely, suggesting subjectivity and difficulty establishing a single, generalized monetary value for an average person's private information. Further, an auction mechanism may introduce unnatural competition among study participants for selling their private information.

Prior research has also argued that the point estimates consumers place on their private information may be misguided due to inherent instability (Acquisti et al. 2015) and uncertainty regarding consequences (Acquisti et al. 2013). Klopfer and Rubenstein (1977) proposed that the instability of valuations stems from the subjective nature of the interpretation of privacy and the reward for revealing information. An individual's subjective interpretation may also be flawed

due to incomplete information about the disclosure (Acquisti and Grossklags 2005). That is, individuals may have little knowledge about the information an organization has already captured and organizations do not always explicitly state their intentions or planned usage for gathered information. Thus, consumers struggle to determine what types and how much information should be disclosed (Acquisti et al. 2015). Even when consumers receive full details in the disclosure decision, they continue to struggle with optimal decision-making due to bounded rationality (Acquisti and Grossklags 2005). In other words, consumers possess cognitive limitations that hinder their ability to acquire and process the dimensions of a disclosure decision (Acquisti 2004).

In addition, empirical studies found that non-normative factors influence privacy disclosure decisions. Tsai et al. (2011) implemented a lab experiment using a search engine to find products from multiple websites. Search results included the price of the product and a rating for each website's privacy policy, among other information. The authors found that consumers pay premium prices for products from websites with greater privacy protection when differences in protection are salient and accessible. Acquisti et al. (2013) used a field experiment to investigate non-normative privacy behavior and the presence of an endowment effect. In their experiment, mall shoppers received one of two types of gift cards with either traceability or non-traceability of the purchases with the gift card. Results show that those with a lower value, non-traceable gift card were unwilling to exchange for a higher value, traceable card. Thus, consumers' value endowed privacy protections greater than non-endowed privacy protections. Given the landscape of prior literature on privacy valuations, we contribute in two important ways. First, we study multi-dimensional information disclosure decisions, which combine important privacy factors that to the best of our knowledge have not been considered together in

the prior literature. We discuss these dimensions and our corresponding hypotheses in the next section. Second, we address the limitations identified in the prior literature for measuring privacy valuations by introducing methodology from experimental economics.

2.1 Research Model and Hypotheses

We focus our attention on three important factors in privacy disclosure: the *context of the information disclosure*, the intended *secondary use of the disclosed information*, and the *requirement to disclose identifying information*. These factors can affect an individual's privacy and potentially the value an individual places on their privacy. Of the commonly studied privacy factors, these three frequently appear in real online information disclosures, whereas other commonly studied privacy factors such as information accuracy and improper access are less common in practice. Information context directly affects an individual's trusting beliefs, risk beliefs, and behavioral intentions toward an information disclosure (Malhotra et al. 2004). The secondary use of information, i.e., the distribution of disclosed information to third parties, affects an Internet user's privacy decisions and can result in stricter privacy settings (Chellappa and Sin 2005; Angst and Agarwal 2009). Requiring the disclosure of personally identifying information increases vulnerability and the likelihood of experiencing negative consequences from disclosure (Solove 2006).

From an economic view, consumers require utility (e.g., a payoff) to transition from a state of high privacy to a state of low privacy. In the case of monetary payoffs, we use willingness-to-accept (WTA) to represent the monetary value a consumer will accept in order to make this transition (Hanemann 1991). A formal utility function of privacy decision-making that includes WTA follows (cf. Acquisti et al. 2013, page 258). Let $u(w, p)$ be a consumer's utility

regarding wealth, w, and privacy, p, with p^- representing less privacy and p^+ representing more privacy. For any consumer in a position of $u(w, p^+)$, the minimum amount that the consumer would require to enter a state of p^- is given by $u(w + \text{WTA}, p^-) = u(w, p^+)$. Thus if a consumer perceives they are transitioning to a state of lower privacy, they should require a gain in utility in return.

Our first hypothesis considers the effect of information context on a privacy valuation. In a review of the privacy literature, Smith et al. (2011) identified eight differing contexts of private information. The list includes behavioral, biographical, financial, medical, consumer, personal, employee, general, and publicly available information. Each type of private information creates different perceptions of privacy among individuals, and prior research has observed that individuals generally perceive information disclosed in the consumer context as less private and information disclosed in the medical context as more private (Nowak and Phelps 1992; Phelps et al. 2000; Sheehan and Hoy 2000). Thus, in an information disclosure decision we expect that a context perceived as more private leads to a higher WTA for making the disclosure in comparison to a context perceived to be less private.[7] Therefore, we hypothesize the following:

> *Hypothesis 1: Participants asked to disclose information in a more private context (e.g., medical information) will exhibit a higher privacy valuation (WTA) than participants asked to disclose information in a less private context (e.g., consumer shopping information).*

[7] We acknowledge that is possible for the specific type of information disclosed to modify the perceived level of privacy for the context. For example, revealing the purchase of adult videos (shopping context) would be perceived as more private than revealing if an individual is a smoker (medical context). We control for these exceptions to the general perceptions of information context in the design of our experiments as described in Section 4.

Our second hypothesis predicts the influence of an organization's secondary use intentions on a privacy valuation. When organizations gather information on consumers and users they have the option to either distribute (or sell) this information to an external party (external secondary use), or restrict this information for use only in internal operations (internal secondary use) (McMillan 2014). The literature has shown that consumers hold a negative attitude towards external secondary information use (Culnan 1993; Sutanto et al. 2013), which influences their purchasing behavior (Hoffman et al. 1999; Sutanto et al. 2013). In contrast, the same research has shown that consumers hold a positive attitude towards internal secondary information use when it results in enhanced custom product offerings. Further, consumers are willing to pay premium prices to prevent the external secondary use of private information (Hann et al. 2007). Therefore, an individual should require greater compensation to disclose their private information to an organization that practices external secondary use. Formally, we hypothesize the following:

> *Hypothesis 2: Participants asked to disclose information that will be distributed to a third party (i.e., external secondary use) will exhibit a higher privacy valuation (WTA) than participants asked to disclose information that will not be distributed to a third party (i.e., internal secondary use).*

Our third hypothesis predicts that the requirement to disclose personally identifiable information (PII) will affect a person's WTA because it can increase direct risk. Individuals have a strong desire to protect identifying information because it increases vulnerability and the likelihood for others to inflict harm (Solove 2006; Ohm 2010). Consumers also prefer using

online services that have high online privacy protections, such as those that do not collect identifying information, unless they receive adequate financial gain or convenience (Hann et al. 2007). The National Institute of Standards and Technologies (NIST) defines (McCallister 2010, p. ES-1) PII as "any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information." We focus our study on NIST's type 1 in the manipulation of the PII factor, and acknowledge the potential impact of linkable information (type 2) by also testing for interaction effects (discussed in the next section). Markets for securities, bonds, and insurance demonstrate the common economic principle that taking on higher risks leads to higher expectations for compensation. Disclosing PII is inherently risky, therefore we expect a market for privacy to act in a similar manner to these other markets and hypothesize the following:

> *Hypothesis 3: Participants asked to disclose information that requires the inclusion of personally identifying information (PII type 1) will exhibit a higher privacy valuation (WTA) than participants asked to disclose information that does not require personally identifying information.*

To the best of our knowledge, we are the first to study the influence of these three privacy factors together in their impact on privacy valuations. Below we discuss the design and

implementation of a set of experiments that allow us to combine these effects into a single disclosure scenario for our participants.

2.2 Experimental Procedures, Analyses, and Results

We designed and implemented three studies, with each study building upon the previous one. As will be discussed, the results of our first study were surprising – none of the factors showed significant effects on privacy valuations. Therefore, in Study 2 we increased the saliency of privacy factors for participants to help ensure the participants were fully aware of the disclosure dimensions. In Study 3, we went one step further and presented a video that explicitly communicated to participants the potential consequences of disclosing private information. In the following subsections, we describe the experimental procedures and post-experiment survey that are common among the three studies, followed by the unique features and analysis of the studies individually.

2.2.1 Experimental Procedures Common to each Study

We used Qualtrics software to implement a randomized, between-subjects design consisting of eight experimental treatments (2^3 factorial design). The factorial design allowed us to combine factors, test interactions, and gain power advantages with our sample sizes. We advertised the study as joint research collaboration between Google and (BLINDED FOR REVIEW) to gather user feedback in the design of new services. Participants were not aware they were participating in a purely academic study until they were debriefed at the conclusion of their participation, as was required by our university review board. Participants were told Google was conducting market research with the help of (BLINDED FOR REVIEW) and they would be paid for the

disclosure of private information to help Google test compensation models. We used Google as the focal organization across all treatments and all participants to limit potential biases manifesting as treatment effects. Google is a real and well-known firm that regularly collects data about their customers, and with which most participants should have prior experience. Alternatively, using a fictitious or unknown firm would introduce uncontrolled uncertainty into the decision scenario.

The experiment treatments consisted of combinations of the following two-level factors: (1) the context of the information Google requires for their new application (high privacy is a request for medical history, low privacy is a request for shopping preferences), (2) the planned secondary use of the private information by Google (will or will not distribute the information to a third party), (3) whether or not Google requires the disclosure of PII. Participants in the high (low) information context condition were told they must disclose their medical (shopping) history information for a new Google medical (shopping) service. To ensure participants understood what to expect and control for the potential that their expectations of disclosure could impact their perceptions of the context privacy level, we provided participants with a list of items for each context condition prior to requiring their disclosure (see Online Appendix Table A2). Participants in the secondary (non-secondary) use condition were told that Google will (will not) distribute the information they disclose to third party marketing and advertising agencies. Participants in the identifying (non-identifying) information condition were told that they must (will not have to) disclose PII including their full name, email address, and phone number.

To elicit WTA, we implemented the incentive compatible Becker-DeGroot-Marschak (BDM) procedure (Becker et al. 1964). In the absence of an incentive compatible procedure such as the BDM, participants tend to overstate their WTA (Coursey et al. 1987; Plott and Zeiler

2005; Shogren et al. 1994; Hanemann 1991; List and Shogren 2002). The BDM procedure is commonly used in experimental and behavioral economics studies when the good in question is subjective and there is not an established market price (List and Shogren 2002), and in abstract experimental contexts eliciting private values (Rivenbark 2011). Private information is representative of such a good.

BDM's incentive compatibility comes from the binding nature of the exchange. BDM incentivizes participants to be truthful in our experiment because those who sell their information must disclose their real private information and receive real payment for doing so.[8] We implemented the BDM procedure as follows. The system draws a random price for each participant from a uniform distribution between \$0 and \$5. Without knowing the drawn price, the participant provides their WTA for disclosing their information. Their WTA is compared with the drawn price and if the participant's WTA is lower than or equal to the drawn price then their information must be sold at the drawn price. For example, if a participant's WTA is \$2.50 (i.e., they are willing to disclose private information for \$2.50), and the system draws a price of \$3.00, the participant receives \$3.00 and must disclose their information. If the participant's WTA is greater than the drawn price, the participant does not sell their information. For example, if a participant's WTA is \$2.50 and the system draws a price of \$1.00, then no transaction occurs. The participant does not disclose their information and is directed to the post-experiment survey. Therefore, the BDM results in accurate stated valuations because stating a WTA that is higher than an actual valuation may result in a missed opportunity to sell information (i.e., disadvantageous non-selling). In contrast, stating a WTA lower than an actual valuation may result in the participant selling for less than their desired value (i.e., disadvantageous selling). To

[8] We asked participants in the debrief if they disclosed false information and if so removed them from the analysis.

ensure understanding of the BDM, we trained participants on the procedure and required them to pass a quiz before moving on to the experiment.

As noted above, we presented participants with a sample of the disclosure form they must complete if they sell their information, prior to entering their WTA. The sample form included the required identifying information fields if they were in that treatment. Participants viewed and disclosed the same number and types of items in both context treatments. Prior presentation of the sample form helped to control for participant uncertainty prior to the WTA decision.

Following the presentation of the sample form, we asked participants to enter their WTA for disclosing their information, given their treatment scenario. Participants had an unlimited amount of time to enter their WTA and they could opt-out at any point during the experiment. The acceptable range of WTA values was $0.00 - $5.00. As would be expected, actual market prices for private information vary based on specific data types, context, and time. For example, legally obtained batches of user profiles sell for as low as $0.005 per account (Madrigal 2012). The service Datacoup, who pays users for to collect and share their social media data and credit card transactions, pays roughly $8 per month for this right. Furthermore, Facebook reported average revenue per user of $5.07 in 2017 (Shinal 2017), a value largely driven by leveraging user information for targeted advertising. Therefore, the $0.00-$5.00 range we chose for WTA measurements is fair and reflects the approximate magnitude of current market prices for private information. We also note that in our post-hoc analyses zero participants reported opting out because of the WTA range. Additionally, we use Tobit regression models in our analysis to control for WTA censoring due to the selected price range.

After participants submitted their WTA, the system reported the drawn price and the sales outcome. Those who successfully sold their private information then viewed a form that required

them to complete all fields before continuing to the post-experiment survey and receiving payment. Participants who did not sell were directed to the post-experiment survey without completing the form. All items in the post-experiment survey used a 7-point Likert-type scale to capture demographics and relevant controls.[9] To control for Internet experience, survey items included Internet usage, online privacy breach history, and propensity to falsify information online. Participants were instructed on how to receive credit and any monetary payment they were owed for their disclosure upon completion of the survey. Participants were always aware they would receive credit regardless of whether or not they disclosed their private information, in addition to monetary payments for their disclosure.

2.2.2 Study 1: Information Disclosure Factors and WTA

We recruited students from large, upper-division courses that frequently engage in research from the college of management and the school of public policy in a large North American university. We provided a link to the Qualtrics site to interested participants and allowed them to complete the study at their convenience. We sample from a fairly homogeneous population of undergraduate students in Study 1 (Druckman and Kam 2009; Compeau et al. 2012), so any deviations in WTA will likely be due to specific treatment manipulations and not underlying heterogeneity of the participants in the sample (e.g., wealth and other factors that vary across a diverse population). However, there may be specific limitations to generalizability due to the student sample. So in Studies 2 and 3 we sample from both homogenous (i.e., student) and heterogeneous (i.e., Amazon Mechanical Turk) populations to generalize results to a broader population.

[9] The survey items can be found in the Online Appendix.

2.2.2.1 Study 1 Data Analysis and Results

In our first study, we collected data from 394 participants and omitted those who failed attention and manipulation-checking questions, had no variation in their survey responses, failed to complete the experiment, or declined to sell their information by opting out and not providing a WTA. In total, we removed 94 participants based on these criteria, leaving data from 300 participants.[10] Table 1 provides the summary statistics.

Table 1. Summary Statistics by Population, Study, and Treatment ($\overline{WTA}$ grouped by context)

Population	Study	Male	Female	$\overline{WTA}$	Treatment	Obs.	Count of Sold	$\overline{WTA}$ Sold
Students	1	170	130	$3.31	Med-SU-Id	33	9 (27.27%)	$2.91
					Med-No SU-Id	43	11 (25.58%)	$2.79
					Med-SU-No Id	40	12 (30.00%)	$2.93
					Med-No SU- No Id	35	11 (31.43%)	$2.45
				$3.05	Shop-SU-Id	39	9 (23.08%)	$2.36
					Shop-No SU-Id	38	11 (28.95%)	$1.72
					Shop-SU-No Id	35	12 (34.29%)	$2.45
					Shop-No SU-No Id	37	15 (40.54%)	$1.98
	2	114	88	$3.72	Med-SU-Id	21	5 (23.81%)	$2.50
					Med-No SU-Id	27	7 (25.93%)	$2.64
					Med-SU-No Id	27	8 (29.63%)	$2.94
					Med-No SU- No Id	25	10 (40.00%)	$2.29
				$3.28	Shop-SU-Id	25	7 (28.00%)	$1.70
					Shop-No SU-Id	24	7 (29.17%)	$2.07
					Shop-SU-No Id	24	9 (37.50%)	$1.85
					Shop-No SU-No Id	29	8 (27.59%)	$2.06
	3	77	63	$3.36	Shop-SU-Id	32	9 (28.13%)	$1.75
					Shop-No SU-Id	36	13 (36.11%)	$2.00
					Shop-SU-No Id	34	11 (32.35%)	$2.24
					Shop-No SU-No Id	38	14 (36.84%)	$1.79
AMT	2	237	199	$3.49	Med-SU-Id	58	9 (15.52%)	$1.76
					Med-No SU-Id	49	13 (26.53%)	$2.06
					Med-SU-No Id	65	26 (40.00%)	$1.88
					Med-No SU- No Id	54	17 (31.48%)	$2.31
				$3.47	Shop-SU-Id	49	12 (24.49%)	$2.08
					Shop-No SU-Id	58	18 (31.03%)	$2.35
					Shop-SU-No Id	54	21 (38.89%)	$2.18
					Shop-No SU-No Id	49	20 (40.82%)	$2.07
	3	91	91	$3.76	Shop-SU-Id	46	10 (21.74%)	$1.95
					Shop-No SU-Id	46	9 (19.57%)	$1.53
					Shop-SU-No Id	45	16 (35.56%)	$1.80
					Shop-No SU-No Id	45	9 (20.00%)	$2.04

Note: Med = Medical Context; Shop = Shopping Context; SU = Secondary Use; Id = Identifying Information

First, we tested for treatment effects using an ANCOVA model that includes the dependent variable WTA, control variables (gender, web usage, propensity to provide false

[10] A limitation of Qualtrics is that a new response is generated each time the hyperlink to begin the study is clicked. Therefore, it is necessary to remove more than 16% of responses due to being substantially incomplete. Please refer to Tables A9 and A10 in the Online Appendix for a detailed breakdown of responses removed from each study.

information, indicated prior privacy breach), and main effects (three indicator variables representing the manipulated factors: information context; secondary use; identifying information). Results indicate a marginal difference (yet not significant at the $\alpha = 0.05$ level) in the mean of WTA for the information context main effect ($F = 3.17$, $p = 0.076$).

Next, we used a Tobit regression to uncover the magnitude of the effects. Tobit estimations are common in modeling willingness-to-pay/accept measurements because of the natural censoring of the dependent variable (e.g., Donaldson et al. 1998). We report only the average marginal effects because the baseline Tobit estimates reflect the marginal effects on the unobserved and uncensored dependent variable (McDonald and Moffitt 1980). Table 2 Model 1 presents the results for Study 1. The results do not suggest any significant main effects.[11]

Table 2. Main Experimental Results Using Tobit Regressions

	Study 1	**Study 2**		**Study 3**	
Population	**Students**		**AMT**	**Students**	**AMT**
Variables	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Gender	0.079 (0.147)	0.029 (0.182)	-0.246* (0.125)	0.244 (0.257)	-0.215 (0.217)
Age			0.205** (0.058)		0.077 (0.096)
Education			-0.052 (0.055)		-0.033 (0.081)
False Information	-0.031 (0.082)	0.143 (0.105)	0.166* (0.066)	0.000 (0.163)	0.040 (0.123)
Web Usage	-0.113 (0.083)	0.055 (0.104)	0.010 (0.057)	-0.059 (0.141)	-0.111 (0.085)
Breach History	0.022 (0.052)	0.044 (0.064)	0.056 (0.047)	-0.040 (0.091)	-0.084 (0.075)
Information Context	0.223 (0.144)	0.416* (0.175)	0.111 (0.129)		
Secondary Use	-0.005 (0.146)	-0.112 (0.176)	-0.106 (0.130)	0.137 (0.251)	-0.091 (0.216)
Identifying Information	0.149 (0.147)	0.150 (0.178)	0.289* (0.130)	0.259 (0.255)	0.204 (0.221)
Constant	3.371** (0.406)	2.621** (0.571)	2.509** (0.552)	3.661** (0.913)	5.281** (1.065)
F-value	0.950	1.550	3.040**	0.520	0.750
Log pseudolikelihood	-498.865	-329.653	-729.350	-245.276	-293.726
Censored Observations	3 censored at 0 43 censored at 5 254 uncensored	1 censored at 0 52 censored at 5 149 uncensored	7 censored at 0 130 censored at 5 299 uncensored	4 censored at 0 41 censored at 5 95 uncensored	8 censored at 0 76 censored at 5 98 uncensored

Robust standard errors in parentheses, †$p \leq 0.10$, *$p \leq 0.05$, **$p \leq 0.01$. Average marginal effects reported for the coefficients.

[11] We also estimated a control variables only model, a fully interacted model, standard OLS regression models, and models incorporating (-1,1) effect coding for the main effects. Results of these models are consistent with those reported in Table 2. No significant interaction effects were found in any of the models.

2.2.3 Study 2: Increasing the Saliency of Treatment Factors

Consumers make privacy conscious decisions when privacy information is salient (Tsai et al. 2011). Furthermore, a common explanation for the privacy paradox is that users lack awareness of the risks involved in disclosing private information (e.g., Acquisti and Gross 2006, Tufekci 2008). Therefore, in Study 2 we increased the saliency of privacy factors as a methodological check to help rule out a lack of awareness as to what was being asked of the participants. According to Taylor and Fiske (1978), changing the color (e.g., bolding) and size of images (e.g., enlarging font) within textual descriptions lead to greater participant attention and recall of information. Further, prior work in computer vision has termed visual saliency as the idea that humans have significant visual arousal from changes in scene (Kadir and Brady 2001). Julesz (1995) and Treisman (1985) identify these changes as 'pop-out' and create visual cues that stand out and aid in human visual saliency (i.e., attention grabbing). Accordingly, we enhanced the text describing each factor in the experiment, using larger bold fonts with italics and underlining, to increase visual saliency. We also included a summary page that reiterated the three manipulations immediately before participants were asked to enter their WTA (please refer to Tables A1, A4, and A6 of the Online Appendix). Otherwise, the procedure remained identical to that of Study 1.

To broaden the generalizability and implications of our research, we conducted Study 2 using two different populations: a homogenous student population and a heterogeneous Amazon Mechanical Turk (AMT) population. We implemented qualification controls and payments on AMT following prior work (e.g., Balebako et al. 2015), required the workers to be aged greater than 18, from the United States (US), have an acceptance rate greater than or equal to 89%, and a time limit of 30 minutes. We paid AMT workers $1.35 for their time (based on the national

average for the US minimum wage of $7.25 and the average time to complete the study in pilot testing) plus a bonus payment for selling their private information using the same BDM mechanism we used with students.

The student sample size for Study 2 is 202 after removing 68 participants for the same reasons as in Study 1.[12] The AMT sample size is 436 participants after removing eight participants for failing manipulation checks. Within the AMT sample, 46% of participants were female, age ranged from 18 to 74 (46% were older than 35), and 51% had at least a Bachelor's degree.

2.2.3.1 Study 2 Data Analysis and Results

Consistent with Study 1, we first tested for treatment effects using an ANCOVA. Regarding the student sample, the results indicate a significant difference in the mean of WTA for the information context main effect ($F = 6.25$, $p = 0.013$). Regarding the AMT sample, the results indicate a marginal difference (yet not significant at the $\alpha = 0.05$ level) for the identifying information main effect ($F = 3.37$, $p = 0.067$). Again, we estimated a Tobit regression (Table 2) and the average marginal effects in Model 2 (students) illustrate a significant impact of information context on WTA. Changing from the shopping context to the medical context results in a $0.42 ($p = 0.017$) increase in WTA. We note the fit for Model 2 is marginal, thus the result is not conclusive. The average marginal effects in Model 3 (AMT) illustrate a significant impact of identifying information on WTA. Changing from not including identifying information in the disclosure to including identifying information results in a $0.29 ($p = 0.026$) increase in WTA.

[12] Power analyses provide evidence that we have sufficient power to detect small to medium effect sizes using our factorial design. Please refer to the Online Appendix for additional details.

Although we observed these two significant main effects, they were not consistent across populations or models.

2.2.4 Study 3: Highlighting the Consequences of Private Information Disclosure

The persistence of null effects in Study 2 led to our design of Study 3. It includes an educational video that clearly highlights and describes the consequences and risks of disclosing private information online (the video can be viewed at https://goo.gl/X2C5lj). The video provides an additional methodological check to help rule out the explanation that our participants were unaware of the risks associated with private information disclosure. Participants were required to watch the video before beginning Study 3.

The video covered three aspects related to the consequences of information disclosure. First, it clearly defined external secondary use of private information and personally identifying information. Second, the video included actual examples of firms using these practices. Third, the video discussed four consequences of disclosing private information when external secondary use and PII are present. The video also presented multiple news article snippets supporting each consequence. Following the video, participants completed a quiz on the consequences of disclosing private information. The experiment did not begin until a participant correctly answered all quiz questions.

We fixed the information context to shopping preferences for all participants in Study 3 to simplify the experimental design and focus on the two factors consumers would generally expect of Google as an organization. Doing so allowed us to focus substantial time during the video on only two factors, instead of spending smaller amounts of time split up among the three

factors. This resulted in a 2x2 design for Study 3 with the remainder of the procedure replicating the procedure of Study 2.

The student sample size for Study 3 was 140 usable participants (initially 175), and the AMT sample size was 182 usable participants (initially 200). We followed the same procedures for removing participants as in Study 2, with the addition of also removing those that did not watch the full video. Within the AMT sample, 50% of participants were female, age ranged from 18 to 74 (51.65% were older than 35), and 50% completed at least a Bachelor's degree. We paid AMT workers $2.66 for their time[13] plus a BDM-based bonus payment for selling their information as with the students.

2.2.4.1 Study 3 Data Analysis and Results

As before, we first tested for treatment effects using an ANCOVA. The results are largely consistent with the previous studies except that we find no significant main effects, regardless of population. The Tobit regressions (Table 2) also demonstrate no significant main effects. Note the intercept in Model 5 (AMT) is greater than $5.00, which is unusual given the censoring of WTA at $5.00. We conducted several post estimation tests to identify the underlying cause, including the Shapiro-Wilk test for normality of the error term, Breusch-Pagan test for heteroscedasticity, and variance inflation for multicollinearity. Based on the results, we conclude that the model contains heteroscedasticity.[14] We estimated two models that are robust against heteroscedasticity and the results were consistent with the Tobit regressions already presented.[15]

[13] The slightly higher base payment in Study 3 was used because the participants had to watch the approximately 7 minute video and complete a quiz on the video prior to starting the study. We again based this payment on the prorated US minimum wage.

[14] The null hypothesis for the Breusch-Pagan test is H_0: Constant Variance. The results from the test provide that $\chi^2 = 3.61$ and $p = 0.058$, suggesting the presence of heteroscedasticity.

[15] Omitted here for sake of brevity. Please refer to the Online Appendix for further details.

2.2.5 Post Hoc Analyses and Comparisons across Studies

With the exception of two population dependent significant main effects in Study 2, null effects persisted across all three studies, even after increasing saliency of the disclosure factors and having participants watch a video on the consequences associated with disclosing information. At face value, it suggests that consumers are largely neutral to these privacy factors in a disclosure. However, it is also possible that there are effects that did not manifest in the analysis of the main factors. Therefore, we conducted several post hoc analyses to determine the manner in which WTA valuations were affected across studies.[16]

We pooled the data across studies to compare effects of each study on WTA as shown in Table 3.[17] Study 3 fixed information context to shopping preferences, so our comparison across studies is limited to that factor level. Tobit regressions included an indicator variable for Study 1, 2, or 3. We used Study 1 as the baseline for comparison in Models 1-2 and Study 2 as the baseline in Models 3-4. Increasing the saliency of the privacy factors associated with a disclosure request resulted in a statistically significant increase in average WTA (Study 2 indicator, Model 1). Interestingly, the average WTA did not significantly change across the studies when Study 2 is the baseline for comparison (Study 1 and Study 3 indicators, Model 3). However, increased saliency combined with consequences shown in the video did increase the average WTA for Study 3 in comparison to Study 1 (Model 2), and Study 3 in comparison to Study 2 (Model 4).

Table 3. Pooled Analysis Using Tobit Regressions				
Context	**Both**	**Shopping Only**		
Pooled Studies	**1, 2**	**1, 2, 3**		**2, 3**
Population	**Students**			**AMT**
Variables	**(1)**	**(2)**	**(3)**	**(4)**
Gender	0.051 (0.114)	0.068 (0.141)	0.068 (0.141)	-0.171 (0.138)
Age				0.115† (0.062)

[16] Heteroscedasticity was not present in any of the post hoc analyses.
[17] We also tested for interaction effects between experimental factors to determine if the increased saliency or the video affected the magnitude of the treatment effects. No interactions were statistically significant.

Education				-0.027 (0.058)
False Information	0.044 (0.065)	0.020 (0.087)	0.020 (0.087)	0.147* (0.070)
Web Usage	-0.033 (0.065)	-0.139† (0.082)	-0.139† (0.082)	-0.051 (0.057)
Breach History	0.032 (0.040)	0.006 (0.050)	0.006 (0.050)	-0.022 (0.050)
Information Context	0.329** (0.112)			
Secondary Use	-0.054 (0.112)	-0.042 (0.138)	-0.042 (0.138)	0.015 (0.137)
Identifying Information	0.147 (0.113)	0.195 (0.143)	0.196 (0.143)	0.316* (0.139)
Study 1 Indicator			-0.267 (0.165)	
Study 2 Indicator	0.378** (0.117)	0.267 (0.165)		
Study 3 Indicator		0.399* (0.165)	0.132 (0.187)	0.311* (0.139)
Constant	2.866** (0.342)	3.349** (0.455)	3.769** (0.468)	3.224** (0.587)
F-value	2.850**	1.440	1.440	2.220*
Log pseudolikelihood	-832.075	-675.572	-675.572	-651.333
Censored Observations	4 censored at 0 95 censored at 5 403 uncensored	8 censored at 0 82 censored at 5 301 uncensored	8 censored at 0 82 censored at 5 301 uncensored	10 censored at 0 127 censored at 5 255 uncensored

Robust standard errors in parentheses, †$p \leq 0.10$, *$p \leq 0.05$, **$p \leq 0.01$.
Average marginal effects reported for the coefficients.

It is also possible that on average, participants price themselves out of the market (e.g., set WTA close to $5 to decrease the probability of selling their private information), due to increased saliency and heightened awareness of consequences. To test for this possibility, we estimated Logit regressions with a binary dependent variable indicating whether a participant entered a high WTA (i.e., greater than $4.95) or not. Table 4 displays the results. Similar to the pooled analysis in Table 3, we include an indicator variable for each study. Note the baseline study of comparison by the omitted study variable in each model. The results in Model 1 indicate that participation in Study 2 led to an increase in the likelihood of participants pricing themselves out of the market as compared to Study 1 for the student population. Model 2 indicates a significant increase in the likelihood of a WTA greater than $4.95 when comparing Study 3 to Study 1 for the student population, and Model 3 indicates a marginal increase in this likelihood

between Studies 2 and 3. Model 4 indicates a significant increase in the likelihood of a WTA greater than $4.95 for the AMT population when comparing Study 3 to Study 2. Thus, these models largely support the finding that increasing the saliency of privacy factors alone (Study 2 compared to Study 1) led to a greater likelihood of a participant pricing themselves out of the market, and heightening awareness using the consequences video (Study 3 compared to Study 2) extended this effect.

Table 4. Logit Pricing Out of Market

Context	Both		Shopping Only					
Pooled Studies	1, 2		1, 2, 3				2, 3	
Population	Students						AMT	
Variables	(1) AME	(1) OR	(2) AME	(2) OR	(3) AME	(3) OR	(4) AME	(4) OR
Gender	0.038 (0.037)	1.278 (0.302)	0.039 (0.043)	1.266 (0.332)	0.039 (0.043)	1.266 (0.332)	-0.083† (0.047)	0.668† (0.150)
Age							0.039* (0.020)	1.211* (0.118)
Education							0.002 (0.018)	1.012 (0.088)
False Information	0.014 (0.019)	1.096 (0.133)	0.014 (0.022)	1.090 (0.151)	0.014 (0.024)	1.090 (0.150)	0.046* (0.022)	1.251* (0.134)
Web Usage	-0.014 (0.022)	0.913 (0.126)	-0.011 (0.024)	0.936 (0.135)	-0.011 (0.024)	0.936 (0.135)	0.004 (0.019)	1.018 (0.094)
Breach History	0.011 (0.012)	1.072 (0.084)	0.010 (0.014)	1.062 (0.092)	0.010 (0.014)	1.062 (0.092)	-0.015 (0.016)	0.930 (0.073)
Information Context	0.058† (0.035)	1.452† (0.331)						
Secondary Use	-0.017 (0.035)	0.895 (0.203)	-0.051 (0.041)	0.732 (0.185)	-0.051 (0.041)	0.732 (0.185)	0.014 (0.046)	1.073 (0.239)
Identifying Information	0.028 (0.036)	1.196 (0.273)	0.046 (0.041)	1.328 (0.337)	0.046 (0.041)	1.328 (0.337)	0.088* (0.046)	1.533† (0.346)
Study 1 Indicator					-0.068 (0.055)	0.658 (0.224)		
Study 2 Indicator	0.096** (0.035)	1.847* (0.421)	0.068 (0.055)	1.519 (0.518)				
Study 3 Indicator			0.151** (0.047)	2.523** (0.755)	0.083† (0.050)	1.661† (0.515)	0.164** (0.043)	2.220** (0.497)
Constant	-2.185** (0.609)	0.113** (0.069)	-2.080** (0.669)	0.125** (0.142)	-1.662* (0.672)	0.190* (0.128)	-2.078** (0.658)	0.125** (0.082)
Wald x^2	14.060†	14.060†	15.400*	15.400*	15.400*	15.400*	29.090**	29.090**
Log Likelihood	-245.001	-245.001	-197.022	-197.022	-197.022	-197.022	-234.506	-234.506

Robust standard errors in parentheses, †$p \leq 0.10$, *$p \leq 0.05$, **$p \leq 0.01$. AME (Average Marginal Effects). OR (Odds Ratio)

Lastly, we implemented manipulation checks in the post-survey to determine if participants believed the privacy factors were risk inducing. Manipulation checks included: (1) "How does the inclusion of Name, Date of birth, and Email with other private information affect the risk associated with disclosing your private information?" (2) "How does the knowledge that

Google will provide your private information to a third party affect the risk associated with disclosing your private information?" We measured each item using a 7-point Likert-type scale: significantly less risk (1) to significantly greater risk (7). Participants only viewed the survey items if assigned to a treatment that had external secondary use for private information or included identifying information in the disclosure. Pooling the data provided the following results: Survey item (1) Students mean response of 5.69, AMT mean response of 4.58; Survey item (2) Student mean response of 5.73, AMT mean response of 4.67. Comparing the means to the risk neutral option (i.e., neither greater nor less risk; value = 4.00) using t-tests provides evidence of greater risk when external secondary use (Students $t = 29.32$, $p < 0.001$; AMT $t = 6.70$, $p < 0.001$) and the requirement of identifying information (Students $t = 30.67$, $p < 0.001$; AMT $t = 7.63$, $p < 0.001$) were present in their disclosure. Based upon these results, participants perceived significant manipulation, but their heightened risk did not consistently translate to higher valuations due to these factors.

2.3 Discussion

Overall, the consistent finding of our research is the persistence of null effects of the three common privacy disclosure factors: requiring the disclosure of personally identifying information, external secondary use of personal information, and the context of the information disclosure. With the exception of two, population-specific main effects, we largely observed null effects across multiple studies and populations. This is in contrast to prior research, which has shown that the presence of these three privacy factors can affect consumers' privacy valuations, when they are studied separately and in isolation (e.g., Culnan 1993; Hoffman et al. 1999; Solove 2006; Hann et al. 2007; McMillan 2014). So, a question arises, why do we observe these

consistent null effects? By combining these factors and requesting multiple pieces of personal information, we presented participants with a more realistic disclosure scenario than is typical in prior research. The persistence of the null effects in this scenario suggests that the effects on privacy valuations get complicated when we combine these factors in a disclosure decision. Even when we heightened participant awareness through the increased saliency and the consequences video null effects largely persisted. Thus, the complexity of the realistic scenario may make it difficult for consumers to disentangle the factors affecting the privacy disclosure, demonstrating the resilience of the privacy paradox. This is a significant finding and contribution to privacy valuation research, which poses a need for continuing research on privacy valuations in the context of complex and realistic scenarios that consider the combination of multiple disclosure factors.

Our post hoc analysis provides additional insights on how privacy valuations operate in a scenario with multiple factors. In our replication studies, we introduced increased saliency of the factors and required participants to watch a video on the risks and consequences of disclosing personal information. We used the increased saliency and consequences video as methodological checks to rule out lack of awareness as an explanation for our null effects. Thus, we did not hypothesize direct effects of these checks on overall privacy valuations. Although including these checks did not counter the null effects of the three privacy factors, we did observe average higher valuations and a higher likelihood of participants pricing themselves out of the market (i.e., stating a WTA greater than $4.95 out of $5.00) as a result. This suggests that even though the main privacy disclosure factors did not have consistent significant effects on privacy valuations, participants did react on average when awareness of privacy disclosure risks was heightened. In other words, consumers may make an all or nothing decision: they either accept or

reject the disclosure opportunity when awareness is heightened, regardless of the conditions. These results provide a second important contribution to privacy valuation research and offer an opportunity for further research around saliency and awareness.

Beyond the null effects, it is also necessary to address the population-specific main effects observed in our studies. First, requiring identifying information did have a significant effect on WTA in the AMT sample for Study 2. Prior research has shown that AMT workers in particular place a high degree of importance on the security of their personally identifiable information (Lease et al. 2013), and thus it is not surprising that we observe that main effect to be significant in our essay as well. The effect does not persist to the AMT sample for Study 3, however. This is likely because the video further heightened awareness and led to an overall higher WTA and likelihood of pricing out of the market, which overcame the specific main effect of identifying information. Second, we observed a significant effect of information context in the student sample for Study 2. One explanation for the result is that the increased saliency of the factors removed the lack of awareness for our participants, resulting in the manifestation of a significant effect of information context, as has been shown in the prior research. The ANCOVA results for Study 2 and pooled regression analysis for studies 1 and 2 provide statistical support for this result. However, the lack of evidence in the AMT sample for Study 2 and the relatively low model fit for the regression analysis of the student sample in Study 2 provide evidence to the contrary. Thus, we have observed only marginal and inconsistent support for the effect of information context on privacy valuation and cannot rule out the possibility that the observed effect is simply due to chance. This inconsistency provides additional evidence that incorporating multiple factors into a privacy disclosure decision complicates effects beyond what prior research has observed by studying factors in isolation.

Legal scholars and policy makers often use the findings from studies on privacy tradeoffs and valuations to inform judicial decisions and aid in the establishment of new precedents (Romanosky and Acquisti 2009). For example, some state courts in the U.S. are considering restricting an organization's ability to gather forms of identifying information and share it with outside parties (Ribeiro 2015). Our results suggest that policy makers may also serve consumers by providing educational services focused on the benefits and consequences of disclosing private information. Policy makers may also consider requiring businesses that collect private information to provide salient descriptions of the private information they capture, their intended uses for that information, and the associated risks involved.

We acknowledge that our findings are not without limitations. The consistent use of Google in all of our studies provides experimental control and suggests the results may be generalizable for large well-known firms but not necessarily new players. It is possible, however, that participants will react differently to unknown or fictitious focal organizations and exhibit differing privacy valuations because of the organization. It is also possible that participants will have mixed reactions or biases about other well-known firms such as Amazon or Apple, Inc. Therefore, future studies should consider investigating the degree to which changing the organization in the disclosure decision affects a privacy valuation.

Another possible limitation of our research is the use of a specified range of $0 - $5 for WTA. We chose this range because it reflects current market prices and our interest in the changes in relative value of a participant's WTA due to the factors we manipulated. We account for any potential censoring of the data due to the range of WTA values by using Tobit regressions. Further, none of our participants reported dissatisfaction with the price range in our post hoc survey. Future work could modify the range of WTA values and consider how the

specified price interval affects WTA in privacy valuations. However, we suspect that any observed differences would likely be due to anchoring effects (e.g., Tversky and Kahneman 1974) and not privacy factors.

3. Essay 2 - Fool Me Twice: An Analysis of Repeat Data Breaches within Firms

One of the key features of data breach notification laws is that they vary across each state with differing policies that comprise the law as well as the time of their adoption. For example, California enacted a data breach notification law in 2003 which requires firms provide notification of breach to the California Attorney General's office. On the other hand, Florida enacted a data breach notification law in 2014 which requires firms provide notification of breach to both the Florida Attorney General's office in addition to specific consumer reporting agencies. As a result, the question of whether deterrence stems from the mere presence of a notification law requiring firms disclose breaches to the public or if specific policies within the law are driving the deterrence effect is of great importance in understanding firms' responses to a data breach (Cohen 2000). We build upon this question by answering *"To what extent do policies within data breach notification laws lower the risk of a future breach?"*

Additional questions arise regarding the number of firms being breached multiple times. Given that almost one out of every three firms experience a future subsequent breach, firms' apparent ineffective means of preventing future breaches calls into question the types of countermeasures they are implementing as well as their usage of resources (e.g., security technologies, monetary funds, employee efforts) for improvements and changes. Thus, we formalize two research questions focused on guiding firms toward a more efficient usage of their information security resources. Specifically, the second and third research questions we address in this paper include *"Are particular industries more susceptible to certain types of breaches?"* and *"Does the breach type increase the likelihood of experiencing a similar breach type in the future?"*

Many scholars (e.g., Schwartz and Janger 2007; Romanosky and Acquisti 2009) argue that the intent behind data breach notification laws is to publicly inform consumers of firm events and practices that organizations may otherwise be unwilling to disclose and subsequently promote change within the organization. Specifically, notification laws seek to increase consumer awareness and force firms to incur further financial costs by mailing notification letters to affected consumers, providing customer support call centers, and administering identity theft protection services. The average cost of a single data breach was roughly $4 million in 2016, but the average cost doubles to almost $10 million for organizations experiencing multiple data breaches (Romanosky 2016). Organizations cover a majority of the financial costs associated with a breach, but consumers face time and effort inconveniences that leave them disgruntled with the organization, thereby leading to the chance of decreased sales (Salane 2009). The annual cyber security report by Cisco Systems reports that breached firms in 2016 experienced a 20% - 29% loss in revenue for the year (Cisco Systems 2017).

In addition to the monetary costs and revenue loss following the public notification of a data breach, scholars have also studied the effects of a data breach announcement on stock market performance. Initial investigations on the impact to stock valuations gave conflicting stories as Hovav and D'Arcy (2003), Campbell et al. (2003), and Kannan et al. (2007) did not find abnormal returns in their stock valuation studies. However, Garg et al. (2003) and Cavusoglu et al. (2004) discovered abnormal returns of -5.3% over a three-day window and -2.1% over a two-day window respectively. Acquisti et al. (2006), Ko and Dorantes (2006), Gatzlaff and McCullough (2010), and Gay (2015) later provided further support of a negative stock market response and a reduction in firm performance following the announcement of a data breach. The reason for conflicting results across studies is likely due to differences in the

methodologies, datasets, market valuation metrics, and the window used by the essay. Gordon et al. (2011) and Kvochko and Pant (2015) appear to have resolved these inconsistencies, as their work offers conclusive evidence suggesting that firms have initial decreases in stock valuations that then normalize over the following months.

The requirement imposed by some regulatory agencies for public breach disclosure also has a negative effect on an organization's brand and generates negative publicity (Leonard and Rubin 2006). Fully compliant organizations may still fall victim to reputation loss and decreased market value following a breach, but the extent of the damage is dependent upon who public opinion deems responsible for the incident (Spiekermann et al. 2015; Acquisti et al. 2006). The damage to an organization's reputation increases when public opinion believes that the breach is the result of negligence. Considering the negative financial and reputational effects of a breach disclosure in combination with the prior findings that firms make operational changes post breach, it is reasonable to assume that firms will take preventive actions (e.g., enhancing their security and implementing new information security policies) to both publicly demonstrate their response to the breach and further diminish their chances of a subsequent breach (i.e., a deterrence effect encourages firms to take action).

3.1 Characteristics of Data Breach Notification Laws

General deterrence theory suggests that strong and enforceable penalties discourage individuals and firms from committing or allowing specific acts (Straub and Welke 1998). Blumstein et al. (1978) identified the discouragement to be comprised of the certainty of a sanction[18] and the severity of a sanction. In the context of data breaches, an emerging stream of literature that examines the presence of deterrence argues that the direct and indirect

[18] Sanctions are threatened penalties for not adhering rules or policies.

penalizations[19] businesses experience because of breach regulations affects the incentives of firms to invest in preventive information security measures (Laube and Bohme 2016). Specifically, maximizing the collection of consumers' personal data offers organizations a significant competitive advantage but also invites malicious activity to compromise the information, which can temporarily negate the benefits derived and result in tremendous short-term monetary loss (Spiekermann et al. 2015). Samuelson Law, Technology, and Public Policy Clinic (2007) and Schneider (2009) were among the first to investigate and identify breach notification's deterrence effect as organizations reported taking significant measures to improve their information and operational security in response to data breach regulations and awareness of consumer reactions. Moreover, Sen and Borle (2015) suggested that deterrence is present because states that have breach notification legislation experienced longer durations between reported breaches in the state. We can then expect the presence of a breach notification law to significantly influence the risk of future subsequent breach. Formally, we hypothesize that:

Hypothesis 1: The presence of a data breach notification law will reduce the risk of future subsequent breach.

We argue that the prior empirical work has demonstrated an aggregate deterrence effect by focusing solely on the presence of a state breach notification law (i.e., an aggregate effect arising from the combined policies within particular laws). As mentioned above, data breach

[19] Direct penalizations are fines or penalties handed down by an overseeing agency for violating a policy or law. The direct penalizations for a firm experiencing a data breach include notifying consumers, providing compromised consumers with identity theft protection, and settling consumer lawsuits. Indirect penalizations are negative consequences stemming from experiencing the direct penalization. Firms experience indirect penalizations following consumer notification of the breach, which include drops in stock valuation, revenue loss from negative publicity, and a decrease in firm performance.

notification laws are enacted and controlled at the state-level. Therefore, laws may be enacted at different points in time and policies within the laws may vary. Thus far, the deterrence effect found in the literature does not account for such discrepancies. Hence, to fully understand the impact of a data breach notification law's influence on the risk of future subsequent breach we disaggregate the effect by delving into the specific policies comprising each state's notification law.

The first policy we investigate is the allowance of firms to refuse public notification of data breach if the firm can demonstrate that consumers are not and will not be harmed by the breach. In the wake of a data breach, firms located in certain states conduct a risk of harm analysis, which is comprised of analyzing the extent of the information breached, the inclusion of identifying information (e.g., name, social security number, or phone number), whether the information was acquired or viewed, and the extent of protection given to the information (e.g., encryption and anonymization methods). Firms that conclude from their risk of harm analysis that consumers are unlikely to experience harm are not required to provide public notification, but must maintain adequate documentation of their investigation and, in some instances, submit the documentation to an overseeing agency. We posit that allowing firms to forego public notification removes the incentive for firms to implement information security countermeasures and organizational change as they are less likely to experience the negative consequences mentioned earlier that are associated with breach notification. Similarly, Arora et al. (2008; 2010) demonstrate that firms operate in a less than socially optimal manner for addressing software vulnerabilities unless those vulnerabilities are publicly disclosed. We then hypothesize:

Hypothesis 2: Allowing firms to opt out of public notification if the breach presents no immediate or future risk of harm to consumers increases the risk of future subsequent breach.

The next notification policy we study is the requirement for firms to provide public notification if the breach involves paper records. Currently the Breach Notification Rule within the Health Insurance Portability and Accountability Act (HIPAA) provides that medical firms and business associates must notify of a breach involving non-electronic (e.g., paper) records. State data breach notification laws containing a provision for public notification of a paper records breach seek to extend the HIPAA regulation by expanding it to other industry types. This provision increases the liability to firms who have yet to implement an electronic system or continue to maintain non-electronic record keeping. Similar to the risk of harm policy, the paper record notification policy may increase public awareness to a security vulnerability within the firm and consequentially the firm experiences negative repercussions. States without a paper record notification policy may feel little to no incentive to take corrective action because negative consequences are unlikely. Thus, we formally hypothesize:

Hypothesis 3: Requiring firms to publicly notify a breach of paper records containing private consumer information reduces the risk of future subsequent breach.

The third policy found in data breach notification laws is the allowance of individuals caught maliciously obtaining or trafficking consumer private information to face criminal

charges. [20] The criminal charges policy is distinct from other breach notification policies because it attempts to deter at the individual-level rather than at the firm-level. The firm itself and top-level management, assuming they did not commit the breach, cannot be charged with criminal penalties. Therefore, the policy is aimed at the employees within the firm.

Those found guilty of committing a data breach can receive a Class A misdemeanor, which includes a jail sentence of one year. The key stipulation of the policy is that the individual must demonstrate malicious intent with the private information. Thus, employees who unintentionally disclose private information through negligence or by mishap are exempt from the provision, but employees who are insider threats or outside individuals who hack a computer system may be found guilty of a criminal misdemeanor. It is common practice for policymakers to use sanctions as a means to deter from not adhering to policies and rules as they have been found useful in corporate settings (Akers 1990; Tyler and Blader 2005).

Extensive work in the IS security literature (e.g., Straub 1990; Straub and Nance 1990; Straub and Welke 1998; Kankanhalli et al. 2003; Pahnila et al. 2007; Bulgurcu et al. 2010) has shown that stringent sanctions deter employee cybercrime and computer abuse when IS security policies are well-defined, monitored, and consistently punished. We then posit that deterrence may not occur prior to the first breach within an organization because it is unlikely that firms routinely inform their employees of the data breach laws and the possible repercussions for not adhering to those laws. However, we argue that deterrence is likely to be present for subsequent

[20] The criminal charge policy found in a data breach notification law is different from the criminal charges individuals face from state level identity theft laws. Every state within the U.S. has a law regarding identity theft or impersonation. Identity theft is the use of another person's personally identifiable information and may occur as a result of a data breach. The difference between the laws is that a data breach involves coming into possession of consumer information from a firm while identity theft is the actual usage of that consumer information to commit fraud. As of this writing, many states have provisions in their identity theft laws such that individuals found in possession of or are trafficking consumer information will face criminal misdemeanor or felony charges. However, states with the criminal charge policy within their data breach notification laws do not have those provisions in their identity theft laws.

breaches following an initial breach because employees may learn of the individual responsible and witness the penalties he or she face. Hence, we formally hypothesize:

Hypothesis 4: Allowing criminal charges to be brought against an individual or party that maliciously breaches consumer information will reduce the risk of future subsequent breach.

The remaining policies include whether or not a breached firm must disclose the breach to a state attorney general or a consumer reporting agency. A state attorney general serves as the counselor for state agencies and legislatures as well as the lawyer representative for the people of that state, and it is their responsibility to enforce state laws and propose new law or amendments to an existing law. In the case of data breaches, an attorney general may represent citizens of that state in bringing civil penalties against the firm for improper handling of consumer information or administer sanctions against firms when there is reason to believe the firm did not adhere to the state's data protection laws. Firms recognize and are likely to respond to such sanctions by increasing their data protection capabilities (Romanosky and Acquisti 2009). Therefore, we posit that firms will take precautions in order to avoid such scrutiny by the attorney general; thereby lowering the risk of future subsequent breach when firms must notify a state attorney general. Formally, we hypothesize:

Hypothesis 5: Requiring firms to disclose a data breach to a state attorney general will decrease the risk of future subsequent breach.

Another entity firms may be required to disclose a data breach to is a consumer reporting agency. Consumer reporting agencies collect and share information regarding a person's private information (e.g., insurance claims, employment history, credit history, banking history). A reason for having firms disclose breach information to the agency is to aid consumers against poor credit scores and history. A breached firm may provide a consumer reporting agency with a list of the consumers who have had their information compromised. The agency may then make note of the breach and associate it with the person's information to explain changes in the individual's reports. With this information, we posit that firms receive little to no incentive to change their business processes as a consumer reporting agency cannot impose sanctions against the firm. Therefore, we believe the lack of sanctions will not deter firms from experiencing future subsequent breaches because firms lack a reason to make changes. Hence, we formally hypothesize:

Hypothesis 6: Requiring firms to disclose a data breach to a consumer reporting agency will increase the risk of future subsequent breach.

Lastly, data breach notification laws may not require firms disclose breach information to an attorney general or a consumer reporting agency. By not requiring firms to disclose the breach to another entity, firms have little responsibility for taking corrective and preventive action to resolve the breach (Dickler 2018). We then posit that the lack of incentive for change will lead to greater risk of future subsequent breach. Thus, we formally hypothesize:

Hypothesis 7: Requiring firms to disclose a data breach to a state attorney general or a consumer reporting agency will decrease the risk of future subsequent breach.

3.2 Predicting Future Breaches

A related stream of research has sought to predict the occurrence and number of records exposed in future data breaches. Early studies such as Curtin and Ayres (2008) struggled to find adequate predictive power due to sample size limitations within their breach data. Widup (2010) later used data breaches between 2005 and 2009 to predict an imminent increase in the number of firms experiencing breaches by internal employees accessing data without authorization or maliciously using organizational resources (i.e., insider threat). Edwards et al. (2016) utilized a Bayesian approach to identify time series trends, which informed their prediction that the coming years would yield several large-scale (greater than 80,000 affected individuals) breaches.

An interesting paper by Sarabi et al. (2016) estimated a data breach risk distribution for predicting types of data breaches within certain industries and discovered a correlation between the two. However, Sarabi et al.'s data limitations are unable to account for significant changes in industry regulations (e.g., the introduction of HIPAA and the HITECH Act in the healthcare sector), capture longitudinal trends in breach type, or evaluate breaches by employees at a granular level; thereby introducing potential biases to their results. Our essay differs and extends the prior work in several ways. First, the extensive historical data used in this paper accounts for significant changes in industry regulations (e.g., the introduction of HIPAA in the healthcare sector) and also captures trends in breach type; thereby alleviating potential yearly biases. Second, we use the IS security threat taxonomy presented by Willison and Warkentin (2013) for coding the breach type, which allows for a granular investigation of both internal and external

security threats. Finally, our analyses extend beyond identifying correlations between breach and industry types to investigate changes in the tendency of firms to experience particular breach types based on their breach history.

3.3 Model Development

3.3.1 Hazard Model

We use an accelerated failure-time (AFT) hazard model to answer our first research question. We chose an AFT model rather than the commonly used Cox proportional hazard model for several reasons. The first reason is because our data contains delayed entry gaps and multiple failure points (subsequent breaches in our case) which have been shown to be better suited to the AFT model (Hosmer et al. 2008). The second reason is because we want to estimate and allow the effect of time to interact with our covariates. That is, we desire to understand how data breach notification policies affect the time duration between organizations experiencing a prior breach and a future breach. The AFT model allows us to take such durations into account when estimating our coefficients whereas the Cox proportional model separates the effect of time from the effect of the covariates by holding the effect of time constant such that the coefficient estimates for the covariates remain proportional at any point in time.

The AFT model expresses the natural logarithm of the survival time, log *t*, as the linear function

$$\log t_{ij} = \boldsymbol{x}_{ij}\beta + \mathrm{z}_{ij}$$

for ith breach observation in the jth firm. The observation time for our model is eleven years and time t_{ij} is the years between either the next breach or, if there is not a subsequent breach following breach i, it is the years until the end of the observation period (January 1, 2017). $\boldsymbol{x}_{ij}$ is

the vector of covariates[21]. β is the vector of estimated coefficients for the covariates. z_{ij} is the error with density f. Appropriate choice of the error term density is one of the most critical selections for the model. To ensure correct density fit, we calculated the Kaplan-Meier estimator of the survival function and plotted the estimation in order to determine the appropriate distributional form of the error term. We then compared the plotted estimation with the plotted survival function for the Weibull, log-normal, and gamma distributions, and found that the plot of the Kaplan-Meier estimator resembles a log-normal distributed survival function.[22] Using the lognormal distribution, we parameterized the survival function as

$$S(t) = 1 - \Phi\left\{\frac{\log(t_{ij}) - x_{ij}\beta}{\sigma}\right\}, \;\; x \geq 0; \sigma > 0$$

where $\phi(z)$ is the standard normal cumulative distribution function and the standard deviation, σ, is an ancillary parameter estimated from the data. We then use the density functions, $f(t)$ and $F(t)$, for the lognormal distribution to obtain our hazard function $h(t)$ as follows.

$$f(t) = \frac{\phi\left(\frac{\log(t_{ij}) - \mu}{\sigma}\right)}{t\sigma}$$

$$F(t) = \Phi\left(\frac{\log(t_{ij}) - \mu}{\sigma}\right)$$

$$\bar{F}(t) = 1 - F(t)$$

$$h(t) = {f(t)}\Big/{\bar{F}(t)}$$

[21] A list of the covariates we use in the model, along with descriptions, can be found in Table 1. Summary statistics for our covariates is in Table 2.

[22] We also estimated the exponential, generalized gamma, Weibull, log-logistic, and Gompertz parametric distributions within our hazard model to obtain the log likelihood and AIC performance measures. The log-normal distribution had the highest log likelihood and the lowest AIC. Thus, we are confident in our use of this distribution.

Where σ is estimated as an ancillary parameter using our data. We also introduced shared-frailty to the model by adding an unobservable, firm specific, effect α_i to the hazard function such that it takes the form

$$h_{ij}(t|\alpha_i) = \alpha_i h(t|x_{ij})$$

The frailty, α_i, is a random positive number with a gamma distribution and is assumed to have mean 1 and variance θ. Since α_i is group specific, the value varies across firms but records involving the same firm use the same α_i value. There are two further assumptions needed to incorporate α_i into our hazard model. First, differences in organizations experiencing data breaches may be caused by unobservable factors such as an organization's managerial differences. Executives or managers within the organization may direct employee efforts toward initiatives unrelated to information security or misidentify information security needs. Second, organizations are independent of one another and handle information security in different ways. An organization's executives may place a higher priority on information security with many directives, but an identical organization under different leadership may place a lower priority on information security with fewer security directives. Having α_i in the model, however, allows us to account for this unobserved heterogeneity so that we avoid this bias.

Since shared frailty is common to a group of observations, the data becomes organized as $i = 1, \dots, n$ groups with the ith group containing $j = 1, \dots, n_i$ breach observations. The log likelihood is then the sum of the log likelihoods for each group. We define $D_i = \sum_j d_{ij}$ as the number of subsequent breaches in a ith group. The log conditional likelihood function with shared gamma frailty can then be written as

$$\log L = \sum \log L_i = \sum \left\{ D_i \log h_{ij}\left(t_{ij}\right) - \left(\frac{1}{\theta} + D_i\right) \log \left\{ 1 - \theta \sum_{j=1}^{n_i} \log \frac{S_{ij}(t_{ij})}{S_{ij}(t_{0ij})} \right\} + D_i \log \theta + \log \Gamma\left(\frac{1}{\theta} + D_i\right) - \log \Gamma\left(\frac{1}{\theta}\right) \right\},$$

where θ is estimated jointly with β and σ. The data are right-censored due to the firms continued risk of experiencing breaches beyond the observation time window (e.g., recent data breaches, such as those in 2017).

3.3.2 Multinomial Logit

We answer our second research question *"Are particular industries more susceptible to certain types of breaches?"* with a multinomial logit model. Multinomial logit models are widely used across economics, marketing, and operations management for studying changes in the likelihood of an outcome in the case of unordered categories. This model allows us to answer our second research question by estimating changes in the average probability of a type of breach according to firm's industry. For our analysis, consider the following multinomial logit model in which the probability that the breach type for the jth breach is equal to the ith breach type.

$$\Pr(y_j = i) = \begin{cases} \dfrac{1}{1 + \sum_{m=2}^{k} \exp(x_j \beta_m)}, & if \quad i = 1 \\ \dfrac{\exp(x_j \beta_m)}{1 + \sum_{m=2}^{k} \exp(x_j)}, & if \quad i > 1 \end{cases}$$

In the model, k is an unordered categorical outcome (i.e., type of breach that occurred) with a reference outcome of 1. We have that x_j represents the firm's industry type covariates and β_m is the estimated coefficient for the corresponding covariate. We use the same breach and industry types from hazard model for consistency.

To answer our third research question *"Does the breach type increase the likelihood of experiencing a similar breach type in the future?"* we use another multinomial logit model but instead estimate the changes in the average probability of a future type of breach according to the firm's previous type of breach. Considering the model derivation above, we have the probability that the subsequent breach type for the jth breach is equal to the ith type of breach. In this model, k is an unordered categorical outcome (i.e., the subsequent breach type that occurs) with a reference outcome of 1. The covariates x_j represent the prior breach type a firm experienced and the estimated coefficients for our covariates is β_m.

3.4 Data

3.4.1 Privacy Rights Clearinghouse

PRC is a non-profit organization that records data breach information from government agencies and news media websites when they are made public. PRC has web crawlers continuously searching outlets that post data breaches in an attempt to provide an up to date look at the data breach landscape. The organization has been gathering breach information since 2005 and has grown into one of the largest and most comprehensive breach data sets available (Edwards et al. 2016). For our analysis, we use data breaches recorded between 2005 and 2016, providing information on 6,088 breaches and include the following fields: *the date in which the breach was recorded, the breached firm's name, industry type, detailed description of the breach, and the number of records compromised in the breach*.

The first date in which the breach was recorded by PRC is often times the date in which the breach was made public. The organization's web crawlers collect the breach information and store it in their database as soon as it becomes available. There are several reasons why the date

tends to be the publication date and not the actual date of the breach. The first reason is because firms may not disclose the actual date that the breach occurred. In some instances, firms are not required to provide the exact date so long as they notify that a breach occurred. The second reason is because the firm's investigation into the breach may not provide an exact date of the event. In these cases, the date the breach is made public is the most information regarding the date of the breach. However, the date used in the field becomes the actual breach date when that information is provided. The date field is used to create the time, t_{ij}, variable in the hazard model. We calculate the variable by first taking the difference in days between the prior breach and the subsequent breach or the prior breach and the end of the observation period. The difference in days is then divided by 365.25 to convert the difference to years. To illustrate, consider the following example. Microsoft experienced three breach incidents, one of which was on February 22, 2013, another on December 26, 2014, and finally on April 3, 2015. Then $t_{1Microsoft} = 1.839836$ (difference in years between breach one and two), $t_{2Microsoft} = 0.268309$ (difference in years between breach two and three), and $t_{3Microsoft} = 1.749487$ (difference in years between breach three and January 1, 2017).

The next field used is the name of the firm experiencing the breach. We assigned a unique identifier to each firm in order to allow grouping of data breaches to firms with multiple breach observations. In addition, we used the combination of the breach date and the firm name to create a binary indicator variable named *event*. The *event* variable took a value of 1 if there was a subsequent data breach following a prior data breach at the same firm. The *event* variable took a value of 0 if there was not a subsequent data breach. Hazard models require an *event* variable to denote the occurrence of an outcome of interest (i.e., a subsequent data breach occurs) and the time until that outcome.

The third field is the firm's industry type. There are seven unique industry types identified by PRC: government-affiliated organizations, educational institutions, healthcare related organizations, retail businesses, financial services firms, non-profit organizations, and non-retail businesses (e.g., Internet firms). A given firm is assigned by PRC to a single industry that it most closely represents. For example, a state university is classified as an educational institution. We created a binary indicator variable for each of the seven industry types. These indicator variables are used as control variables in the hazard model and covariates in the first multinomial logit model.

The fourth field from the PRC data set is the detailed description of the breach. PRC includes a description of the breach when the firm discloses information regarding the event. The descriptions vary in their specificity, but one can decipher how the information was compromised. Although PRC includes its own breach type designation, we seek to provide a more granular identification of the breach event. The primary motivator for this is because many of the breaches are caused by internal employees, which is not clearly identified with the PRC breach type categorization. Additionally, we wanted to use a well-established taxonomy for identifying types of information security breaches. Therefore, breach types are coded using the taxonomy provided in Willison and Warkentin (2013). Their taxonomy expands an earlier mapping by Loch et al. (1992) to include a continuum of internal and external sources of threat with greater attention on internal sources.

The information security violators in the taxonomy include internal humans, internal nonhumans, external humans, and external nonhumans. Internal humans are employees or individuals that have been given access to sensitive data within the firm. The types of breaches an internal human violator may cause can be regarded as either non-volitional compromise,

volitional noncompliance, or malicious computer abuse. Non-volitional compromises can be described as when an employee unintentionally compromises sensitive data. Volitional noncompliance is when an employee disregards information security safeguards or policies but without the intent to compromise data. Malicious computer abuse is the theft or corruption of sensitive data with the intent to sell or harm. Internal nonhuman violators are internal events that may compromise data but are not linked to a human insider (e.g., hardware failure and printer errors). External human violators are people who are not employees and do not have access to sensitive data within a firm but maliciously infiltrate a firm's information system and compromise data. External human violators may gain access to the data through technological (e.g., hackers) or nontechnological (e.g., device theft) means. External nonhuman violators are typically outside programs such as malware that infiltrate a firm's information system and compromise data. We coded the observation as one of these breach types based on the description of the reported breach. We also created an additional indicator labelled as unknown, which was used for coding breach observations in which the firm did not disclose information regarding the breach. The indicator variables for breach type are used as control variables in the hazard model, the outcome variable in both multinomial logit models, and as the covariates for the second multinomial logit model.

The last field we use from the PRC data set is the number of records compromised in the breach. The number of records ranged anywhere from zero to in the billions. If a firm was unaware of the number of compromised records or did not disclose number, the field takes the value of "Unknown." Because we were unable to distinguish whether "Unknown" was actually a few or many records, we created a binary indicator to represent whether the firm reported the

number of compromised records. The indicator variable is used as a control variable in the hazard model.

3.4.2 Identity Theft Laws

All states in the U.S. have implemented a form of identity theft law[23] prior to the start of PRC's data collection, but similar to data breach notification laws they differ from state to state. Identity theft is the use of another individual's sensitive information (e.g., Social Security number, credit card number, medical insurance ID number, etc.) with the intention to commit fraud. The distinction between identity theft laws and data breach laws is that identity theft laws are at the individual level and data breach laws are, for the most part, at the firm level. A state's identity theft law specifies what constitutes identify theft, the penalties for committing identity theft, and additional provisions to further protect people's information. Since identity theft may occur as a result of some data breach types such as internal employees maliciously abusing data or external hackers attempting to compromise firm data, we must control for the impact of identity theft laws on subsequent data breaches.

The controls implemented in the hazard model include the penalty type associated with identity theft, whether or not restitution may be filed against a guilty party, and the trafficking of data from one party to another. The penalties for identity theft include degrees of misdemeanors or felonies. As previously mentioned, misdemeanors are crimes typically punishable by one year of jail time. Felonies, on the other hand, are the most serious type of crimes and punishable for longer than one year of prison. Next, parties convicted of identity theft may be ordered to make restitution for any financial losses by the victims such as lost wages, attorney's fees, and credit

[23] The website we obtain our information on identity theft laws by state is http://www.ncsl.org/research/financial-services-and-commerce/identity-theft-state-statutes.aspx.

corrections. Lastly, parties may be found guilty of identity theft through trafficking of sensitive information. The act of identity theft specifies an intent to commit fraud, which does not necessarily fall into the distribution of sensitive information. Some states however, specify that trafficking sensitive information to other parties who intend to commit fraud is just as culpable. For our analysis, we create four indicator variables for each of these aspects of identity theft laws and coded them based on the information provided by each state's law. We supplemented the PRC data set by pairing the identity theft indicators with breaches that occurred based on the state.

3.4.3 Data Breach Notification Policies

Finally, we dissect data breach notification laws by each state into their individual policies.[24] Policies included the presence of a data breach notification law, risk of harm to the consumer, notification of paper records breach, allowing firms to face criminal penalties, and the overseeing body to which firms must disclose breach information (state attorney general, consumer reporting agency, or no notification to a governing body). States may have unique policies according to their individual needs or requirements, but we chose these policies because of their prevalence across all notification laws. The policy information was incorporated into the PRC data set according to the state in which the breach occurred and the date of the incident by using indicator variables.[25] If the breach incident occurred prior to a state's implementation of a

[24] We gathered policy information from Hennessy et al. (2018)'s chart of data breach notification policies and date of implementation by Perkins Cole (https://www.perkinscoie.com/en/news-insights/security-breach-notification-chart.html).

[25]We must note that data breach notification is dependent upon the individual's residence and not the site of the breached firm. In other words, breached firms must abide by the state notification law according to where each affected individual resides. Unfortunately, without information on each individual's residence, it is infeasible to use all of the state notification laws associated with a breach. However, we argue that using the state notification law for the location of the breached firm is useful and meaningful because it is reasonable to assume that a firm will comply with those laws in the event of a breach. For example, a hospital breach is likely to affect individuals within the

notification law, the notification law indicator and all policy fields were set to zero because a law had not established the policies. Table 4 provides a list of the variables used in this paper as well as descriptions for the variables.

Table 5. Variable Descriptions

Variables	**Description**
Control Variables	
Industry Type	
Government	The organization is a governmental organization (e.g., Department of Health and Human Services, U.S. Army, etc.).
Education	The organization is an educational institution (e.g., University of Arizona).
Medical	The organization is a medical organization (e.g., Arizona Medical Center).
Financial	The organization is a financial institution (e.g., Wells Fargo).
Business (Retail)	The organization is a retail business (e.g., Target, Costco, etc.).
Business (Other)	The organization is not a retail business but operates for profit and has paying customers (e.g., Netflix, Google, etc.).
Non-profit	The organization is a non-profit organization (e.g., Amnesty International).
Breach Type	
Internal Nonhuman	Internal breach events that compromise data but are not caused by an employee (e.g., hardware failure and printer error).
Internal Human – Malicious	Theft of corruption of sensitive data with an intent to sell or harm by an employee (e.g., insider threat)
Internal Human – Volitional	Sensitive data is compromised as a result of an employee not complying with information security safeguards or policies but without malintent.
Internal Human – Nonvolitional	Sensitive data is compromised unintentionally by an employee (e.g., employee mistakes).
External Nonhuman	Outside programs that infiltrate a firm's information system and compromise sensitive data (e.g., malware).
External Human – Technology	External party that uses a technology to penetrate a firm's information system and compromise data (e.g., system hacking).
External Human – Nontechnology	External party that does not use technology to obtains sensitive data (e.g., laptop computer theft).
Unknown	The organization did not provide a description of the incident or little definitive information is known.
Identity Theft Law	
Felony Charge	Parties found guilty of identity theft face felony punishments.
Misdemeanor Charge	Parties found guilty of identity theft face misdemeanor punishments.
Restitution	Parties found guilty of identity theft must pay victims' fees and losses.
Identity Trafficking	Parties distributing stolen identities to another party may also be convicted of identity theft.
Other Variables	
Records Disclosed	The organization has disclosed the number of records compromised in the incident.
Publicly Traded	Firm is publicly traded in the United States.
Independent Variables	
Breach Notification Law	
Notification Law Indicator	The state in which the breach occurred has enacted a data breach notification law.

surrounding area with few exceptions. Furthermore, breaches at larger organizations are likely to affect individuals within the same state because of the national customer base.

Risk of Harm	The organization is not required by law to notify the public of a data breach if the organization believes that the breach has not and will not cause harm to the individuals.
Paper Records	The form of information covered for a data breach includes paper records.
Criminal Penalties	The organization may incur criminal penalties as a result of the breach.
Attorney General Notification	Organizations must provide notification of a breach to the state Attorney General's office.
CRA Notification	Organizations must provide notification of a breach to nationwide Consumer Reporting Agencies (CRA) such as a credit bureau.
No Notification Requirement	Organizations are not required to provide notification of a breach to any government or consumer agencies.

3.5 Data Analysis and Results

The following sections present our analyses and results. Section 5.1 addresses the analysis and results for the first research question. Sections 5.2 and 5.3 present the analyses and results for research questions two and three respectively. Summary statistics are in Table 5.

Table 6. Summary Statistics

Variables	**Total Records**		**Subsequent Breach Records**	
	Obs.	**Percent Yes**	**Obs.**	**Percent Yes**
Control Variables				
Industry Type				
Government	6088	11.81	1231	8.93
Education	6088	12.60	1231	20.63
Medical	6088	41.22	1231	38.34
Financial	6088	10.48	1231	13.08
Business (Retail)	6088	8.85	1231	8.69
Business (Other)	6088	13.50	1231	9.50
Non-profit	6088	1.48	1231	1.00
Breach Type				
Internal Nonhuman	6088	3.15	1231	2.92
Internal Human – Malicious	6088	11.33	1231	14.22
Internal Human – Volitional	6088	7.38	1231	5.12
Internal Human – Nonvolitional	6088	16.90	1231	19.82
External Nonhuman	6088	4.00	1231	3.74
External Human – Technology	6088	25.54	1231	19.17
External Human – Nontechnology	6088	23.98	1231	25.99
Unknown	6088	7.65	1231	9.02
Identity Theft Law				
Misdemeanor Charge	6088	22.73	1231	22.99
Felony Charge	6088	30.37	1231	29.16
Restitution	6088	71.71	1231	71.24
Trafficking Identities	6088	36.02	1231	36.80
Other Variables				
Records Disclosed	6088	74.39	1231	77.09
Publicly Traded Firm	6088	8.44	1231	24.37
Independent Variables				
Breach Notification Law				

Notification Law Indicator	6088	86.56		1231	80.58
Risk of Harm	6088	42.20		1231	36.72
Paper Records	6088	16.72		1231	16.33
Criminal Penalties	6088	7.79		1231	8.45
Attorney General Notification	6088	46.39		1231	48.50
CRA Notification	6088	52.99		1231	46.06
No Notification Requirement	6088	10.64		1912	11.70
Dependent Variables					
	Obs.	**Mean**	**St. Dev**	**Min**	**Max**
$\log t$	6088	0.980	1.273	-5.901	2.483

3.5.1 What impacts the survival of a firm that experiences repeat breaches?

We have 6,088 data breach observations in our data set. Among the observations, there are 4,857 firms and 1,231 subsequent breaches. We include all control and notification policy variables in the estimated model. Table 6 displays the results from our estimation.

The coefficients in an AFT hazard model are interpreted by first looking at the direction of the effect. Negative coefficients represent an increase in survival time; thereby lowering the risk of future subsequent breach as it takes longer for a subsequent breach to occur. Positive coefficients represent a decrease in survival time; thereby increasing the risk of future subsequent breach as it takes less time for a subsequent breach to occur.

3.5.1.1 Control Variables

We begin our discussion of the results with the control variables. The first set of controls is firm industry type. For the industry type, we establish the medical industry type as the baseline for comparison. Interestingly, all other industry types varied significantly from the medical industry. We highlight that educational institutions were the only industry type to have significantly less risk of future subsequent breach compared to the medical industry with $\beta = -0.975, p = 0.000$. To our surprise, non-retail businesses and non-profit organizations are

at the highest risk of future subsequent breach compared to the medical industry with $\beta = 1.342, p = 0.000$ and $\beta = 1.322, p = 0.031$ respectively.

The second set of controls is breach type. For the breach type, we establish nontechnological breaches by an external party as the baseline for comparison. The two breach types which significantly increased the risk of future subsequent breach, compared to nontechnological breaches by an external party, are volitional breaches by an internal employee and technological breaches by an external party with $\beta = 0.594, p = 0.41$ and $\beta = 0.538, p = 0.007$ respectively. No other breach types significantly affect the risk of future subsequent breach compared to nontechnological breaches by an external employee.

The third set of controls is the characteristics within a state's identity theft laws. Among these characteristics, classifying identity theft as a felony offense ($\beta = 0.475, p = 0.020$) or as a misdemeanor offense ($\beta = 0.572, p = 0.079$) increases the risk of future subsequent breach. Requiring a party guilty of identity theft to pay restitution to victims and including the trafficking of sensitive information to parties committing fraud as identity theft do not significantly affect the risk of future subsequent breach.

The fourth set of controls includes the indicator for whether the number of records affected in the breach was disclosed and the indicator for whether the firm is publicly traded. The estimated results provide that both variables significantly reduce the risk of future subsequent breach. Specifically, the results are $\beta = -0.566, p = 0.000$ for disclosing the number of records compromised in the breach and $\beta = -3.640, p = 0.000$ for firms that are publicly traded.

3.5.1.2 Hypothesized Variables

Moving to our hypotheses, the estimated coefficients from our model indicate that three hypotheses are supported, three hypotheses are not supported, and one hypothesis is not supported with the effect in the opposite direction of our prediction. Allowing firms to not disclose breach information if consumers are not at immediate risk of harm marginally increases, $\beta = 0.347, p = 0.086$, the risk of future subsequent breach (*Hypothesis 2 – Supported*). Enabling consumers to file criminal charges against the guilty party responsible for the breach significantly reduces, $\beta = -0.600, p = 0.030$, the risk of future subsequent breach (*Hypothesis 4 – Supported*). Requiring firms to disclose breach information to a state attorney general marginally reduces, $\beta = -0.409, p = 0.070$, the risk of future subsequent breach (*Hypothesis 5 – Supported*). Requiring firms to disclose breaches involving paper records, having firms provide breach information to a consumer report agency, and allowing firms to not disclose breach information to an overseeing body did not significantly affect the risk of future subsequent breach (*Hypotheses 3, 6, and 7 – Not Supported*). Lastly, the presence of a data breach notification law significantly increases, $\beta = 0.963, p = 0.006$, the risk of future subsequent breach (*Hypothesis 1 – Not Supported, Opposite Direction*).

The σ value in our estimated model was significant ($p = 0.000$), indicating that the log-normal distribution was appropriate for our estimation. The θ value in our estimated model was significant ($p = 0.000$), indicating the presence of unobserved heterogeneity.

Table 7. Log-normal Hazard Model	
Variables	
Notification Law Indicator	0.962** (0.353)
Risk of Harm	0.350† (0.202)
Paper Records Notification	-0.187 (0.246)
Criminal Penalties	-0.600* (0.276)
Attorney General Notification	-0.411† (0.226)

CRA Notification	0.254 (0.248)
No Notification Requirement	-0.394 (0.326)
Constant	4.120 (2.928)
Log(σ)	0.988** (0.025)
Log(θ)	1.493** (0.076)
Obs.	6088
Organizations	4857
Number of Failures	1231
Log Likelihood	-3725.360
x^2	317.480**
Controls	Yes

†$p \leq 0.10$, *$p \leq 0.05$, **$p \leq 0.01$

3.5.2 Are particular industries more susceptible to certain types of breaches?

In the first multinomial logit model, we establish the medical industry as the relative industry type and nontechnological breaches by an external party as the relative breach type. Table 7 displays the average marginal effects for the estimated coefficients.

Our results reveal interesting changes in the likelihood of a particular type of breach according to the firm's industry. First, we find that financial institutions are more likely to experience breaches by employees maliciously compromising their data. Non-retail businesses are more likely to experience technological breaches by an external party or malware affecting an information system. Retail businesses are more likely to be breached by employees maliciously compromising data, technological breaches by an external party or malware affecting an information system. Educational institutions are more likely to experience a breach caused by an employee making a mistake or a technological breach by an external party. Government agencies are more likely to be breached by an employee than an external threat as all four internal threat variables are positive and significant while external threats are not significant. Surprisingly, non-profit organizations are less likely to be breached by an employee

mistake but may be susceptible to the remaining breach types. Overall, we discover evidence of clear differences in breach types affecting some industries but not others.

Table 8. Multinomial Logit by Industry Type

Variables	Finance	Business (Other)	Retail	Education	Government	Non-Profit
Internal_Nonhuman	0.037 (0.021)	-0.017 (0.031)	-0.019 (0.032)	0.017 (0.025)	0.098** (0.020)	-0.022 (0.015)
Internal Human – Malicious	0.061** (0.012)	0.009 (0.017)	0.103** (0.013)	-0/117** (0.022)	0.028† (0.015)	-0.004 (0.005)
Internal Human - Volitional	0.011 (0.016)	0.014 (0.019)	0.036* (0.017)	-0.046* (0.021)	0.066** (0.015)	-0.012 (0.007)
Internal Human – Nonvolitional	-0.003 (0.013)	-0.025 (0.016)	-0.005 (0.013)	0.090** (0.012)	0.080** (0.012)	-0.014* (0.006)
External Nonhuman	-0.009 (0.022)	0.077** (0.020)	0.092** (0.017)	0.034 (0.021)	0.004 (0.023)	0.010 (0.009)
External Human – Technology	-0.003 (0.011)	0.135** (0.011)	0.098** (0.011)	0.043** (0.011)	-0.008 (0.012)	-0.003 (0.003)
Unknown	0.004 (1.302)	-0.102 (1.817)	0.037 (1.154)	-0.126 (1.505)	-0.040 (1.332)	-0.227 (11.728)
Log Likelihood	-9477.813					
LR x^2	1303.500**					
N	6088					

$\dagger p \leq 0.10$, $*p \leq 0.05$, $**p \leq 0.01$

3.5.3 Does the breach type affect the risk of experiencing a particular repeat breach?

In the second multinomial logit model, we establish nontechnological breaches by an external party as the reference breach type. Table 8 displays the average marginal effects for the estimated coefficients.

The results from our model suggest a relationship between a future subsequent breach type and prior breach type. Specifically, we find that, with the exception of hardware failure and employees disregarding information security policies, future breaches are more likely to be of the same breach type as a previous breach at the same firm. For instance, the average marginal effect for an employee maliciously compromising data in a prior breach indicates a 15% greater chance

of experiencing an employee maliciously compromising data in a subsequent breach. Interestingly, we see a decrease in the likelihood of a subsequent breach by employees disregarding information security policy when the prior is of the same breach type. Hardware failure as a prior breach type does not affect the likelihood of any subsequent breach type.

Table 9. Multinomial Logit for Next Data Breach Type

Variables	Internal Nonhuman	Internal Human - Malicious	Internal Human - Volitional	Internal Human - Nonvolitional	External Nonhuman	External Human - Technology	Unknown
Internal Nonhuman	0.024 (0.028)	-0.172 (0.118)	-0.049 (0.051)	-0.024 (0.078)	0.175* (0.068)	0.024 (0.027)	0.086* (0.044)
Internal Human – Malicious	-0.002 (0.019)	0.152** (0.027)	-0.033 (0.021)	-0.010 (0.037)	0.023 (0.043)	-0.029 (0.027)	0.030 (0.027)
Internal Human – Volitional	0.000 (0.028)	0.013 (0.048)	0.016 (0.022)	0.067 (0.049)	-0.032 (0.067)	0.004 (0.027)	0.056 (0.036)
Internal Human – Nonvolitional	0.012 (0.016)	0.016 (0.029)	-0.041* (0.020)	0.068* (0.030)	0.103** (0.035)	-0.021 (0.020)	0.012 (0.026)
External Nonhuman	-0.014 (0.038)	-0.029 (0.063)	-0.030 (0.037)	-0.032 (0.066)	0.170** (0.058)	0.046* (0.020)	-0.013 (0.054)
External Human - Technology	0.018 (0.015)	-0.016 (0.032)	-0.030 (0.019)	-0.108** (0.039)	0.281** (0.030)	0.025† (0.014)	-0.041 (0.032)
Unknown	0.009 (0.021)	0.035 (0.038)	-0.037 (0.027)	-0.037 (0.046)	-0.064 (0.058)	0.032† (0.018)	0.126** (0.026)
Log Likelihood	-2200.431						
LR x^2	261.140**						
N	1231						

†$p \leq 0.10$, *$p \leq 0.05$, **$p \leq 0.01$

3.6 Discussion and Implications

Data breach notification laws yield a deterrence effect. However, the effect manifests through specific policies within the law and not through the mere presence of the notification law. We find evidence that there are three policies which affect the degree of deterrence. The first policy is the ability to file criminal charges against the party responsible for the breach if said party had malicious intent for the compromised data. Our results indicate that implementing

the criminal charges policy leads to a reduction in firms' risk of future subsequent breach. We can decompose the change in risk (estimated coefficient) into the attributable implementation of the policy across multiple states through the use of a counterfactual. We begin with the parameterization of the hazard model into the following linear function.

$$\log t_{ij} = \boldsymbol{x}_{ij}\beta + \mathrm{z}_{ij}$$

We may then decompose the change in risk as:

$$\log t_{ij} = \delta criminal\_policy_{ij} + \beta x_{ij} + z_{ij}$$

$$\log t_{1j} - \log t_{0j} = \delta criminal\_policy_{1j}$$

$$t_{1j} = t_{0j} e^{\delta criminal\_policy_{1j}}$$

We can now use the model to construct a counterfactual where t_{1j} is the duration with breaching a firm being a criminal offense while $t_{0j}$is the duration in which breaching is not a criminal offense. In particular, we are interested in what happens to the time between breaches if breaching data becomes a criminal offense. We first calculate the average number of years between subsequent breaches for states that have not implemented the criminal charge policy, which is $t_{0j} = 4.3046$ years. We then use the estimated coefficient for the presence of the criminal charge policy as $e^{-0.6001}$ because $criminal\_policy_{1j} = 1$. Calculating t_{1j} we find that implementing the criminal charge policy leads to an additional 2.3622 years between a prior breach and a subsequent breach. Specifically, the average duration between breaches for states without the criminal charge policy is 4 years and 4 months and the average duration between breaches for states with the criminal charge policy is 6 years and 8 months.

Using the same linear decomposition, we calculate the counterfactual for the remaining two policies that significantly affected deterrence. The next policy counterfactual we calculate is the risk of harm policy. We calculate the average number of years between subsequent breaches

for states that have not implemented the risk of harm policy to be $t_{0j} = 4.4165$ years. We use the estimated coefficient for the presence of the risk of harm policy as $e^{0.3500}$ because $risk\_of\_harm_{1j} = 1$. Calculating t_{1j} we find that implementing the risk of harm policy leads to a decrease of 1.8506 years between a prior breach and a subsequent breach. Specifically, the average duration between breaches for states without the risk of harm policy is 4 years and 5 months and the average duration between breaches for states with the risk of harm policy is 2 years and 7 months.

The last policy counterfactual we calculate is the requirement that firms disclose breach information to a state attorney general. We calculate the average number of days between subsequent breaches for states that have not implemented the disclosure to an attorney general policy to be $t_{0j} = 4.5618$ years. We use the estimated coefficient for the presence of the disclosure to an attorney general policy as $e^{-0.4109}$ because $attorney\_general_{1j} = 1$. Calculating t_{1j} we find that implementing the disclosure to an attorney general policy leads to an additional 3.0247 years between a prior breach and a subsequent breach. Specifically, the average duration between breaches for states without the disclosure to an attorney general policy is 4 years and 7 month and the average duration between breaches for states with the disclosure to an attorney general policy is 7 years and 7 months.

Earlier studies have shown increased spending on information and operational security following a breach and therefore we hypothesized that the presence of a data breach notification law will reduce the risk of future subsequent breach. Although our results do not support our hypothesis, the significance of the effect in the opposite direction (i.e., having a data breach notification law increases the risk of future subsequent breach) suggests that firms are complying and thereby increasing the predicted risk. While it is reassuring that organizations appear to be

complying with regulations, breach notifications may lose their impact (e.g., promoting organizational change and increased spending on information security) over time. Growing visibility and consumer awareness of data breaches may lead to a sense of inevitability among consumers, weakening their reaction to breach notifications and holding firms less accountable to the event.

Among the types of breaches, our results demonstrate that system hacks and employee mistakes present the greatest risk of a future subsequent breach. A possible explanation for the greater risk of system hacks is the growth in the tool sets available to current and aspiring hackers. The hacker community appears to be shifting toward a broader audience as the opportunity to learn and perform small system hacks has increased with simple, easy to download tools that walk novices through step-by-step. Novice hackers are then utilizing the hacking tools to gain experience and test many different organizations information systems. Security professionals must also understand that hackers are members of a greater hacking community that is willing to share knowledge with one another in order to better themselves and the collective group. It is likely that hackers share their information system exploits within the community and draws novice hacker's attention to vulnerable organizations. Firms must be persistent and continuously revisit their information security policies and procedures as well as training employees.

The results from our multinomial logit analyses provide evidence of relationships between breach and industry types as well as the increased likelihood of similar breach types as subsequent breaches. For example, financial institutions are more likely to be breached by employee-related insider threats and their subsequent breaches will continue to be employee-related insider threats. From a managerial perspective, the results in this paper offer evidence for

investing in targeted information security to address common vulnerabilities within the firm's industry. In addition, our results suggest that management may significantly lower their risk of experiencing a data breach by focusing their efforts on relevant breach types associated with its industry. For instance, online businesses may better allocate information security resources through system intrusion detection and prevention as opposed to training employees in handling paper records.

The insights gained from our analyses offer interesting and actionable implications for information security and privacy. Beginning with the impact to the risk of future breach, we discovered that educational institutions are at the lowest risk of future subsequent breach while government agencies and non-profit organizations are at the highest risk of future subsequent breach. Higher education institutions such as colleges and universities have received considerable attention over recent years to improve safeguarding their data. This has proved to be a difficult task as summarized in a statement by University of Maryland's president, in which he explains: "Security in a university is very different than the private sector because we are an open institution. In the private sector, you can centralize cyber security to a specific information system. You cannot do that at a university with the many points of access (multiple information systems and networks within each college and department) that promote the free flow of information. So, we have to find that proper balance between security and access" (Harris and Hammargren 2016). Federal regulations in the United States (e.g., the Family Education Rights & Privacy Act) have helped guide schools in protecting and handling their data through amendments that lay out specific rules for all schools receiving federal funding. The U.S. Department of Education has even published its own Data Breach Response Checklist to promote educational institutions toward improving their data privacy, confidentiality, and

security practices by clearly outlining best practices for the industry (Bathon 2013). Institutions at all educational levels appear to be benefiting from these guidelines and practices as they are experiencing fewer subsequent breaches than other industries.

The higher risk of subsequent breach among non-retail businesses and non-profits was somewhat unexpected. Non-retail businesses contain the largest volume and breadth of consumer data. It is for this very reason that one might expect non-retail businesses to possess the most stringent security measures. It is also the reason why it garners the most attention among cybercriminals. Compromising the largest amounts of data offers the greatest bragging rights among cybercriminal peers and the greatest monetary gain. It is to this end that once a breach within a non-retail business becomes publicized it will draw more attention from cybercriminals looking to exploit the vulnerabilities before they can be resolved.

Non-profit agencies may be easily penetrated because some of them lack the necessary monetary funds, employee skill sets, or security technologies to protect themselves. Non-profits must closely monitor their investments outside of the fundraising cause in order to appease donors because donors may prefer to offer their donations toward a cause and not daily operations. Yet, security investments at non-profit organizations are essential because they record and maintain personal information on donors such as names, addresses, phone numbers, and credit card numbers; and donors must feel their personal information is safe within the organization otherwise he or she will make less frequent or smaller donations (Insureon 2015). Such delicate security interests significantly slow the process of identifying the cause of a data breach as well as resolving the issue; thereby increasing the non-profit's window of vulnerability to additional breaches (Schaffhauser 2017). Non-retail businesses and non-profit organizations may lower their risk of future breaches by taking proactive information security measures that

have been shown to benefit firms in other industries such as appointing a Chief Information Security Officer (CISO), frequent mandatory information security training for employees, and formalizing a written information security program for safeguarding customers' personal information (Harris and Hammargren 2016).

As with all research, our essay has limitations. The first limitation is our lack of extensive covariates to control for individual firm characteristics such as organizational size and the number of employees within each firm. Unfortunately, a majority of the records within our data set are private businesses and we cannot retrieve further information regarding the organization. Luckily, we remain capable of estimating a full AFT hazard model because the AFT model is robust to missing covariates in the mode. The second limitation is the usage of PRC's breach descriptions and organization labels. With this limitation, we are unable to be certain that we have a complete story for the analysis. A small portion of the data set (approximately 7.65% of the data) has Unknown listed as the breach type. Therefore, the number of breach types could be slightly biased downward, and it is possible but not likely that they are explaining some of the null effects. The third limitation is our reliance on truthful breach disclosures by organizations. Some firms may not disclose a breach if not required by law, or may not disclose due to poor ethical practices. Therefore, our data set is limited to those organizations who reported a breach, but we are unable to verify truthfulness in the organization's disclosure reporting. However, we are confident that these limitations do not weaken our analysis or alter the essay's findings.

4. Essay 3 - Impact of Data Breaches on the Benefits of Healthcare Information Technology

Patients are expressing that health information privacy is a priority of concern now that EMRs have been adopted in more than eighty percent of hospitals across the U.S. (Anderson and Agarwal 2011; Meingast et al. 2006; Kruze et al. 2017; Hiller et al. 2011; Kim et al. 2015). The Privacy Rule and Security Rule for the Health Information Portability and Accountability Act (HIPAA) were the first means to address the growing need for information privacy. The HIPAA Privacy Rule provides a formal definition for protected health information (PHI) – any information held by a healthcare organization containing personally identifiable information associated with health status, provision of health care, or health care payment. The HIPAA Security Rule requires administrative, physical, and technical safeguards be adopted to protect patients' PHI. However, information privacy and security within healthcare is an increasingly complex topic due to the digitization of healthcare data; which generates a larger and more vulnerable repository of data to compromise (Kwon and Johnson 2014b). Legal scholars have identified several deficiencies in the HIPAA Security Rule. One deficiency is that the implementation of data safeguards is at the discretion of the hospital and the safeguard requirements lack detail and specificity (Hoffman and Podgurski 2007), which offers little guidance and may lead hospital management to implement minimal security efforts (Appari and Johnson 2010). Huang et al. (2014) finds support for this notion through a survey finding that indicates over one-quarter of the hospitals sampled do not employ information security staff.

Another deficiency of the HIPAA Security Rule is its lack of enforcement because the regulations had little power to sanction noncompliant hospitals (Hoffman and Podgurski 2007). Thus, the Enforcement Rule and eventually the Health Information Technology for Economic and Clinical Health (HITECH) Act were enacted to strengthen penalties for healthcare

organizations disregarding information privacy and security provisions as well as mandate that hospitals notify DHHS following a data breach. DHHS is then able to investigate hospital breaches and administer fines to hospitals that experience a breach. Thus far, the prior literature on data breaches in hospitals has focused on the financial aspects surrounding a breach. For instance, Kwon and Johnson (2013) found that external data breaches (i.e., a breach caused by an outside individual rather than a hospital employee) spurred the hospital's investment toward furthering their HIT implementation to meet government standards. Kwon and Johnson (2014a) discovered that hospitals with proactive information security efforts significantly reduced the magnitude of the breach as well as federal penalties and fines from DHHS. Furthermore, Angst et al. (2017) contributed to the literature with their findings that higher information security investments in hospitals lead to greater risk of breach; meaning that greater investment potentially flags a hospital as having information that is more valuable. Angst et al. found conflicting evidence with Kwon and Johnson (2014a) by demonstrating that hospitals with substantial IS security investment do not outperform or reduce their likelihood of breach more than hospitals with less IS security investment.

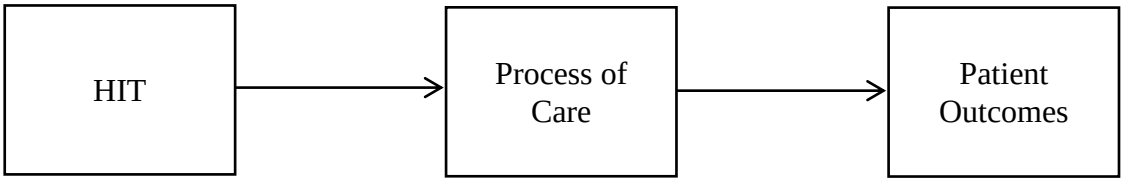
In addition to monetary consequences, hospitals typically change organizational policies at the request of DHHS's findings following their breach investigation or in an attempt to ward off an impending breach investigation. According to data provided by DHHS, over ninety percent of hospitals report organizational policy changes following a data breach. When hospitals implement policy changes of any kind, there is often disruption to nurses' and physicians' ability to care for patients (Angst et al. 2012), which may be detrimental to patient outcomes. We then ask two unique research questions, (1) *"To what extent do data breaches*

affect the hospital process of care for certain conditions?" and (2) *"To what extent do data breaches then affect patient outcomes (e.g., mortality and readmission rates)?"*

4.1 Research Model and Hypotheses

Angst et al. (2012) suggest that the direct relationship between HIT implementation and patient care is complex. Through a structure-process-outcome framework, they found evidence of a mediating effect in which HIT affects processes involved with providing patient care and leads to changes in patient outcomes. We utilize their proposed framework and extend it in several unique ways. First, we theorize that achieving meaningful use of EMR within a hospital is indicative of the hospital's HIT and impacts processes of patient care. Second, we hypothesize that a hospital data breach affects processes of patient care. Finally, we expand upon their process and outcome measures to include other common medical conditions such as heart attack and pneumonia. Figure 1 illustrates the proposed relationships from Angst et al. (2012) and Figure 2 provides our extensions to the relationships.

Figure 1. Role of HIT on Processes and Outcomes



We draw upon organizational information processing theory (OIPT) (Gattiker and Goodhue 2005) to posit that hospitals derive several overall benefits from HIT implementation. OIPT provides that organizations in complex environments with a high degree of uncertainty

seek to enhance performance and decision making by improving quality and communication of information. Premkumar et al. (2005) suggest that one strategy organizations have for achieving these improvements is to implement integrated information systems to generate efficient knowledge transfer. For instance, patients frequently experience fragmented care because they receive care from numerous physicians and specialists whom are often times in different medical groups. This leads to disjointed information which may result in incomplete medical information, prescription medication problems, and improper management of illness (Pinsonneault et al. 2017). However, HIT has the potential to lower the barrier of fragmented care for patients by improving communication between practices and managing comprehensive patient records; which in turn alters the processes throughout the healthcare value chain to provide better patient care (Lim et al. 2015; McCullough et al. 2016; Angst et al. 2011). A designation signifying that a hospital utilizes HIT and has demonstrated its ability to both improve its quality of patient information and shares that information with other healthcare providers is meaningful use attestation. Hospitals attest to meaningful use once they reach a defined set of government standards.[26]

Scholars and industry leaders believe these standards to be essential for improving patient care. Jones et al. (2014) conducted a large-scale meta-analysis on meaningful use and found that achieving meaningful use, particularly with clinical decision support and computerized physician order entry systems, led to improvements in overall patient care, processes of care and cost reductions. Specifically, they discovered that a majority of the studies reported fewer medication

[26] CMS established the standards for attesting meaningful use to promote the adoption and continued usage of EMR. CMS offers incentive payments to hospitals who achieve the minimum standards for meaningful use and further payments to hospitals that exceed the minimum standards. To establish a long term incentive plan for growing hospitals' usage of EMR, CMS classifies hospitals as meeting various stages of meaningful use with each stage having a different set of requirements that build upon one another. CMS currently has seven stages of meaningful use. Hospitals that meet the minimum standards are in Stage 1 of meaningful use and become eligible for CMS incentive payments.

errors leading to a reduction in future complications from overdosing or incorrect prescriptions. Jones et al. also found that over 70 articles reported significant process improvements; one of which demonstrated a 30% increase in adhering to surgical guidelines (Bell et al. 2010). Thus, we use meaningful use attestation as an indicator for HIT in the Angst et al. (2012) framework and hypothesize the following:

Hypothesis 1: Meaningful use attestation within a hospital will positively improve process of care.

4.1.1 Hospital Data Breach

We argue that the penalties administered by DHHS to breached hospitals act as a deterrent to future breach and promote organizational change. As shown in Essay 2, firms across all industries respond to strict sanctions by taking measures to reduce the risk of future breach. We find evidence of this in hospitals according to hospitals' breach notification to DHHS. Hospitals provide a unique information security setting because of its vulnerability to both internal and external threats (Kwon and Johnson 2014a). Internal threats are typically breaches caused by hospital staff with examples including an employee maliciously stealing patient information (i.e., insider threat) or an employee mistakenly disclosing patient information to an unauthorized party. Hospitals typically resolve breaches from internal threats by altering business processes and retraining employees on the new processes. For instance, a large number of breaches reported to the DHHS were the result of employees unintentionally disclosing patient information to an unauthorized party and, in each of these instances, hospitals responded with policy changes and employee training. External threats, on the other hand, are breaches caused

by an individual(s) who is not a member of hospital staff with examples such as burglary of hospital devices or an information systems hacker infiltrating HIT systems in the hospital. Hospitals typically resolve breaches from external threats by altering hospital safeguards such as physical security measures or improved information security systems. In the event that a breach is caused by an information systems hacker, hospitals respond by analyzing the intrusion and, if possible, correcting the vulnerability in the HIT. Hospital staff then receives training on the changes made to the HIT system. Thus, hospital data breaches of any type often result in policy change and, as mentioned previously, changes in hospital policies typically lead to disruptions in the process of care for patients. Hence, we hypothesize the following:

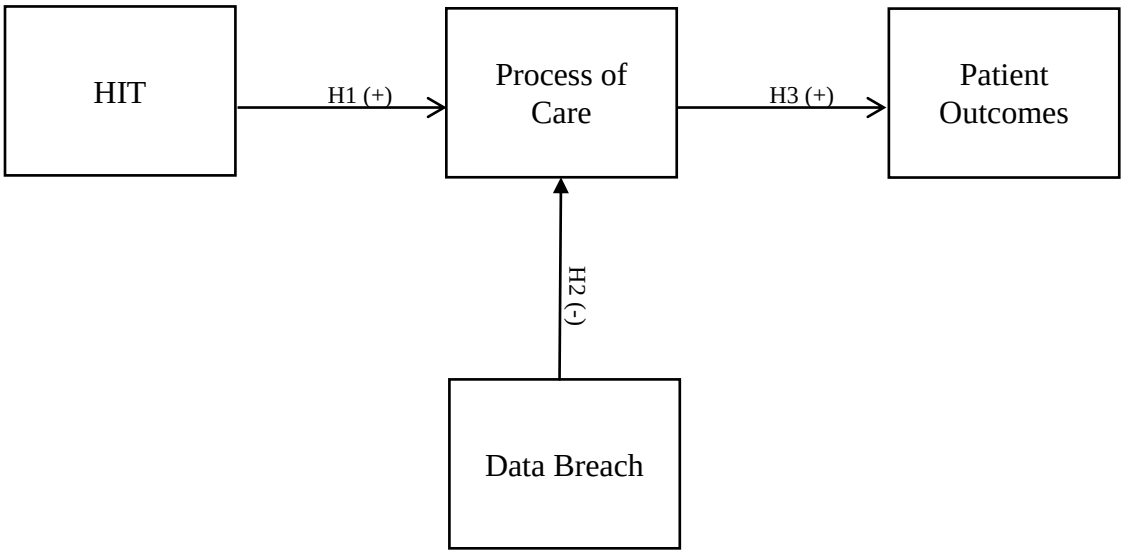
Hypothesis 2: Experiencing a hospital data breach will negatively affect process of care.

4.1.2 Process of Care

A core component of the Angst et al. (2012) framework is the argument that changes in process of care affect patient outcomes. Several studies exhibit robust findings that well executed processes of care for common conditions (e.g., heart attack, heart failure and pneumonia) reduce mortality and readmission rates for those same conditions (Fichman et al. 2011). Additionally, Williams et al. (2016) analyzed the top performing hospitals in the U.S. to study the relationship between process of care and patient outcomes. They discovered that improving the efficiency of process of care and ensuring process of care execution led to better patient outcomes. Hence, we hypothesize the following:

Hypothesis 3: Improvements to process of care will positively affect patient outcomes.

Figure 2. Hypothesized Role of HIT and Data Breach on Process of Care



4.2 Research Methodology

We use a two-stage least squares (2SLS) model to answer our research question. We selected the 2SLS model because our hypothesized framework proposes that meaningful use attestation and experiencing a data breach affect patient outcomes through the process of care. Therefore, both (1) regressing patient outcome on process of care without first accounting for data breach and meaningful use attestation and (2) regressing patient outcome on process of care, data breach, and meaningful use attestation will introduce endogeneity. We must then use data breach and meaningful use attestation as instrumental variables[27] in the estimation process. The 2SLS model corrects for endogeneity by splitting the estimation into two stages. The first stage estimates a regression of process of care on our instruments meaningful use attestation and data

[27] An instrumental variable, z, is a variable that has the property that changes in z lead to changes in x but do not change y through. We estimated a fixed effects panel data model to test for the direct effects of data breach and meaningful use attestation on patient outcome. We found that data breach is not correlated with patient outcome ($F = 1.20, p = 0.27$) but meaningful use attestation is correlated with patient outcome ($F = 21.97, p = 0.00$). Although, meaningful use attestation is correlated with patient outcome we apply the exclusion restriction (i.e., prior theory presents that the instrument is not correlated with the error term in the second stage of 2SLS), which allows us to use it as an instrumental variable for our estimation. Thus, we are confident in using both data breach and meaningful use attestation as instrumental variables in our model. We discuss the limitations of the meaningful use instrument in the Discussion section.

breach as well as other HIT controls. The predicted values of the regression are used for the second stage in which we regress patient outcome on the instrument for process of care. 2SLS is a popular method for resolving endogeneity and produces consistent estimates (Woolridge 2015).

Since we have yearly observations for each hospital, we must estimate 2SLS for a panel data model. The type of panel data model we use is a fixed effects model. We chose a fixed effects model for two reasons. First, the data set for analysis is a panel data set spanning a variety of different hospitals across the U.S. Therefore, it is unlikely to control for all hospital level characteristics that may affect our dependent variable. The second reason is the result of Hausman tests for determining whether to use a fixed effects or random effects model.[28] The Hausman test[29] analyzes the difference in estimators for both the fixed and random effects models and determines if a significant difference is present. The results from our Hausman tests for each model are $p < 0.001$ indicating that the fixed effects model is most appropriate. We then write 2SLS for a fixed effects panel data model as

$$patient\_outcome_{it} = process\_of\_care_{it}\gamma + X_{it}\beta + \mu_i + \upsilon_{it} = Z_{it}\delta + \mu_i + \upsilon_{it}$$

for the ith hospital in year t. We use X_{it} to represent the vector of exogenous variables including meaningful use attestation, hospital breach indicator, and other controls. Z_{it} is the vector of predicted observations for the process of care with instruments X_{it}. μ_i is the error between hospitals and υ_{it} is the error for the model. Since we are using a fixed effects panel data model,

[28] The fixed effects model yields a consistent estimator, but it is not always efficient. The random effects model provides an efficient estimator, but it is not always consistent. The Hausman test allows us to analyze the estimator from each model and determine the most consistent and efficient model for the data.

[29] We conducted the Hausman test in the following manner. First, we estimated a fixed effects model to test our hypotheses. Next, we stored the estimated coefficients for the fixed effects model. Third, we estimated a random effects model with the same variables. Finally, we calculate the Hausman test comparing the two estimators with a chi-square. The null hypothesis states that the estimators are not statistically different from one another. Therefore, if we reject the null hypothesis, the estimators are significantly different, and the appropriate model is the fixed effects model. If we fail to reject the null hypothesis, the appropriate model is the random effects model.

we include a transformation to control for variation between hospital characteristics (i.e., fixed effect) in the form of

$$\widetilde{w}_{it} = w_{it} - \overline{w}_i + \overline{w}$$

$$\overline{\overline{w}}_i = \frac{1}{n}\sum_{t=1}^{T_i} w_{it}$$

$$\overline{w} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T_i} w_{it}$$

with n as the number of hospitals and N is the total number of observations. The transformation removes μ_i and is obtained from the following estimated observations from the second stage.

4.3 Healthcare Data

A majority of our data comes from the Hospital Compare program funded by CMS, which has been tracking quality of care information on thousands of hospitals across the U.S. since 2005. CMS collects data from hospitals through a survey that hospitals complete and submit back to CMS on both a quarterly and annual basis. The survey measures CMS gathers are well established and frequently used throughout the medical and HIT literature.

CMS obtains data on the process of care and patient outcomes for three medical conditions: heart attack, heart failure and pneumonia. CMS focuses on these specific conditions because they are the most common conditions admitted to hospitals in the U.S. (Angst et al. 2012). The reason for collecting data on process of care and patient outcomes is to provide data for research and establish a federal incentive program. Data from the Hospital Compare program is publicly available to provide patients a glance at hospital performance and empower patients to make informed decisions on where they receive care. Medical researchers commonly study

process of care to analyze the effectiveness of different processes on treatments. Additionally, CMS uses Hospital Compare to track hospitals' progress in improving patient outcomes for heart attack, heart failure and pneumonia. Hospitals that demonstrate significant improvement in patient outcomes receive federal grants and funding.

Process of care and patient outcome are comprised of several different survey measures. Process of care consists of nine processes for treating a heart attack, three processes for treating heart failure, and two processes for treating pneumonia. Each process is independent of another. Patient outcome consists of the mortality rate and readmission rate for heart attack, heart failure and pneumonia. Table 10 provides a detailed description of the individual measures comprising process of care and patient.

Table 10. Index Variable Descriptions

Variables	Measure ID	Condition	Description
Process of Care	OP-1	Heart Attack	Median time to fibrinolysis
	OP-2	Heart Attack	Outpatients with chest pain or possible heart attack who got drugs to break up blood clots within 30 minutes of arrival
	OP-3b	Heart Attack	Average number of minutes before outpatients with chest pain or possible heart attack who needed specialized care were transferred to another hospital
	OP-4	Heart Attack	Outpatients with chest pain or possible heart attack who got aspirin within 24 hours of arrival
	OP-5	Heart Attack	Average number of minutes before outpatients with chest pain or possible heart attack got an ECG
	AMI-2	Heart Attack	Heart attack patients given aspirin at discharge
	AMI-7a	Heart Attack	Heart attack patients given fibrinolytic medication within 30 minutes of arrival
	AMI-8a	Heart Attack	Heart attack patients given PCI within 90 minutes of arrival
	AMI-10	Heart Attack	Heart attack patients given a prescription for a statin at discharge
	HF-1	Heart Failure	Heart failure patients given discharge instructions
	HF-2	Heart Failure	Heart failure patients given an evaluation of left ventricular systolic (LVS) function
	HF-3	Heart Failure	Heart failure patients given ACE inhibitor or ARB for left ventricular systolic dysfunction (LVSD)
	PN-3b	Pneumonia	Pneumonia patients whose initial emergency room blood culture was performed prior to the administration of the first hospital dose of antibiotics
	PN-6	Pneumonia	Pneumonia patients given the most appropriate initial antibiotic(s)
Mortality and Readmission	READM-30-AMI	Heart Attack	Rate of readmission for heart attack patients
	MORT-30-	Heart Attack	Death rate for heart attack patients

	AMI		
	READM-30-HF	Heart Failure	Rate of readmission for heart failure patients
	MORT-30-HF	Heart Failure	Death rate for heart failure patients
	READM-30-PN	Pneumonia	Rate of readmission for pneumonia patients
	MORT-30-PN	Pneumonia	Death rate for pneumonia patients

For testing our hypotheses, we create an index variable for process of care and an index variable for patient outcome. We construct each index in the following way.

$$\sigma_i = \sqrt{\frac{1}{N_i}\sum_j (x_{ij} - \bar{x}_i)^2}$$

$$\gamma_{ij} = \frac{x_{ij}}{\sigma_i}$$

$$\rho_j = \frac{\sum_i \gamma_{ij}}{n}$$

Applying the equations to the process of care, we first calculate the standard error σ for the ith process. Thus, we calculate fourteen standard errors (i.e., one for each process associated with treating heart attack, heart failure, and pneumonia). Next, we divide the jth observation for the ith process by the standard error for that ith process. Finally, we sum all of the ith processes for the jth observation and divide by n total number of processes (i.e., fourteen for the process of care index and six for the patient outcome index). We repeat the calculations for constructing the patient outcome index using mortality rate and readmission rate. Hospitals that did not respond to a process of care measure or patient outcome measure in the survey were removed from the data set.

The next source of data we use is the American Hospital Association's Healthcare data set. The data set provides HIT indicators that examine the degree and type of HIT implementation within a hospital. Over 3,500 hospitals respond to the survey each year and we

have data from 2010-2015. We use the data for meaningful use attestation and other control variables. Meaningful use attestation is identified with a binary indicator signaling that the hospital has achieved at least Stage 1 of meaningful use attestation. The control variables include hospital size and HIT implementation characterization. Hospital size is measured by taking the log of number of beds within a hospital. HIT implementation characterization consists of three binary variables representing whether the hospital uses a single purchased HIT system, multiple purchased HIT systems, or an internally created HIT system.

The third data source we use is the collection of reported data breaches to DHHS. 1,746 data breaches have been reported between the Breach Notification Rule's enactment in 2009 and December 31, 2016. The information contained within the data set includes the date the breach was reported, the type of breach that occurred, the hospital's name, the hospital's address, the hospital's Medicare ID (i.e., the number assigned by CMS), and a brief description of the breach. For the purpose of this study, we focus on the date the breach occurred and the hospital's Medicare ID.

Before testing our hypotheses, we cleaned and merged the data from our three sources in order to create a panel data set spanning 2012 through 2015. Cleaning the data required bringing all variables to a related period of measure. CMS and AHA survey measures are collected on an annual basis but, in some cases, they are collected quarterly. A majority of the CMS measures and all AHA measures are collected between March of the prior year and March of the following year. Because of this discrepancy, we adjust the annual data breach range to be from the prior March to the subsequent March. The adjustment improves consistency for our interpretation of the analysis. Next, we established the breach indicator variable as the presence of a breach in the prior year$(t-1)$. We use a lagged breach indicator for several reasons. First, the other

covariates in the model are reported annually. Therefore, breaches that occurred closer to the March submission deadline may have a minimal effect of the measures reported. Another reason for using a lagged variable is the slow-to-change nature of the hospital work environment as new processes take a significant amount of time for employees to incorporate into their routines (Rivers et al. 1997).

We merged the data over several steps. First, we utilized the full CMS data set to generate our index variables for each year. Once the index variables were created we merged the CMS data with the AHA data. We matched hospitals in the two data sets using the Medicare ID assigned to them. Although the Medicare ID is used as a unique identifier and generally matches, we checked to ensure the hospital name, address, and state matched. Afterward, we merged the data breach data from DHHS by matching the Medicare ID and creating a binary variable indicating that the hospital experienced a breach in the appropriate year. The complete merged data set consisted of 6,560 observations for 2,479 hospitals across the U.S. Of the 2,479 hospitals in the sample, 29% have one year of responses, 17% have two years of responses, 21% have three years of responses, and 35% have four years of responses as well as 109 data breaches at 95 different hospitals.

4.4 Results

We begin our analysis by providing evidence of the structure proposed by Angst et al. (2012) and for our model. Specifically, we estimate a series of fixed effects panel data models to establish the relationships between HIT, process of care, and patient outcome prior to using 2SLS. Although the positive or negative nature of the relationships is significant to note, the

purpose of these models is to establish the existence of the relationships for subsequent analysis. Table 11 displays the results from these estimations.

Model 1 provides the results for the effect of meaningful use attestation on process of care. The results show that there is a significant relationship between meaningful use attestation and process of care after controlling for hospital and HIT characteristics. Model 2 provides the results for the effect of meaningful use attestation on patient outcomes. The results show that there is a significant relationship between meaningful use attestation and patient outcomes after controlling for hospital and HIT characteristics. Model 3 provides the results for the effect of process of care on patient outcomes. The results show that there is a significant relationship between process of care and patient outcomes after controlling for hospital and HIT characteristics. Finally, Model 4 provides the results for the effect of meaningful use attestation and process of care on patient outcomes in the same model. The results continue to show a significant relationship between meaningful use attestation and patient outcomes as well as a significant relationship between process of care and patient outcomes. Based on our findings, we lend support for Angst et al. (2012)'s framework suggesting that HIT implementation affects patient outcomes, but the effect may be a result of its altering process of care. We also support the presence of endogeneity with the significant relationship between meaningful use attestation and process of care and strengthen our reasoning for using a 2SLS approach.

Table 11. Relationships between HIT, Process, and Outcome

	Model			
Variable	(1) Outcome	(2) Process	(3) Outcome	(4) Outcome
Control				
Hospital Size	0.1252 (0.0902)	0.2334 (0.2082)	0.2595 (0.2073)	0.2678 (0.2068)
HIT – Self Implemented	-0.2603† (0.1553)	0.1416 (0.3585)	-0.0507 (0.3559)	0.0700 (0.3561)
HIT – Single System	-0.1474† (0.0886)	0.0289 (0.2046)	0.0149 (0.2036)	-0.0117 (0.2032)
HIT – Multiple Systems	-0.1328	0.0650	0.0398	0.0285

	(0.0899)	(0.2075)	(0.2066)	(0.2061)
Independent				
Process of Care			-0.2824** (0.0359)	-0.2749** (0.0359)
Meaningful Use	-0.1445** (0.0477)	0.5232** (0.1102)		0.4835** (0.1096)
Constant	1.0943 (0.2237)	8.9983** (0.5164)	9.7772** (0.5039)	9.2991** (0.5143)
F-Value	3.10**	4.80**	12.64**	13.82**
Observations	6560	6560	6560	6560
# of Hospitals	2479	2479	2479	2479

$\dagger p \leq 0.10$, $*p \leq 0.05$, $**p \leq 0.01$

Since we demonstrate evidence that HIT affects process of care, we use a 2SLS specification to compare the relative effect of process of care on patient outcomes. We first estimate process of care as a function of HIT characteristics, meaningful use attestation, and experiencing a data breach in the prior year. We then estimate the effect of process of care and our control variable hospital size. Model 1 provides the results of estimating the first stage and Model 2 provides the results of estimating the second stage. According to our estimation (seen in Table 12), we find evidence that meaningful use attestation and experiencing a breach significantly affect process of care. Furthermore, the results also show that process of care directly influences patient outcomes.

The first stage of the 2SLS estimation is a panel data regression using fixed effects. Through this estimation, we discover that meaningful use attestation negatively affects process of care and experiencing a data breach in the prior year positively affects process of care. Specifically, hospitals that achieve a minimum of Stage 1 meaningful use see a decrease in the process of care index by 0.1445 standard deviations. Hospitals that experience a data breach in the prior year see an increase in the process of care index by 0.0793 standard deviations. Thus, we must reject *Hypothesis 1*, which states that meaningful use attestation will have a positive relationship with process of care. We also reject *Hypothesis 2*, which states that experiencing a

data breach will have a negative relationship with process of care. In the second stage, we find evidence that improvements in process of care are associated with reducing mortality and readmission rates (i.e., improving patient outcomes). Specifically, increasing the process of care index by 1 point leads to a decrease in the mortality and readmission rates by 1.7957 standard deviations. The results in Model 2 support *Hypothesis 3*, which states that a raise in process of care will have a positive relationship with patient outcome.

Table 12. 2SLS for Fixed Effects Panel Data Model

	Stages	
Variable	**(1)**	**(2)**
Control		
Hospital Size	0.1234 (0.0902)	0.4532† (0.2630)
HIT – Self-Implemented	-0.2606† (0.1552)	
HIT – Single System	-0.1456 (0.0886)	
HIT – Multiple Systems	-0.1307 (0.0899)	
Independent		
Process of Care		-1.7957** (0.6788)
Meaningful Use	-0.1445** (0.0477)	
Breach t-1	0.0793† (0.0462)	
Constant	1.0952** (0.2236)	11.0093** (0.7735)
F-Value	3.0800**	
x^2		628387.9800**
Observations	6560	6560
Number of Hospitals	2479	2479

†$p \leq 0.10$, *$p \leq 0.05$, **$p \leq 0.01$

For additional robustness, we consider the effect of a breach in both the current year and the year prior. Our results for the lagged data breach variable could be affected if there is an initial and immediate impact of the breach in year t. We test for the effect using a fixed effects panel data model that regresses process of care on an indicator for a breach occurring in the same year and an indicator for a breach occurring in the prior year. The results from our model

indicate that a breach in the same year, t, does not affect process of care ($\beta = 0.0252, p = 0.38$). The model also indicates marginal significance that a breach in the prior year, $t - 1$, does affect process of care ($\beta = 0.0750, p = 0.10$). Next, we question if the two coefficients are significantly different from one another and if their effect on process of care is weakened by the presence of both variables in the same model. We perform an F-test to determine whether the difference in the coefficients is equal to 0. The result from the F-test ($F = 0.74, p = 0.39$) indicates that there is inconclusive evidence to suggest that the difference in the coefficients is not equal to 0. Therefore, we are confident in our usage of the lagged breach variable for testing the effect of a data breach on process of care.

4.5 Discussion

According to our results, data breaches affect the process of care for heart attack, heart failure, and pneumonia, which subsequently affects patient outcomes for those conditions. However, the relationship is the opposite of our hypothesis. As mentioned previously, data breaches are often disruptive to the organization because of lengthy audits needed to identify the cause of the breach as well as time spent addressing the breach. Therefore, we expected a hinderance to process of care. Our results suggest that the fallout from a breach may bring beneficial improvements to process of care. The question arises of why we find this improvement to process of care. Currently hospitals may be fined millions of dollars by DHHS if the hospital or its employees are found not complying with HIPAA guidelines for securing patient information. The effectiveness of DHHS fines and audits as sufficient motivation for hospitals to actively change business processes has been debatable over the past several years and become an area of concern for law makers (Sullivan and Davis 2017). Our findings suggest

that a deterrence effect may be present and promote business process reengineering and organizational change that otherwise may not have occurred.

Literature shows that hospitals operate in a hierarchical structure with hospital support staff being at the bottom tier, nurses at the middle tier, and doctors and specialists at the upper tier (Raman and Bharadwaj 2012; Abraham and Reddy 2008; Calvin et al. 2009). Hospital support staff tend to follow the lead of nurses and doctors, while nurses follow the lead of doctors. For instance, the adherence to hospital policies among all hospital staff has been shown to be higher when doctors place greater value on those policies (Raman and Bharadwaj 2012). Therefore, in order to promote and implement successful organizational change hospitals must target higher tiers (e.g., doctors and specialists) for a trickledown effect. Nearly half of the breaches reported to DHHS that were caused by a hospital employee were mistakes made by a doctor or specialist. These doctors typically face a high degree of scrutiny during the following audit processes (Torrieri 2013). Thus, doctors may take additional precautions in the form of added patient care to ensure they do not draw attention to themselves in other areas, which leads to further changes in process of care among nurses and support staff.

Prior studies focused their attention on HIT improvements to medication errors when doctors are capable of submitting prescriptions through a computerized physician order entry system, which involves the final steps of treatment. However, in this study, many of the measures included in our process of care index are associated with the initial steps for treating heart attack, heart failure, and pneumonia. Thus, our finding that meaningful use attestation weakens the process of care is interesting and extends the HIT literature because it suggests that the improvements to coordinating health services may hinder the initial stages of care. Hospitals require nurses and doctors to be efficient by attending patients quickly because of high patient

volumes and life-threatening conditions. Nurses and doctors have indicated that they are divided on the usage of HIT throughout the process of care. For instance, nurses who heavily use HIT and adhere to hospital HIT policies acknowledge that it can reduce the number of patients they attend at a given time (Chisholm et al. 2000; Mellott et al. 2013). This decrease in efficiency would then appear in our data set as a reduction in process of care because several of our measures involve the time it takes to administer care. Our finding is unique among the HIT literature because it highlights the need for balance between HIT usage and efficient patient care.

The need for balance between achieving meaningful use and efficient patient care offers practical implications for policy makers and hospital management. Through incentive programs CMS has sought to encourage hospitals to progress toward greater sophistication and communication with HIT. CMS's objective for the end of 2018 is to have some degree of HIT implementation in all hospitals within the U.S. However, our results suggest that enhancing HIT to the point of meaningful use may lead to more harm than good. It is then important for hospital management to monitor the effects of HIT implementation and be flexible in its timeline of release. One possible explanation for the negative effects on process of care is physicians' displeasure in HIT. Physicians report significant dissatisfaction with their HIT because of its poor functionality and rushed implementation (Peckham 2016). Therefore, the deadlines imposed by CMS to achieve meaningful use may leave smaller, less profitable hospitals scrambling to catch up, and in turn reducing their ability to care for patients. Federal policy makers may find our results useful in restructuring their timelines for HIT incentive payments and legislative mandates.

We must address several limitations in the essay. First, the 2SLS model relies on an instrumental variable that is ideally supported by theory and prior literature. Although we

provide theoretical evidence for using HIT as an instrumental variable based on Angst et al. (2012)'s structure-process-outcome framework, we use meaningful use attestation as a binary variable encompassing many different aspects of HIT. The variable is limited because it does not account for all aspects of HIT, meaning that a better representation of HIT may exist, and prior literature has yet to use meaningful use attestation in this manner. However, we are confident in its usage for this study. A requirement for 2SLS is that the instrumental variable must not directly cause the dependent variable, patient outcome in our case. We argue that meaningful use attestation does not directly lead to better or worse patient outcomes but rather is a compilation of tools that enable change in patient outcomes. Therefore, the HIT implemented under meaningful use is dependent upon a care providers ability to use and incorporate the tools and information gained and translate it to change in patient outcomes.

Future research could explore additional HIT variables to potentially identify an alternative instrument. For example, our binary variable representing meaningful use attestation does not distinguish the hospital's stage of meaningful use for a given year nor the specific HIT used at the hospital. Future studies may discover that process of care is affected differently according to the stage of meaningful use such that it is not until later stages of meaningful use that hospitals begin to see improvements to process of care.

Additionally, a limitation in our analysis is the use of HIT control variables. The use of the HIT control variables in the first stage of the 2SLS model places them with our meaningful use attestation indicator, which may introduce endogeneity. However, we are confident that the type of system implemented does not influence meaningful use attestation because our data set contains only two hospitals that did not have some HIT system implemented and less than two percent of hospitals did not achieve meaningful use. Furthermore, the number of HIT systems

and the self-development of the system are not requirements for meaningful use attestation. Thus, we have instances in which multiple HIT systems are implemented but do not meet the criteria for meaningful use.

Another limitation is that our data is reliant on the truthful measures reported by hospitals to CMS and AHA. The Affordable Care Act, enacted in 2010, began incentivizing hospitals to reduce costs and improve patient care by implementing a CMS program that offers payment incentives to hospitals that reduce their readmission rate for heart failures. Healthcare policy experts became concerned that hospitals may take steps to improve readmission rates on the surface by cutting corners. Their fears were supported by a recent study that showed hospitals were reducing readmissions by simply declining patients for readmittance into the hospital (Gupta et al. 2018). Unfortunately, we are unable to verify the veracity of the reported measures to CMS and AHA, but we attempt to mitigate misleading reporting by including mortality and readmission rates for heart attack and pneumonia, which are conditions not associated with an incentive program.[30] Furthermore, the reported mortality and readmission rates are commonly used throughout the healthcare and HIT literature and are therefore a limitation among all studies using the measure.

Although we use heart attack and pneumonia mortality and readmission rates to offset the possible bias of false reporting, we acknowledge that it limits our interpretation of the relationships between data breaches, HIT, process of care, and patient outcome. Our current interpretation suggests that each of the conditions are affected in the relationship (e.g., a data breach leads to improvements in process of care for heart attack, heart failure, and pneumonia). However, it is possible that changes in the relationships may be attributable to only one of the

[30] CMS began offering incentive payments for improving mortality and readmission rates for heart attacks under the Cardiac Rehabilitation Incentive Payment Model. These incentive payments were enacted on January 3, 2017 and therefore do not affect the 2012 through 2015

conditions and it is driving the significance for our index variables. Yet, we remain confident in the usage of all three conditions in our indexes because hospitals are incentivized to improve the mortality and readmission rate for each condition through CMS using the rates for nationwide hospital ranking and patients' ability to publicly view the rates. Future research may take a more granular approach to tease out the effects on individual conditions. Specifically, researchers could study the effects of meaningful use attestation and experiencing a data breach on the process of care for heart attack and, subsequently, the influence of process of care for heart attacks on heart attack mortality and readmission rates. Studying the individual effects will benefit the HIT literature by identifying the areas of care most influenced by HIT and meaningful use attestation.

5. Conclusion

Throughout the thesis, we investigated the extent to which consumers and firms respond to information privacy and security factors during consumer information disclosures and organizational data breaches. Essay 1 analyzed relative consumer valuations for an information disclosure. Through three experiments we demonstrated that consumer privacy valuations are largely unaffected by requiring the disclosure of personally identifying information, the information context, and the intended secondary use of the disclosed information, when these factors are combined in an online disclosure decision. Our results contrast the prior research, which has shown these factors to have significant effects on privacy valuations when studied in isolation. Our results are robust, and the experiments used incentive compatible techniques from experimental economics across two separate samples (students and Amazon Mechanical Turk). We did find that increasing the saliency of privacy factors in the disclosure and highlighting the consequences of disclosing private information increases privacy valuations. The results from Essay 1 offer useful implications for consumers, organizations that capture consumer information, and policy makers seeking to improve consumers' privacy protections.

Essay 2 analyzed subsequent data breaches within firms. Prior data breach literature has shown the existence of a deterrence effect in that firms take precautions and preventive measures to avoid experiencing subsequent data breaches. However, after reviewing data breaches over the past decade, many firms experience multiple breaches. Therefore, we disaggregate the deterrence effect and use survival analysis to investigate which policies actually deter firms from future breaches. To conduct the analysis, we utilize the publicly available PRC dataset for breach incidents between 2005 and 2016. After interpreting the results from a hazard model and using counterfactuals, we discover that establishing a breach as a criminal offense lengthens the

duration between breaches by over two years and having firms disclose to a state attorney general lengthens the duration between breaches by over three years. Furthermore, we learn that permitting firms to forego breach notification when the compromised information is not harmful to consumers leads to more frequent breaches within a firm.

We also find that educational institutions have the longest duration (i.e., low risk) between subsequent data breaches while non-retail businesses and non-profit organizations have significantly shorter durations (i.e., greater risk) between breaches, suggesting that firms within certain industries may be more responsive to data breaches. Following the survival analysis, we studied the susceptibility of given industries to particular breach types using several multinomial logit models. Our results suggest that firms can specialize their breach prevention efforts toward industry specific breach types. Our findings provide useful results for guiding future data breach research and policy regulations. For instance, privacy researchers can take these findings and apply additional firm-level characteristics such as firm size and other economic firm variables to aid with firm distinction to test whether the results hold true.

Finally, Essay 3 analyzed the impact of data breaches within hospitals. We predicted that data breaches negatively affect the process of care while attesting to meaningful use positively affects the process of care. We estimated a 2SLS fixed effects model using hospital level data from over 2,000 hospitals through the U.S. Our model indicates that data breaches actually improve the process of care for heart attack, heart failure, and pneumonia. Surprisingly, meaningful use attestation reduces the process of care for heart attack, heart failure, and pneumonia. Furthermore, improvements to process of care lead to improvements in patient outcome. Thus, our findings suggest that experiencing a data breach in a hospital may lead to better patient outcomes in the future. These findings offer useful insight for hospital

management's HIT and information security investments. Essay 3 also draws attention to and suggests further exploration of the potential negative consequences of HIT implementation.

Overall, the takeaway from our three essays is that consumers and firms do not alter their information privacy and security behavior unless proper incentive is given, which may bring detrimental consequences in the near future. For example, the Cambridge Analytica scandal with Facebook embodies the information sharing relationship between consumer and firm, and when left unchecked both sides can experience negative consequences at the hands of malicious parties. Cambridge Analytica was a political consulting firm that, as mentioned previously, compiled personal data from approximately 87 million Facebook users without their consent in order to sway public opinion and behavior (Confessore 2018). Facebook claims to be unaware of the malintent behind Cambridge Analytica's data capturing. The scandal has been labeled as a data breach because Cambridge Analytica took personal information from Facebook users that was not in agreement with Facebook's rules and policies. Although consumers have expressed outrage at the mishandling of their information our findings suggest that they are unlikely to change their behavior. The dichotomous nature of consumer privacy behavior we found in Essay 1 was affirmed by the early events following the release of the scandal. Users who were angered by the situation refused to continue using Facebook and deleted their user account. However, the overwhelming majority of users remain loyal to the social media platform and continue with normal usage (Zetlin 2018).

Facebook's response to the scandal supports our findings in Essay 2 and Essay 3 such that firms facing scrutiny and severe penalties in the aftermath of a data breach will act to prevent future breach and provide positive signals to overseeing agencies. Facebook's stock valuation dropped by 24%, a loss of $134 billion, within eight days of the Cambridge Analytica

scandal being reported. Facebook continues to see significant monetary losses months later (e.g., setting the record for the largest drop in stock value in a single day) as it attempts to send positive signals to users. In addition to the monetary loss, Facebook executives drew scrutiny from the U.S. Congress and negative publicity across the world. Facebook has responded to the scandal by hiring over 20,000 employees who will specialize in data privacy and information security throughout the social media platform and nearly doubling its information security expenditures (Cherney 2018).

In conclusion, the research presented in this thesis significantly contributes to the information privacy and security literature by presenting a holistic view of the information sharing relationship between consumers and firms. Consumers consistently demonstrate an inability to comprehend and consider privacy implications in a disclosure. Therefore, consumers rely on firms to protect their data from misuse. However, there is a misalignment of interests as firms are hesitant to invest resources into protecting consumer data until there is a need to do so. Firms respond in a positive way when federal oversight is present by strengthening information security and, in the case of healthcare, reengineering organizational processes. Evidence of our findings can be seen in consumer and firm actions surrounding real world events. Thus, by demonstrating how both parties respond to information privacy and security factors, we offer the framework for a new stream of research to examine and promote the alignment of information privacy and security interests between consumers and firms through federal regulation.

APPENDIX A: Essay 1 Supplementary Data

Overview

Contained within Appendix A are an overview of the experiment protocol, relevant figures and tables describing each of the three studies we conducted, and additional details regarding experiment participants and empirical analyses. Figures A1 and A2 provide the experiment protocols participants completed, in order from left to right. All participants followed the same path through the protocols independent of their treatment assignment. With the exception of Study 3, participants began the experiment by reading a series of pages that created a scenario of market research by Google. Participants in Study 3 also went through these scenario pages, but prior to these pages they watched a video presentation on the consequences of disclosing private information and completed a quiz on the topics discussed in the video (viewable at https://goo.gl/X2C5lj). Tables A1, A4, and A6 demonstrate the manipulations found in the scenarios.

As regards the experiment procedure, participants were given instructions on how they may sell their private information by entering a valid WTA. Following the instructions, we quizzed participants to ensure they understood how the selling mechanism (i.e., Becker-DeGroot-Marschak procedure) operates. Participants were then given a list of sample items (Table A2) and told that they may be asked to disclose private information contained in the list or any additional private information Google may require. Following the list of sample items, participants entered their WTA between $0.00 and $5.00. Participants were prevented from continuing if an invalid WTA was entered.

Immediately after submitting their WTA, participants were informed whether their information sold. If participants sold their private information (indicated by dashed arrows), they

were directed to a form that contained the list of sample items presented to them earlier. Participants had to disclose private information for each item in the form before continuing the experiment. Participants who did not sell their private information were directed to the post-experiment survey without completing a form. Lastly, participants answered several questions on a post-experiment survey. At the end of the survey, participants were thanked for their time and informed on how they may receive payment if their private information sold.

We also conducted an a priori power analysis to determine how many participants were required per cell in our factorial design to have sufficient power for detecting at least a medium effect size with an alpha of 0.05. With three factors and two levels of each factor, a minimum participant count per cell is 20 in order to obtain a power of 0.88. The minimum cell count in Study 1 is 34 (37.5 average per cell for students), which is sufficient for a power of 0.98. The minimum cell count in Study 2 is 20 (25.25 average per cell for students; 54.5 average for AMT). Study 3 has two factors with two levels of each factor, so 35 participants are required per cell to obtain a power of 0.84. The minimum cell count in Study 3 is 32 (35 average per cell for students; 45.5 average for AMT). A post hoc power analysis for the student samples show that even with an alpha of 0.10, the power for Study 3 is approaching 0.90. Further, after pooling the studies, the post hoc power achieved is > 0.99 for detecting a medium effect at an alpha of 0.05, and > 0.90 for detecting a small effect. Overall, we are confident that we have sufficient power to detect even small effect sizes, especially given the very high p-values for our null results.

Our post-experiment survey captured demographics, control information, and participants' privacy concerns.[31] To control for participants' experience using the Internet, survey items included Internet usage, online privacy breach history, and propensity to falsify

[31] See Hong and Thong (2013) for the survey measures.

information online. To measure privacy concerns, we used validated and reliable measures from the Internet Privacy Concerns (IPC) model by Hong and Thong (2013).

The IPC model illustrates a person's Internet privacy concerns as a third order construct that is a reflection of the (i) trust one places in an organization to handle their private information, (ii) the risk one perceives for disclosing information to an organization, (iii) their awareness of privacy-related issues, (iv) how one manages their interaction with others, and (v) how one manages their private information. The model further captures how a person manages their interaction with others as a reflection of first-order constructs that measure an individual's attitude toward an organization's data collection practices, an organization's secondary use for the information, and the desire for controlling one's own private information. In addition to a person's interaction management, how a person manages their private information is also a reflection of first-order constructs. These constructs are the individual's belief regarding the organization's responsibility for the protection of the individual's private information against unauthorized access and the degree to which an organization should ensure the information is error free. We reused the measurement items for the factors in the IPC model, and implemented a 7-point Likert-type scale for all survey items.

We estimated the IPC model using PLS-SEM (SmartPLS v3.2). One reason we chose SmartPLS in particular was because of its ability to generate a factor score for an n-order construct. The factor score represents a single measure of a participant's overall Internet privacy concerns, which we can use as a predictor variable in the analysis of our experimental data. Table 1 provides the summary statistics for the model. To strengthen parameter and estimate stability for our reliability and validity statistics, we ran a bootstrap procedure and PLS algorithm with inflated settings, 2000 and 1000 respectively. Consistent with Hong and Thong (2013), all

items loaded correctly on their respective constructs. The results show that each construct demonstrates sufficient reliability (Cronbach's $\alpha > 0.7$; Composite Reliability > 0.7) and significance ($p < 0.001$). In the covariance matrix, we find that each construct is valid in the model with all AVE scores greater than 0.5 and the square root of the AVE for each construct (shown in italics) greater than the covariance values. After successfully recreating the IPC factor structure and verifying the model's reliability and validity, we generated a factor score representing each participant's Internet privacy concerns for use in the analyses that follow. We were then able to analyze the effects from increased saliency and informed consequences on individuals' privacy concerns. We learned that increasing the saliency does not result in heightened privacy concerns among participants, but informing participants the consequences associated with disclosure does heighten privacy concerns. Our usage of the IPC model further validates the IPC model by Hong and Thong (2013) and extends the prior research by applying it in a unique and interesting new context such as predicting privacy valuations.

Tables A3, A5, and A7 show the text of each page participants viewed throughout the experiments. Within the text, we indicate the manipulation text from the corresponding manipulations table with brackets and italics (e.g., [*Secondary Use*]). Table A8 includes the items of our post-experiment survey. All items are measured using a 7-point Likert scale unless otherwise specified. Table A11 provides descriptive statistics and PLS measures for creating the IPC latent variable constructs. Table A12 contains the results from regressing increased saliency (i.e., participation in Study 2) and informed consequences (i.e., participation in Study 3) on IPC.

Participation, Opt Outs, and Incomplete Observations

We outline in Table A9 those observations that were removed from each study. As we can see by the Opt Out column, the rates of explicitly opting out are quite low, less than 2% in every case except AMT Study 3 at 3%. It appears the opt out rate does slightly increase as we increase the saliency of consequences in each study, as may be expected. We know that increasing the saliency of consequences does increase the WTA, and that the increase is greatest in Study 3. Next, we discuss the issue of incomplete observations, as that is the source of the bulk of removed observations for the student sample. Almost all the incomplete observations are driven by limitations of the Qualtrics system used for data collection. Qualtrics records an observation each time a potential participant opens the survey. Therefore, subjects who click on the link to participate, regardless of whether they continue with the study, create a potentially incomplete observation in the data set. Upon inspection, almost all of the unfinished observations were abandoned prior to the request to sell private information. Thus, subjects were assigned to a treatment but did not advance to a point in which they could tease out the purpose of the study. Also, incomplete observations may occur for a multitude of reasons. For example, students may start the study and then become distracted, click on a recruitment link in email using a phone and then later click the link for a second time on a computer, or a participant may have uncertainty if he/she wants to participate at all. In contrast to the student sample, we believe the AMT workers intend to participate in the study once it is started so that their time is used efficiently.

Last, we consider those that failed the attention and/or manipulation checks. We report a bulk rate of ~20-25% removed observations but after removing the incomplete and opt out observations, the percentage of participants that failed attention and/or manipulation checks is less than 9% for students. Regarding AMT, the difference between Study 2 and Study 3 is driven by eight participants that did not play the entire video. Therefore, only 4 of the 12 participants

failed the attention or manipulation checks in similar ways for AMT Study 3, putting the rate percentage of observations at 2.00%, which is almost identical to the 1.78% rate for AMT Study 2. In addition, instead of implementing attention questions for the AMT workers, we instead used time checking for each page to discover fraudulent workers (e.g., bot workers). If a worker spent a significantly smaller amount of time on the page than the average, we flagged the observation for later review. Any flagged AMT observation was dropped from the study during review if the worker completed the study in an impossibly fast amount of time.

A further breakdown of participants removed from each study is shown by Table A10. There are not any clear indications of differences in participation rates between treatments, and there seems to be no repeatable pattern between the invasiveness of the treatment groups and the incomplete responses and/or failed attention or manipulation checks. For example, in Student Study 1, the group with the highest failed attention checks has medical context with no secondary use but with identifying information, whereas in Student Study 2, the medical context with secondary use but not identifying information is highest. A similar situation occurs for incompletes, where Student Study 1 has the highest number of incompletes for medical information with no secondary use and no identifying information, but Student Study 2 has medical secondary use identifying with the largest number of incompletes. In addition, the overall results are consistent between the Student population and the AMT population, and the AMT population also has fewer dropped observations. The overall breakdown of participants gives us confidence in our random assignment of participants to treatments and effectiveness of experimental manipulation.

Details Regarding Empirical Analysis

To address heteroscedasticity in Study 3 for the AMT sample, we estimated two additional models. The first model estimated is Generalized Method of Moments (GMM) due to its ability to generate efficient parameter estimates in the presence of heteroscedasticity (Baum et al. 2003). We also estimated a Symmetrically Trimmed Least Squares (STLS) model, based upon the ability for STLS estimators to address heteroscedasticity in the Tobit model (Powell 1986). Results are shown in Table A13 and are qualitatively consistent with the Tobit regressions already presented in the paper.

Tables and Figures

Figure A1: Study 1 and Study 2 Experiment Protocol

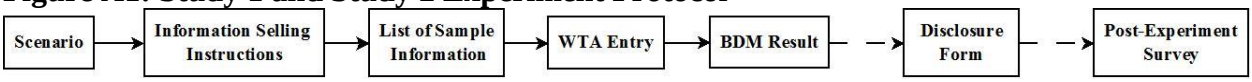


Figure A2: Study 3 Experiment Protocol

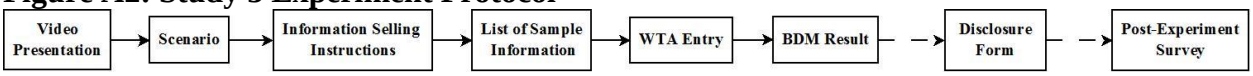


Table A1: Study 1 Manipulations	
Information Context	
Medical	medical history
Shopping	shopping preferences
Secondary Use	
Secondary Use	will distribute the information you provide to outside marketing and advertising agencies for various purposes
No Secondary Use	will not distribute information to any third parties. The information will be for internal application use only
Identifying Information	
Identifying Information	will store identifying information, such as name, email, and phone number, with the medical information you provide
No Identifying Information	will not store identifying information, such as name, email, and phone number, with the medical information you provide

Table A2: List of Sample Information Items	
Medical	**Shopping**
Allergies Illnesses	Frequent Retail Stores Frequent Purchases

Diseases	Recent Purchase History
Family History	Preferred Shipping Method
Sexual Activity	Product Attributes
Smoking Habits	Preferred Method of Payment
Drug Use	Frequent E-commerce Websites
Alcohol Use	Common Grocery Purchases
Blood Type	Mobile App Purchases

Table A3: Study 1 Outline

Page 1

Welcome

The following is a study on the valuation of information. Google Inc. is currently developing a new [*Information Context*] application. **Google Inc. wishes to begin paying users** for the information they provide when registering for the application. However, Google Inc. does not know how much to compensate users for their information. In order to capture an appropriate compensation value, we will ask you to enter your selling value for **ALL** of your medical information. **The value entered must be between $0.00 and $5.00.** Throughout the study, we will refer to this value as your selling price.

It is important to note that Google Inc. [*Secondary Use*].

The application Google Inc. is developing will require users to enter [*Information Context*] information about themselves. A sample of the information Google Inc. may request appears later. Due to a nondisclosure agreement with Google Inc., we cannot disclose the specifics of the new application. However, the application will provide a quality service for its users.

You may opt out of this study at any time.

Page 2

Information Selling Instructions

After viewing the list of information Google Inc. may request, you have two options.

1. You may enter your selling price for the information Google Inc. requests. **Remember that the value entered is your selling price for all of your information, not individual pieces.**
2. You may opt out if you do not wish to participate or if your selling price is greater than $5.00. A opt out option is available on the selling price page. If you choose to opt out, you must provide a reason for doing so.

If you choose to participate and enter a selling price then you will type your selling value into a text box. The value must be between $0.00 and $5.00 and in the format X.XX. The following occurs after you submit your selling price:

Google Inc.'s information buying algorithm calculates a buying price between $0.00 and $5.00.

If Google Inc.'s buying price is greater than or equal to your selling price, you will sell your information to Google Inc. for the buying price and must provide the information Google Inc. requests. If Google Inc.'s buying price is less than your selling price, you will not sell your information and do not provide your information to Google Inc.

Example 1:
Your selling value is $1.00, Google Inc. buying price is $2.50 ⇒ You will sell your information for $2.50.

Example 2:
Your selling value is $3.00, Google Inc. buying price is $2.00 ⇒ You will **NOT** sell your information.

Example 3:
Your selling value is $2.50, Google Inc. buying price is $2.50 ⇒ You will sell your information for $2.50.

You will receive course credit for participating in this experiment and finishing the survey at the end. If you sell your information to Google Inc., you will receive the course credit and the buying price. If you do not sell your information to Google Inc. or choose to opt out, you will only receive the course credit.

****IMPORTANT****
Before you receive your payment from Google Inc., Google Inc. will verify the information you provide for truthfulness.

Page 3

Information Selling Instructions

We will now demonstrate how pricing works. It is in your best interest to accurately state your true valuation as your selling price for your information. The following are two examples of why:

Example 1: What happens if your stated selling price is HIGHER than your true value:
Imagine you value your information at $3.00, but you enter a selling value of $4.50. We will say that the Google Inc. buying price is $4.25.

Since the buying price, $4.25, is less than your selling price, $4.50, you will not sell the information to Google Inc. and will not earn the $4.25. Therefore, you will miss the opportunity to sell your information for a price you deem as reasonable.

Example 2: What happens if your stated selling price is LOWER than your true value:
Imagine you value your information at $1.75, but you enter a selling value of $0.75. We will say

that the Google Inc. buying price is $1.00.

Due to your selling price being lower than the Google Inc. buying price, you must sell your information to Google Inc., even though you value the information much more than the $1.00 you will receive. You will forfeit your information for less than what you think it is worth.

Page 4

Information Selling Instructions Quiz

Before proceeding, we wish to ensure you understand all of the instructions clearly. Below are four example scenarios, please choose the best answers to the questions:
Scenario 1: Your stated selling value is $1.50. The Google Inc. buying price is $2.50. What will happen next?

- You will not sell your information or complete the form
- You will sell your information for $1.50 and complete the form
- You will sell your information for $2.50 and complete the form

Scenario 2: Your stated selling value is $1.50. The Google Inc. buying price is $1.00. What will happen next?

- You will not sell your information or complete the form
- You will sell the information for 0.50 and complete the form
- You will sell the information for 1.50 and complete the form

Scenario 3: Google Inc.'s buying price is $0.75 and your selling value is $0.25. What must you do after winning?

- Do not complete the information form
- The information sells so you must complete the information form

Is the Google Inc. buying price (and your selling value) based on each individual piece of information or all information on the form?

- Each individual piece of information
- All information on the form

Page 5

Thank you for completing the tutorial.

The following page will request your selling value for your medical information. Below, we provide a brief list of possible information items Google Inc. will request. This list is not comprehensive and Google Inc. reserves the right to request [*Information Context*] that is not

shown below.

Please be aware that Google Inc. [*Identifying Information]*

[*List of Medical/Shopping Information Items]*

Page 6

Enter a selling value between $0.00 and $5.00, in the format X.XX, for the information presented in the sample form.

If you wish to opt out of the study, you may do so. To opt out please select the Opt Out option and click >>.

Opt Out

Page 7 (if information was sold)

SOLD!

Your value, $[*participant's WTA*], is less than the randomly drawn value, $[*system generated WTA*]. Therefore, you will earn an additional $[*Random WTA minus the WTA entered*] at the end of this experiment. Please proceed to the next page and enter the requested information.

Page 7 (if information was not sold)

NO SALE

Unfortunately your selling price, $[*participant's WTA*] was greater than the randomly drawn value, $[*system generated WTA*]. You are not required to enter your information. Please proceed to the next page.

Table A4: Study 2 Manipulations	
Information Context	
Medical	***medical history***
Shopping	***shopping history***
Secondary Use	
Secondary Use	***will distribute the information you provide to outside marketing and advertising agencies for various purposes***
No Secondary Use	***will not distribute information to any third parties. The information will be for internal application use only***
Identifying Information	
Identifying Information	***will store identifying information, such as name, email, and phone number, with the medical information you provide***

No Identifying Information	***will not store identifying information, such as name, email, and phone number, with the medical information you provide***

Table A5: Study 2 Outline

Page 1

Welcome

Google Inc. is currently developing a new [Information Context] application and wishes to pay its users for the information they provide when registering for the application. However, Google Inc. does not know how much to compensate users for their information. In order to capture an appropriate compensation value, we will ask you to enter your selling value for ***your [Information Context]***. The value entered must be between $0.00 and $5.00. This value will be referred to as your selling price.

The application Google is developing will require users to enter [*Information Context*] about themselves. A sample of the information Google Inc. may request appears later. Due to a nondisclosure agreement with Google Inc., we cannot disclose the specifics of the new application. However, the application will provide a quality service for its users.

You may opt out of participation at any time.

Page 2

Welcome

We must inform you that ***Google Inc. [Secondary Use]*** Disclosing your information represents consent for Google Inc. to share the information with other [*medical/advertising*] companies.

Page 3

Welcome

Shown below is a brief list of possible information items you will provide when registering for an account. In addition to [*Information Context*], ***you [Identifying Information]***. This list is not comprehensive and Google Inc. reserves the right to request [*Information Context*] that is not shown below.

[*List of Medical/Shopping Information Items*]

Page 4

Welcome

In summary, Google Inc. is creating a new application in which
Users *provide* [*Information Context*]
Google [*Secondary Use*]

Users provide [*Identifying Information*]

Compensation for your information can fall between $0.00 and $5.00. You may opt out at any time by closing your web browser.

Page 5

Information Selling Instructions

****IMPORTANT**** Before you receive your payment from Google Inc., Google Inc. will verify the information you provide for truthfulness.

After viewing the list of information Google Inc. may request, you have two options.

1. You may enter your selling price for the information Google Inc. requests. **Remember that the value entered is your selling price for all of your information, not individual pieces.**
2. You may opt out if you do not wish to participate or if your selling price is greater than $5.00. A opt out option is available on the selling price page. If you choose to opt out, you must provide a reason for doing so.

If you choose to participate and enter a selling price then you will type your selling value into a text box. The value must be between $0.00 and $5.00 and in the format X.XX. The following occurs after you submit your selling price:

Google Inc.'s information buying algorithm calculates a buying price between $0.00 and $5.00.

If Google Inc.'s buying price is greater than or equal to your selling price, you will sell your information to Google Inc. for the buying price and must provide the information Google Inc. requests. If Google Inc.'s buying price is less than your selling price, you will not sell your information and do not provide your information to Google Inc.

Example 1:
Your selling value is $1.00, Google Inc. buying price is $2.50 ⇒ You will sell your information for $2.50.

Example 2:
Your selling value is $3.00, Google Inc. buying price is $2.00 ⇒ You will **NOT** sell your information.

Example 3:
Your selling value is $2.50, Google Inc. buying price is $2.50 ⇒ You will sell your information for $2.50.

You will receive course credit for participating in this experiment and finishing the survey at the end. If you sell your information to Google Inc., you will receive the course credit and the buying price. If you do not sell your information to Google Inc. or choose to opt out, you will

only receive the course credit.

Page 6

Information Selling Instructions

We will now demonstrate how pricing works. It is in your best interest to accurately state your true valuation as your selling price for your information. The following are two examples of why:

Example 1: What happens if your stated selling price is HIGHER than your true value:
Imagine you value your information at $3.00, but you enter a selling value of $4.50. We will say that the Google Inc. buying price is $4.25.

Since the buying price, $4.25, is less than your selling price, $4.50, you will not sell the information to Google Inc. and will not earn the $4.25. Therefore, you will miss the opportunity to sell your information for a price you deem as reasonable.

Example 2: What happens if your stated selling price is LOWER than your true value:
Imagine you value your information at $1.75, but you enter a selling value of $0.75. We will say that the Google Inc. buying price is $1.00.

Due to your selling price being lower than the Google Inc. buying price, you must sell your information to Google Inc., even though you value the information much more than the $1.00 you will receive. You will forfeit your information for less than what you think it is worth.

Page 7

Information Selling Instructions Quiz

Before proceeding, we wish to ensure you understand all of the instructions clearly. Below are four example scenarios, please choose the best answers to the questions:
Scenario 1: Your stated selling value is $1.50. The Google Inc. buying price is $2.50. What will happen next?

- You will not sell your information or complete the form
- You will sell your information for $1.50 and complete the form
- You will sell your information for $2.50 and complete the form

Scenario 2: Your stated selling value is $1.50. The Google Inc. buying price is $1.00. What will happen next?

- You will not sell your information or complete the form
- You will sell the information for 0.50 and complete the form

○ You will sell the information for 1.50 and complete the form Scenario 3: Google Inc.'s buying price is $0.75 and your selling value is $0.25. What must you do after winning? ○ Do not complete the information form ○ The information sells so you must complete the information form Is the Google Inc. buying price (and your selling value) based on each individual piece of information or all information on the form? ○ Each individual piece of information ○ All information on the form

Page 8
Enter a selling value between $0.00 and $5.00, in the format X.XX, for the information presented in the sample form. [ ] If you wish to opt out of the study, you may do so. To opt out please select the Opt Out option and click >>. ○ Opt Out

Page 9 (if information was sold)
SOLD! Your value, $[*participant's WTA*], is less than the randomly drawn value, $[*system generated WTA*]. Therefore, you will earn an additional $[*Random WTA minus the WTA entered*] at the end of this experiment. Please proceed to the next page and enter the requested information.

Page 9 (if information was not sold)
NO SALE Unfortunately your selling price, $[*participant's WTA*] was greater than the randomly drawn value, $[*system generated WTA*]. You are not required to enter your information. Please proceed to the next page.

Table A6: Study 3 Manipulations	
Secondary Use	
Secondary Use	will distribute the information you provide to outside marketing

	and advertising agencies for various purposes
No Secondary Use	will not distribute information to any third parties. The information will be for internal application use only
Identifying Information	
Identifying Information	will store identifying information, such as name, email, and phone number, with the shopping information you provide
No Identifying Information	will not store identifying information, such as name, email, and phone number, with the shopping information you provide

Table A7: Study 3 Outline
Page 1
Before you begin, please watch the brief video below. Following the video, you will be asked several questions regarding its content. [*Embedded video*]

Page 2 (Video quiz)
Online businesses are able to perform _____ (varying the price of a product according to who the buyer is), with the information they obtain from consumers. ○ Price matching ○ Price equity ○ Price discrimination ○ Fair pricing Advertising agencies are capable of targeting unwanted advertisements with the information they are able to obtain. ○ True ○ False With the distribution of private information to third parties, there are more targets for hackers and identity thieves to attack. Thus, people experience an increased chance of experiencing _____. ○ Harm ○ Inconvenience ○ Both harm and inconvenience Sharing private information over the Internet leads to greater vulnerability to which of the following events?

- Identity theft
- Hacked online accounts
- Data loss
- All of the above

Which of the following is considered identifying information and can magnify the effects of a security breach?

- Full name
- Shopping habits
- Hobbies
- Special interests

Page 3

Welcome

Google Inc. is currently developing a new shopping application and wishes to pay its users for the information they provide when registering for the application. However, Google Inc. does not know how much to compensate users for their information. In order to capture an appropriate compensation value, we will ask you to enter your selling value for ***your shopping information***. The value entered must be between $0.00 and $5.00. This value will be referred to as your selling price.

The application Google is developing will require users to enter shopping information about themselves. A sample of the information Google Inc. may request appears later. Due to a nondisclosure agreement with Google Inc., we cannot disclose the specifics of the new application. However, the application will provide a quality service for its users.

You may opt out of participation at any time.

Page 4

Welcome

We must inform you that ***Google Inc. [Secondary Use]*** Disclosing your information represents consent for Google Inc. to share the information with other advertising and marketing companies.

Page 5

Welcome

Shown below is a brief list of possible information items you will provide when registering for an account. In addition to the shopping information, ***you [Identifying Information]***. This list is

not comprehensive and Google Inc. reserves the right to request shopping information that is not shown below.

[*List of Shopping Information Items*]

Page 6

Welcome

In summary, Google Inc. is creating a new application in which
Users *provide shopping information*
Google *[Secondary Use]*
Users provide [*Identifying Information*]

Compensation for your information can fall between $0.00 and $5.00. You may opt out at any time by closing your web browser.

Page 7

Information Selling Instructions

****IMPORTANT**** Before you receive your payment from Google Inc., Google Inc. will verify the information you provide for truthfulness.

After viewing the list of information Google Inc. may request, you have two options.

1. You may enter your selling price for the information Google Inc. requests. **Remember that the value entered is your selling price for all of your information, not individual pieces.**
2. You may opt out if you do not wish to participate or if your selling price is greater than $5.00. A opt out option is available on the selling price page. If you choose to opt out, you must provide a reason for doing so.

If you choose to participate and enter a selling price then you will type your selling value into a text box. The value must be between $0.00 and $5.00 and in the format X.XX. The following occurs after you submit your selling price:

Google Inc.'s information buying algorithm calculates a buying price between $0.00 and $5.00.

If Google Inc.'s buying price is greater than or equal to your selling price, you will sell your information to Google Inc. for the buying price and must provide the information Google Inc. requests. If Google Inc.'s buying price is less than your selling price, you will not sell your information and do not provide your information to Google Inc.

Example 1:
Your selling value is $1.00, Google Inc. buying price is $2.50 ⇒ You will sell your information for $2.50.

Example 2:
Your selling value is $3.00, Google Inc. buying price is $2.00 ⇒ You will **NOT** sell your information.

Example 3:
Your selling value is $2.50, Google Inc. buying price is $2.50 ⇒ You will sell your information for $2.50.

You will receive course credit for participating in this experiment and finishing the survey at the end. If you sell your information to Google Inc., you will receive the course credit and the buying price. If you do not sell your information to Google Inc. or choose to opt out, you will only receive the course credit.

Page 8

Information Selling Instructions

We will now demonstrate how pricing works. It is in your best interest to accurately state your true valuation as your selling price for your information. The following are two examples of why:

Example 1: What happens if your stated selling price is HIGHER than your true value:
Imagine you value your information at $3.00, but you enter a selling value of $4.50. We will say that the Google Inc. buying price is $4.25.

Since the buying price, $4.25, is less than your selling price, $4.50, you will not sell the information to Google Inc. and will not earn the $4.25. Therefore, you will miss the opportunity to sell your information for a price you deem as reasonable.

Example 2: What happens if your stated selling price is LOWER than your true value:
Imagine you value your information at $1.75, but you enter a selling value of $0.75. We will say that the Google Inc. buying price is $1.00.

Due to your selling price being lower than the Google Inc. buying price, you must sell your information to Google Inc., even though you value the information much more than the $1.00 you will receive. You will forfeit your information for less than what you think it is worth.

Page 9

Information Selling Instructions Quiz

Before proceeding, we wish to ensure you understand all of the instructions clearly. Below are four example scenarios, please choose the best answers to the questions:
Scenario 1: Your stated selling value is $1.50. The Google Inc. buying price is $2.50. What will happen next?

- You will not sell your information or complete the form

- You will sell your information for $1.50 and complete the form
- You will sell your information for $2.50 and complete the form

Scenario 2: Your stated selling value is $1.50. The Google Inc. buying price is $1.00. What will happen next?

- You will not sell your information or complete the form
- You will sell the information for 0.50 and complete the form
- You will sell the information for 1.50 and complete the form

Scenario 3: Google Inc.'s buying price is $0.75 and your selling value is $0.25. What must you do after winning?

- Do not complete the information form
- The information sells so you must complete the information form

Is the Google Inc. buying price (and your selling value) based on each individual piece of information or all information on the form?

- Each individual piece of information
- All information on the form

Page 10

Enter a selling value between $0.00 and $5.00, in the format X.XX, for the information presented in the sample form.

If you wish to opt out of the study, you may do so. To opt out please select the Opt Out option and click >>.

- Opt Out

Page 11 (if information was sold)

SOLD!

Your value, $[*participant's WTA*], is less than the randomly drawn value, $[*system generated WTA*]. Therefore, you will earn an additional $[*Random WTA minus the WTA entered*] at the end of this experiment. Please proceed to the next page and enter the requested information.

Page 11 (if information was not sold)

NO SALE

Unfortunately your selling price, $[*participant's WTA*] was greater than the randomly drawn value, $[*system generated WTA*]. You are not required to enter your information. Please proceed to the next page.

Table A8: Post-Experiment Survey Items
Demographics
Sex
Age
Highest level of education
Average daily Internet usage
Some websites ask you to register with the website by providing personal information. When asked for such information, how often do you falsify the information? (5-point Likert scale)
How frequently have you personally been the victim of what you felt was an improper invasion of privacy?
Additional Measures
Do you trust Google Inc. to follow through with what they tell consumers?
How does the inclusion of Name, Date of birth, and Email with other private information affect the risk associated with disclosing your private information?
How does the knowledge that Google Inc. will provide your private information to a third party affect the risk associated with disclosing your private information?

**All items use a 7-point Likert-type scale unless otherwise specified.

Table A9: Breakdown of Participants Removed from Each Study

Population	Study	Total Obs	Failed Attention or Manipulation Checks	Incomplete	Opt Out	Usable Obs
Student	1	394	26 (6.60%)	66 (16.75%)	2 (0.51%)	300 (76.14%)
Student	2	270	23 (8.52%)	41 (15.19%)	4 (1.48%)	202 (74.81%)
Student	3	175	12 (6.86%)	20 (11.43%)	3 (1.71%)	140 (80.00%)
AMT	2	450	8 (1.78%)	0	6 (1.33%)	436 (96.89%)
AMT	3	200	12 (6.00%)	0	6 (3.00%)	182 (91.00%)

Table A10: Breakdown of Participants Removed from Each Study by Treatment

Population	Study	Total Obs	Failed Attention or Manipulation Checks	Incomplete	Opt Out	Usable Obs
Student	**1**	**394**	**26**	**66**	**2**	**300**
Med-SU-Id		47	4	8	2	33
Med-No SU-Id		57	7	7	0	43

Med-SU-No Id		52	3	9	0	40
Med-No SU- No Id		46	1	10	0	35
Shop-SU-Id		50	3	8	0	39
Shop-No SU-Id		48	1	9	0	38
Shop-SU-No Id		47	5	7	0	35
Shop-No SU-No Id		47	2	8	0	37
Student	**2**	**270**	**23**	**41**	**4**	**202**
Med-SU-Id		33	3	7	2	21
Med-No SU-Id		36	1	8	0	27
Med-SU-No Id		39	5	6	1	27
Med-No SU- No Id		36	4	6	1	25
Shop-SU-Id		33	4	4	0	25
Shop-No SU-Id		29	2	3	0	24
Shop-SU-No Id		32	3	5	0	24
Shop-No SU-No Id		32	1	2	0	29
Student	**3**	**175**	**12**	**20**	**3**	**140**
Shop-SU-Id		44	7	3	2	32
Shop-No SU-Id		44	2	6	0	36
Shop-SU-No Id		41	1	5	1	34
Shop-No SU-No Id		46	2	6	0	38
AMT	**2**	**450**	**8**	**0**	**6**	**436**
Med-SU-Id		61	0	0	3	58
Med-No SU-Id		52	2	0	1	49
Med-SU-No Id		65	0	0	0	65
Med-No SU- No Id		55	1	0	0	54
Shop-SU-Id		53	2	0	2	49
Shop-No SU-Id		58	0	0	0	58
Shop-SU-No Id		55	1	0	0	54
Shop-No SU-No Id		51	2	0	0	49
AMT	**3**	**200**	**12**	**0**	**6**	**182**
Shop-SU-Id		50	2	0	2	46
Shop-No SU-Id		51	2	0	3	46
Shop-SU-No Id		51	5	0	1	45
Shop-No SU-No Id		48	3	0	0	45

Note: Med = Medical Context; Shop = Shopping Context; SU = Secondary Use; Id = Identifying Information

Table A11: Study 1 IPC Descriptive Statistics, Reliability, and Validity

Construct	**Composite**	**AVE**	**Access**	**Awareness**	**Collection**	**Control**	**Error**	**Risk**	**Sec. Use**	**Trust**
Access	0.943	0.846	*0.920*							
Awareness	0.892	0.734	0.619	*0.857*						
Collection	0.888	0.726	0.738	0.684	*0.852*					
Control	0.893	0.736	0.630	0.839	0.682	*0.858*				
Error	0.913	0.778	0.667	0.493	0.598	0.471	*0.882*			
Risk	0.863	0.611	0.612	0.612	0.668	0.665	0.507	*0.782*		
Secondary Use	0.932	0.820	0.791	0.637	0.762	0.644	0.589	0.681	*0.906*	
Trust	0.864	0.680	-0.304	-0.171	-0.280	-0.196	-	-0.297	-0.383	*0.825*
Construct	**Cronbach's**	**R^2**	**Mean**	**Std. Error**	**T-value**					
Access	0.909	0.857	0.926	0.010	94.029***					
Awareness	0.818	0.675	0.821	0.024	34.293***					
Collection	0.810	0.830	0.912	0.012	76.923***					
Control	0.820	0.736	0.860	0.021	40.894***					
Error	0.857	0.809	0.898	0.018	49.278***					
Secondary Use	0.890	0.825	0.909	0.013	68.478***					
Risk Beliefs	0.788	0.667	0.817	0.021	38.871***					
Trust Beliefs	0.765	0.147	-0.383	0.099	3.890***					
Interaction M.	0.918	0.939	0.969	0.004	269.52***					
Information M.	0.904	0.772	0.878	0.017	50.995***					

Table A12: OLS Regression with IPC as Dependent Variable		
Variables	**(1)** **Study 1 Baseline**	**(2)** **Study 2 Baseline**
Gender	-0.112 (0.081)	-0.113 (0.081)
False Information	0.058 (0.045)	0.058 (0.045)
Web Usage	0.098** (0.047)	0.098** (0.047)
Breach History	0.036 (0.031)	0.037 (0.031)
Study 1 Indicator		0.011 (0.091)
Study 2 Indicator	-0.010 (0.091)	
Study 3 Indicator	0.224** (0.102)	0.235** (0.108)
Constant	0.070 (0.074)	0.060 (0.083)
F-value	3.510***	3.510***

Standard errors in parentheses, *$p \leq 0.10$, **$p \leq 0.05$, ***$p \leq 0.01$.
All models use robust standard error estimation.

Table A13: AMT Study 3 Heteroscedasticity Robustness Models				
	GMM		**Symmetric LS**	
Variables	**(1)**	**(2)**	**(3)**	**(4)**
Gender	-0.086 (0.216)	-0.107 (0.216)	-0.086 (0.229)	-0.107 (0.229)
Age	0.081 (0.091)	0.089 (0.096)	0.081 (0.095)	0.089 (0.100)
Education	-0.027 (0.079)	-0.008 (0.080)	-0.027 (0.082)	-0.008 (0.084)
False Information	0.063 (0.118)	0.051 (0.117)	0.063 (0.129)	0.051 (0.129)
Web Usage	-0.147 (0.089)	-0.144 (0.088)	-0.147 (0.093)	-0.144 (0.092)
Breach History	-0.063 (0.076)	-0.082 (0.078)	-0.063 (0.081)	-0.082 (0.083)
Secondary Use		-0.119 (0.212)		-0.119 (0.226)
Identifying Information		0.318 (0.221)		0.318 (0.233)
Constant	4.223** (0.544)	4.109** (0.554)	4.223** (0.571)	4.109** (0.579)
Observations	182	182	182	182

Robust standard errors in parentheses, †$p \leq 0.10$, * $p \leq 0.05$, ** $p \leq 0.01$.

APPENDIX B: Essay 3 Mediation Analysis

We performed mediation analysis to provide further evidence of the structure of our hypothesized framework in which data breach and meaningful use attestation affect patient outcomes through the process of care. Due to the use of panel data we restructure our data by using the index or binary variable for each individual year. For example, a given hospital that reported their process of care measures for all four years will have a 2012 process of care variable, 2013 process of care variable, 2014 process of care variable, and a 2015 process of care variable. A given hospital that reported their process of care measure for 2012 and no other years will only have a 2012 process of care variable. We then fit the structural equation model by estimating the following effects: (1) the effect of a hospital's data breach indicator for a given year on the hospital's data breach indicator for the subsequent year (e.g., the effect of 2012 data breach indicator on the 2013 data breach indicator); (2) the effect of a hospital's data breach indicator for a given year on the hospital's process of care index for the same year (e.g., the effect of 2012 data breach indicator on 2012 process of care index); (3) the effect of a hospital's data breach indicator for a given year on the hospital's process of care index for the subsequent year (e.g., the effect of 2012 data breach indicator on 2013 process of care index); (4) the effect of a hospital's meaningful use indicator for a given year on the hospital's meaningful use indicator for the subsequent year (e.g., the effect of 2012 meaningful use indicator on the 2013 meaningful use indicator); (5) the effect of a hospital's meaningful use indicator for a given year on the hospital's process of care index for the same year (e.g., the effect of 2012 meaningful use indicator on 2012 process of care index); (6) the effect of a hospital's meaningful use indicator for a given year on the hospital's process of care index for the subsequent year (e.g., the effect of 2012 meaningful use indicator on 2013 process of care index); (7) the effect of a hospital's

process of care index for a given year on the hospital's process of care index for the subsequent year (e.g., the effect of 2012 process of care index on the 2013 process of care index); and finally (8) the effect of a hospital's patient outcome index for a given year on the hospital's patient outcome index for the subsequent year (e.g., the effect of 2012 patient outcome index on 2013 patient outcome index).

The results from our model provide that there is a significant direct effect between meaningful use attestation and process of care as well as experiencing a data breach and process of care. We also find that process of care directly affects patient outcome. Next, we find that there exists an indirect effect between meaningful use attestation and patient outcome. The results also show that there is an indirect effect between experiencing a data breach and patient outcome. Finally, we find that meaningful use attestation and process of care have significant total effects on patient outcome. Thus, we further support our hypothesized framework and interpretation of the effects in our 2SLS model. Tables A14, A15, and A16 provide the results from our structural equation model estimation.

Table A14. Structural Equation Model Estimation

Variable	Coefficient	Std. Error	*p*
Meaningful Use 2013			
Meaningful Use 2012	0.108	0.014	0.000
Constant	0.889	0.014	0.000
Meaningful Use 2014			
Meaningful Use 2013	0.374	0.020	0.000
Constant	0.625	0.020	0.000
Meaningful Use 2015			
Meaningful Use 2014	0.247	0.033	0.000
Constant	0.750	0.020	0.000
Breach 2013			
Breach 2012	0.086	0.028	0.002
Constant	0.014	0.004	0.001
Breach 2014			
Breach 2013	-0.021	0.038	0.585
Constant	0.021	0.005	0.000
Breach 2015			
Breach 2014	0.030	0.038	0.432
Constant	0.026	0.005	0.000
Process of Care 2012			

Meaningful Use 2012	0.002	0.092	0.983
Breach 2012	0.280	0.136	0.039
Constant	1.151	0.090	0.000
Process of Care 2013			
Process of Care 2012	0.663	0.025	0.000
Meaningful Use 2012	0.116	0.070	0.098
Meaningful Use 2013	0.232	0.163	0.153
Breach 2012	-0.040	0.101	0.691
Breach 2013	0.082	0.120	0.495
Constant	0.062	0.160	0.699
Process of Care 2014			
Process of Care 2013	0.701	0.025	0.000
Meaningful Use 2013	0.404	0.254	0.111
Meaningful Use 2014	-0.010	0.358	0.978
Breach 2013	0.167	0.164	0.307
Breach 2014	0.262	0.145	0.070
Constant	0.665	0.310	0.032
Process of Care 2015			
Process of Care 2014	0.759	0.021	0.000
Meaningful Use 2014	-0.034	0.198	0.864
Meaningful Use 2015	0.073	0.198	0.712
Breach 2014	-0.056	0.091	0.538
Breach 2015	0.044	0.081	0.585
Constant	0.192	0.243	0.429
Patient Outcome 2012			
Process of Care 2012	0.368	0.079	0.000
Meaningful Use 2012	0.064	0.215	0.765
Breach 2012	0.296	0.318	0.353
Constant	9.296	0.229	0.000
Patient Outcome 2013			
Patient Outcome 2012	1.024	0.15	0.000
Process of Care 2012	-0.018	0.047	0.708
Process of Care 2013	0.055	0.048	0.246
Meaningful Use 2012	0.049	0.099	0.621
Meaningful Use 2013	0.363	0.229	0.113
Breach 2012	0.059	0.142	0.678
Breach 2013	-0.025	0.169	0.881
Constant	-0.422	0.265	0.111
Patient Outcome 2014			
Patient Outcome 2013	1.044	0.012	0.000
Process of Care 2013	0.018	0.047	0.698
Process of Care 2014	0.016	0.046	0.734
Meaningful Use 2013	0.246	0.236	0.297
Meaningful Use 2014	-0.024	0.332	0.943
Breach 2013	0.197	0.151	0.195
Breach 2014	-0.031	0.134	0.818
Constant	-0.074	0.312	0.812
Patient Outcome 2015			
Patient Outcome 2014	0.963	0.018	0.000
Process of Care 2014	0.006	0.082	0.936
Process of Care 2015	0.023	0.083	0.781
Meaningful Use 2014	-0.157	0.488	0.748
Meaningful Use 2015	0.063	0.487	0.897
Breach 2014	0.163	0.225	0.469
Breach 2015	0.066	0.199	0.741
Constant			
Variance(Meaningful Use 2013)	0.008	0.000	
Variance(Meaningful Use 2014)	0.003	0.000	
Variance(Meaningful Use 2015)	0.004	0.000	

Variance(Breach 2013)	0.015	0.001	
Variance(Breach 2014)	0.020	0.001	
Variance(Breach 2015)	0.026	0.001	
Variance(Process of Care 2012)	0.360	0.017	
Variance(Process of Care 2013)	0.196	0.009	
Variance(Process of Care 2014)	0.368	0.018	
Variance(Process of Care 2015)	0.146	0.007	
Variance(Patient Outcome 2012)	1.976	0.094	
Variance(Patient Outcome 2013)	0.387	0.018	
Variance(Patient Outcome 2014)	0.317	0.015	
Variance(Patient Outcome 2015)	0.888	0.042	

Table A15. Indirect Effects from SEM

Variable	**Coefficient**	**Std. Error**	***p***
Meaningful Use 2014			
Meaningful Use 2012	0.040	0.006	0.000
Meaningful Use 2015			
Meaningful Use 2012	0.010	0.002	0.000
Meaningful Use 2013	0.092	0.005	0.000
Breach 2014			
Breach 2012	-0.002	0.003	0.591
Breach 2015			
Breach 2012	-0.000	0.000	0.657
Breach 2013	-0.001	0.001	0.585
Process of Care 2013			
Meaningful Use 2012	0.026	0.063	0.679
Breach 2012	0.192	0.091	0.034
Process of Care 2014			
Meaningful Use 2012	0.043	0.024	0.071
Meaningful Use 2013	-0.004	0.000	0.000
Breach 2012	0.014	0.015	0.348
Breach 2013	-0.005	0.010	0.585
Process of Care 2015			
Meaningful Use 2012	0.032	0.019	0.103
Meaningful Use 2013	0.298	0.193	0.122
Meaningful Use 2014	0.011	0.272	0.969
Breach 2012	0.011	0.011	0.344
Breach 2013	0.124	0.124	0.319
Breach 2014	0.200	0.110	0.068
Patient Outcome 2012			
Meaningful Use 2012	0.001	0.034	0.983
Breach 2012	0.103	0.055	0.060
Patient Outcome 2013			
Process of Care 2012	0.414	0.081	0.000
Meaningful Use 2012	0.113	0.225	0.614
Meaningful Use 2013	0.013	0.009	0.153
Breach 2012	0.408	0.331	0.215
Breach 2013	0.005	0.007	0.495
Patient Outcome 2014			
Patient Outcome 2012	1.030	0.015	0.000
Patient Outcome 2013	1.006	0.012	0.000
Process of Care 2012	0.410	0.094	0.000
Process of Care 2013	0.058	0.049	0.246
Meaningful Use 2012	0.198	0.255	0.437
Meaningful Use 2013	0.394	0.239	0.100
Meaningful Use 2014	-0.000	0.006	0.978
Breach 2012	0.509	0.375	0.175

Breach 2013	-0.017	0.177	0.923
Breach 2014	0.004	0.002	0.070
Patient Outcome 2015			
Patient Outcome 2012	1.030	0.015	0.000
Patient Outcome 2013	1.006	0.011	0.000
Process of Care 2012	0.410	0.094	0.000
Process of Care 2013	0.073	0.066	0.267
Process of Care 2014	0.032	0.044	0.462
Meaningful Use 2012	0.186	0.362	0.450
Meaningful Use 2013	0.574	0.324	0.077
Meaningful Use 2014	-0.008	0.321	0.979
Meaningful Use 2015	0.002	0.005	0.712
Breach 2012	0.491	0.362	0.175
Breach 2013	0.174	0.224	0.439
Breach 2014	-0.018	0.130	0.885
Breach 2015	0.001	0.002	0.585

Table A16. Total Effects from SEM

Variable	Coefficient	Std. Error	*p*
Meaningful Use 2013			
Meaningful Use 2012	0.108	0.014	0.000
Meaningful Use 2014			
Meaningful Use 2012	0.040	0.006	0.000
Meaningful Use 2013	0.374	0.020	0.000
Meaningful Use 2015			
Meaningful Use 2012	0.010	0.002	0.000
Meaningful Use 2013	0.092	0.005	0.000
Meaningful Use 2014	0.247	0.032	0.000
Breach 2013			
Breach 2012	0.086	0.028	0.002
Breach 2014			
Breach 2012	-0.002	0.003	0.591
Breach 2013	-0.021	0.038	0.585
Breach 2015			
Breach 2012	-0.000	0.000	0.657
Breach 2013	-0.001	0.001	0.585
Breach 2014	0.030	0.038	0.432
Process of Care 2012			
Meaningful Use 2012	0.002	0.092	0.983
Breach 2012	0.279	0.136	0.039
Process of Care 2013			
Process of Care 2012	0.663	0.025	0.000
Meaningful Use 2012	0.142	0.091	0.119
Meaningful Use 2013	0.233	0.163	0.153
Breach 2012	0.152	0.135	0.259
Breach 2013	0.082	0.120	0.495
Process of Care 2014			
Process of Care 2013	0.701	0.025	0.000
Meaningful Use 2012	0.043	0.024	0.071
Meaningful Use 2013	0.401	0.254	0.114
Meaningful Use 2014	-0.010	0.358	0.978
Breach 2012	0.014	0.015	0.348
Breach 2013	0.162	0.164	0.324
Breach 2014	0.262	0.145	0.070
Process of Care 2015			
Process of Care 2014	0.760	0.021	0.000
Meaningful Use 2012	0.032	0.019	0.103

Meaningful Use 2013	0.298	0.193	0.122
Meaningful Use 2014	-0.024	0.336	0.944
Meaningful Use 2015	0.073	0.198	0.712
Breach 2012	0.011	0.011	0.344
Breach 2013	0.124	0.124	0.319
Breach 2014	0.144	0.143	0.314
Breach 2015	0.044	0.081	0.585
Patient Outcome 2012			
Process of Care 2012	0.368	0.079	0.000
Meaningful Use 2012	0.065	0.218	0.766
Breach 2012	0.399	0.322	0.215
Patient Outcome 2013			
Patient Outcome 2012	1.024	0.015	0.000
Process of Care 2012	0.396	0.094	0.000
Process of Care 2013	0.055	0.048	0.246
Meaningful Use 2012	0.162	0.242	0.505
Meaningful Use 2013	0.376	0.229	0.101
Breach 2012	0.468	0.359	0.192
Breach 2013	-0.021	0.169	0.902
Patient Outcome 2014			
Patient Outcome 2012	1.069	0.016	0.000
Patient Outcome 2013	1.045	0.012	0.000
Process of Care 2012	0.426	0.098	0.000
Process of Care 2013	0.076	0.068	0.267
Process of Care 2014	0.016	0.046	0.734
Meaningful Use 2012	0.198	0.255	0.437
Meaningful Use 2013	0.641	0.336	0.057
Meaningful Use 2014	-0.024	0.332	0.942
Breach 2012	0.509	0.375	0.175
Breach 2013	0.179	0.233	0.440
Breach 2014	-0.027	0.134	0.842
Patient Outcome 2015			
Patient Outcome 2012	1.030	0.015	0.000
Patient Outcome 2013	1.006	0.012	0.000
Patient Outcome 2014	0.963	0.018	0.000
Process of Care 2012	0.410	0.094	0.000
Process of Care 2013	0.073	0.066	0.267
Process of Care 2014	0.039	0.094	0.675
Process of Care 2015	0.023	0.083	0.781
Meaningful Use 2012	0.186	0.246	0.450
Meaningful Use 2013	0.574	0.342	0.077
Meaningful Use 2014	-0.165	0.584	0.777
Meaningful Use 2015	0.065	0.488	0.895
Breach 2012	0.491	0.362	0.175
Breach 2013	0.174	0.224	0.439
Breach 2014	0.144	0.259	0.578
Breach 2015	0.067	0.199	0.737

REFERENCES

Abraham, J., and Reddy, M. C. 2008. "Moving Patients Around: A Field Study of Coordination Between Clinical and Non-Clinical Staff in Hospitals," in *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, ACM: San Diego, CA, pp. 225-228.

Acquisti, A. 2004. "Privacy in Electronic Commerce and the Economics of Immediate Gratification," in *Proceedings of the Fifth ACM Conference on Electronic Commerce*, Breese, J., Feigenbaum, J., and Seltzer, M. (eds.), pp. 21-29.

Acquisti, A., Friedman, A., and Telang, R. 2006. "Is there a cost to privacy breaches? An event study," *International Conference on Information Systems Proceedings*, pp. 1563-1580.

Acquisti, A., & Gross, R. 2006. "Imagined communities: Awareness, information sharing, and privacy on the Facebook," in *Privacy enhancing technologies*, Springer Berlin Heidelberg, pp. 36-58.

Acquisti, A. and Grossklags, J. 2005. "Privacy and Rationality in Individual Decision Making," *IEEE Security and Privacy* (2), pp. 24-30.

Acquisti, A., Leslie, J. K., and Loewenstein, G. 2013. "What is Privacy Worth?," *The Journal of Legal Studies* (42:2), pp. 249-274.

Acquisti, A., Brandimarte, L., and Loewenstein, G. 2015. "Privacy and Human Behavior in the Age of Information," *Science* (347:6221), pp. 509-514.

Agarwal, R., Gao, G. G., DesRoches, C., and Jha, A. K. 2010. "The Digital Transformation of Healthcare: Current Status and the Road Ahead," *Information Systems Research* (21:4), pp. 796-809.

Akers, R. L. 1990. "Rational Choice, Deterrence, and Social Learning Theory in Criminology: The Path Not Taken," *Journal of Criminal Law and Criminology* (81), pp. 653-676.

Anderson, C. L., and Agarwal, R. 2011. "The Digitization of Healthcare: Boundary Risks, Emotion, and Consumer Willingness to Disclose Personal Health Information," *Information Systems Research* (22:3), pp. 469-490.

Angst, C. M., and Agarwal, R. 2009. "Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion," *MIS Quarterly* (33:2), pp. 339-370.

Angst, C. M., Block, E. S., D'Arcy, J., and Kelley, K. 2017. "When do IT Security Investments Matter? Accounting for the Influence of Institutional Factors in the Context of Healthcare Data Breaches," *MIS Quarterly* (41:3), pp. 893-916.

Angst, C. M., Devaraj, S., and D'Arcy, J. 2012. "Dual Role of IT-Assisted Communication in Patient Care: A Validated Structure-Process-Outcome Framework," *Journal of Management Information Systems* (29:2), pp. 257-292.

Angst, C. M., Devaraj, S., Queenan, C. C., and Greenwood, B. 2011. "Performance Effects Related to the Sequence of Integration of Healthcare Technologies," *Production and Operations Management* (20:3), pp. 319-333.

Appari, A., and Johnson, M. E. 2010. "Information Security and Privacy in Healthcare: Current State of Research," *International Journal of Internet and Enterprise Management* (6:4), pp. 279-314.

Arora, A., Telang, R., and Xu, H. 2008. "Optimal Policy for Vulnerability Disclosure," *Management Science* (54:4), pp. 642-656.

Arora, A., Krishnan, R., Telang, R., and Yang, Y. 2010. "An Empirical Analysis of Software Vendors' Patch Release Behavior: Impact of Vulnerability Disclosure," *Information Systems Research* (21:1), pp. 115-132.

Balebako, R., Schaub, F., Adjerid, I., Acquisti, A., and Cranor, L.F. 2015. "The Impact of Timing on the Salience of Smartphone App Privacy Notices," *Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices*.

Bathon, J. 2013. "How Little Data Breaches Cause Big Problems for Schools," *The Journal*, November 5.

Baum, C. F., Schaffer, M. E., and Stillman, S. 2003. "Instrumental Variables and GMM: Estimation and Testing," *Stata Journal* (3:1), pp. 1-31.

Becker, G. M., DeGroot, M. H., and Marschak, J. 1964. "Measuring utility by a single-response sequential method," *Behavioral Science* (9:3), pp. 226-232.

Bell, L. M., Grundmeier, R., Localio, R., Zorc, J., Fiks, A. G., Zhang, X., Stephens, T. B., Swietlik, M., and Guevara, J. P. 2010. "Electronic Health Record-Based Decision Support to Improve Asthma Care: A Cluster-Randomized Trial," *Pediatrics* (125), pp. 770-777.

Berendt, B., Gunther, O., and Spiekermann, S. 2005. "Privacy in E-Commerce: Stated Preferences vs. Actual Behavior," *Communications of the ACM* (48:4), pp. 101-106.

Bever, Lindsey. 2018. "Why Apple co-founder Steve Wozniak is joining the #DeleteFacebook movement," Washington Post, published April 9th. Available online: https://www.washingtonpost.com/news/the-switch/wp/2018/04/09/why-apple-co-founder-steve-wozniak-is-joining-the-deletefacebook-movement/?utm_term=.23fd8cffcf64.

Blumstein, A., Cohen, J., and Nagin, D. 1978. "Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates." *National Academy of Sciences*.

Bulgurcu, B., Cavusoglu, H., and Benbasat, I. 2010. "Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness," *MIS Quarterly* (34:3), pp. 523-548.

Buntin, M. B., Burke, M. F., Hoaglin, M. C., and Blumenthal, D. 2011. "The Benefits of Health Information Technology: A Review of the Recent Literature Shows Predominantly Positive Results," *Health Affairs* (30:3), pp. 464-471.

Burke, A. J. 2015. "Data, Data Everywhere, But Not a Bit Your Own," *Huffington Post,* June 25.

Burns, E. 2017. "Cost of a Data Breach Soars to 20% of Revenue as Hacking Goes 'Classic' and Corporate," *Computer Business Review,* January 31.

Calvin, A. O., Lindy, C. M., and Clingon, S. L. 2009. "The Cardiovascular Intensive Care Unit Nurse's Experience with End-of-Life Care: A Qualitative Descriptive Study," *Intensive and Critical Care Nursing* (25:4), pp. 214-220.

Campbell, K., Gordon, L. A., Loeb, M. P., and Zhou, L. 2003. "The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market," *Journal of Computer Security* (11), pp. 431-448.

Caudill, E. M., and Murphy, P. E. 2000. "Consumer Online Privacy: Legal and Ethical Issues," *Journal of Public Policy & Marketing* (19:1), pp. 7-19.

Cavusoglu, H., Birendra, M., and Raghunathan, S. 2004. "The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers," *International Journal of Electronic Commerce* (9:1), pp. 70-104.

Chellappa, R. K., and Sin, R. G. 2005. "Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma," *Information Technology and Management* (6:2-3), pp. 181-202.

Cherney. M. A. 2018. "Facebook stock drops roughly 20%, loses $120 billion in value after warning that revenue growth will take a hit," *Market Watch,* July 26.

Chisholm, C. D., Collison, E. K., Nelson, D. R., and Cordell, W. H. 2000. "Emergency Department Workplace Interruptions Are Emergency Physicians "Interrupt-driven" and "Multitasking"?," *Academic Emergency Medicine* (7:11), pp. 1239-1243.

Cisco Systems. 2017. "2017 Annual Cybersecurity Report," Retrieved from http://www.cisco.com/c/en/us/products/security/security-reports.html.

Cohen, M., A. 2000. "Empirical research on the deterrent effect of environmental monitoring and enforcement," *Environmental Law Reporter News and Analysis* (30:4), pp. 10245-10252.

Compeau, D., Marcolin, B., Kelley, H., and Higgins, C. 2012. "Generalizability of Information Systems Research Using Student Subjects – A Reflection on Our Practices and Recommendations for Future Research," *Information Systems Research* (23:4), pp. 1093-1109.

Confessore, N. 2018. "Cambridge Analytica and Facebook: The Scandal and Fallout So Far," *New York Times,* April 4.

Coursey, D. L., Hovis, J. L., and Schulze, W. D. 1987. "The Disparity between Willingness to Accept and Willingness to Pay Measures of Value," *The Quarterly Journal of Economics,* pp. 679-690.

Culnan, M. J. 1993. "How Did They Get My Name?: An Exploratory Investigation of Consumer Attitudes Toward Secondary Information Use," *MIS Quarterly* (17:3), pp. 341-363.

Culnan, M. J., and Armstrong, P. K. 1999. "Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation," *Organization Science* (10:1), pp. 104-115.

Curtin, M., and Ayres, L. T. 2008. "Using science to combat data loss: Analyzing breaches by type and industry," *ISJLP* (4), pp. 569.

Cvrcek, D., Kumpost, M., Matyas, V., and Danezis, G. 2006. "A Study on the Value of Location Privacy," in *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society,* pp. 109-118.

Daly, J. 2016. "Expenses from the Home Depot and Target Data Breaches Surpass $500 Million," *Digital Transactions,* May 26.

Danezis, G., Lewis, S., and Anderson, R. J. 2005. "How Much is Location Privacy Worth?" in *Workshop on the Economics of Information Security* (5).

Dickler, J. 2018. "In the Wake of the Equifax Data Breach, Consumers More at Risk," *CNBC,* March 11.

Dinev, T., and Hart, P. 2006. "An Extended Privacy Calculus Transactions Model for E-Commerce Transactions," *Information Systems Research* (17:1), pp. 61-80.

Donaldson, C., Jones, M. A., Mapp, J. T., and Olson, A. J. 1998. "Limited Dependent Variables in Willingness to Pay Studies: Applications in Health Care," *Applied Economics* (30:5): pp. 667-677.

Druckman, J., and Kam, C. 2009. "Students as Experimental Participants: A Defense of the "Narrow Data Base"," *Institute for Policy Research Northwestern University Working Paper Series,* WP-09-05 Available at papers.ssrn.com/sol3/papers.cfm?abstract_id=1498843.

Edwards, B., Hofmeyr, S., and Forrest, S. 2016. "Hype and Heavy Tails: A Closer Look at Data Breaches," *Journal of Cybersecurity* (2:1), pp. 3-14.

Feldman, B. 2017. "The Unroll.me Controversy Is a Good Reminder That Tech Companies and Consumers Don't Understand Each Other," New York Magazine, April 26th, Available online at: http://nymag.com/selectall/2017/04/unroll-mes-lesson-legal-ass-covering-isnt-enough.html. Last Accessed April 27, 2017.

Fichman, R. G., Kohli, R. and Krishnan, R. 2011. "Editorial Review: The Role of Information Systems in Healthcare: Current Research and Future Trends," *Information Systems Research* (22:3), pp. 419-428.

Garg, A., Curtis, J., and Harper, H. 2003. "Quantifying the Financial Impact of IT Security Breaches," *Information Management and Computer Security* (11:2), pp. 74-83.

Gattiker, T. F., and Goodhue, D. L. 2005. "What Happens After ERP Implementation: Understanding the Impact of Inter-Dependence and Differentiation on Plant-Level Outcomes," *MIS Quarterly* (29:3), pp. 559-585.

Gatzlaff, K. M. and McCullough, K. A. 2010. "The Effect of Data Breaches on Shareholder Wealth," *Risk Management and Insurance Review* (13:1), pp. 61-83.

Gay, S. 2015. "Strategic News Bundling and Privacy Breach Disclosures," *Workshop on the Economics of Information Security (2016)*.

Gordon, L. A., Loeb, M. P., and Zhou, L. 2011. "The Impact of Information Security Breaches: Has There Been a Downward Shift in Costs?," *Journal of Computer Security* (19), pp. 33-56.

Gupta, A., Allen, L. A., Bhatt, D. L., Cox, M., DeVore, A. D., Heidenreich, P. A., Hernandez, A. F., Peterson, E. D., Matsouaka, R. A., Yancy, C. W., and Fonarrow, G. C. 2018. "Association of the Hospital Readmissions Reduction Program Implementation with Readmission and Mortality Outcomes in Heart Failure," *Journal of the American Medical Association* (3:1), pp. 44-53.

Hanemann, W. M. 1991. "Willingness to Pay and Willingness to Accept: How Much Can They Differ?," *The American Economic Review* (81:3), pp. 635-647.

Hann, I., Hui, K., Lee, S. T., and Png, I. 2007. "Overcoming Information Privacy Concerns: An Information Processing Theory Approach," *Journal of Management Information Systems* (24:2), pp. 13-42.

Harris, C. E., and Hammargren, R. L. 2016. "Establishing a Written Information Security program to address exposure," *University Business*, September 6.

Help Net Security. 2016. "Impact of Security Breaches on Consumer Trust," *Help Net Security*, May 12.

Hennessy, J. J., Howell, C. T., Millendorf, S. M., Overly, M. R., Rathburn, J. L., and Tantleff, A. K. 2018. "State Data Breach Notification Laws," *Foley and Lardner LLP*, January 17.

Henry, J., Pylypchuk, Y., Searcy, T., and Patel, V. 2016. "Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2015," *The Office of the National Coordinator for Health Information Technology* at https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php.

Hiller, J., McMullen, M. S., Chumney, W. M., and Baumer, D. L. 2011. "Privacy and Security in the Implementation of Health Information Technology (Electronic Health Records): US and EU Compared," *Boston University Journal of Science & Technology Law* (17), pp. 1-40.

Hoffman, S., and Podgurski, A. 2007. "In Sickness, Health, and Cyberspace: Protecting the Security of Electronic Private Health Information," *Boston College Law Review* (48:2), pp. 331-386.

Hoffman, D. L., Thomas, P. N., and Marcos, P. 1999. "Building Consumer Trust Online," *Communications of the ACM* (42:4), pp. 80-85.

Hong, W., and Thong, J. Y. L. 2013. "Internet Privacy Concerns: An Integrated Conceptualization and Four Empirical Studies," *MIS Quarterly* (37:1), pp. 275-298.

Hosmer, D. W., Jr., Lemeshow, S. A., and May, S. 2008. Applied Survival Analysis: Regression Modeling of Time to Event Data. 2nd ed. New York: Wiley.

Hovav, A. and D'Arcy, J. 2003. "The Impact of Denial-Of-Service Attack Announcements on the Market Value of Firms," *Risk Management and Insurance Review* (6:2), pp. 97-121.

Huang, C. D., Behara, R. S., Goo, J. 2014. "Optimal Information Security Investment in a Healthcare Information Exchange: An Economic Analysis," *Decision Support Systems* (61), pp. 1-11.

Huberman, B. A., Adar, E., and Fine, L. 2005. "Valuating Privacy," *IEEE Security and Privacy* (3:5), pp. 22-25.

Hui, K., Teo, H. H., and Lee, S. T. 2007. "The Value of Privacy Assurance: An Exploratory Field Experiment," *MIS Quarterly* (31:1) pp. 19-33.

Identity Theft Resource Center. 2016. "Identity Theft Resource Center Breach Report Hits Near Record High in 2015," *Identity Theft Resource Center,* January 25.

Insureon. 2015. "Nonprofit Nightmare: Data Breach Exposes 10,000 Donors' Financial Records," *Insureon*. November 4.

Jones, S. S., Rudin, R. S., Perry, T., and Shekelle, P. G. 2014. "Health Information Technology: An Updated Systematic Review with a Focus on Meaningful Use," *Annals of Internal Medicine* (160:1), pp. 48-54.

Julesz, B. 1995. "Dialogues on Perception," MIT Press.

Kadir, T., and Brady, M. 2001. "Saliency, Scale, and Image Description," *International Journal of Computer Vision* (45:2), pp. 83-105.

Kankanhali, A., Teo, H., Tan, C.Y., B., Wei, K. 2003. “An Integrative Study of Information Systems Security Effectiveness,” *International Journal of Information Management* (23), pp. 139-154.

Kannan, K., Rees, J., and Sridhar, S. 2007. “Market reactions to information security breach announcements: an empirical study,” *International Journal of Electronic Commerce* (12:1), pp. 69-91.

Kim, K. K., Joseph, J. G., and Ohno-Machado, L. 2015. “Comparison of Consumers’ Views on Electronic Data Sharing for Healthcare and Research,” *Journal of American Information Association* (22), pp. 821-830.

Klopfer, P. H., and Rubenstein, D. I. 1977. “The Concept of Privacy and Its Biological Basis,” *Journal of Social Issues* (33:3), pp. 52-65.

Ko, M., and Dorantes, C. 2006. “The impact of information security breaches on financial performance of the breached firms: an empirical investigation,” *Journal of Information Technology Management* (17:2), pp. 13-22.

Kruze, C. S., Smith, B., Vanderlinden, H., and Nealand, A. 2017. “Security Techniques for Electronic Health Records,” *Journal of Medical Systems* (41:8).

Kvochko, E., and Pant, R. 2015. “Why Data Breaches Don’t Hurt Stock Prices,” *Harvard Business Review,* March 31.

Kwon, J., and Johnson, M. E. 2014a. “Proactive Versus Reactive Security Investments in the Healthcare Sector,” *MIS Quarterly* (38:2), pp. 451-471.

Kwon, J., and Johnson, M. E. 2014b. “Meaningful Healthcare Security: Does “Meaningful-Use” Attestation Improve Information Security Performance?,” in *Workshop on the Economics of Information Security (WEIS) 2014, Penn State University*.

Lafky, D. B., and Horan, T. A. 2011. “Personal Health Records: Consumer Attitudes Toward Privacy and Security of Their Personal Health Information,” *Health Informatics Journal* (17:1), pp. 63-71.

Laube, S., and Bohme, R. 2016. “The economics of mandatory security breach reporting to authorities,” *Journal of Cybersecurity*.

Laufer, R. S., and Wolfe, M. 1977. “Privacy as a Concept and a Social Issue: A Mutlidimensional Developmental Theory,” *Journal of Social Issues* (33:3), pp. 22-42.

Lease, M., Hullman, J., Bigham, J. P, Bernstein, M., Kim, J., Lasecki, W. S., Bakhshi, S., Mitra, T., and Miller, R. C. “Mechanical Turk Is Not Anonymous,” SSRN. doi: 10.2139/ssrn.2228728.

Leonard, T. M., and Rubin, P. 2006. “Much Ado about Notification,” *Regulation* (29), pp. 44-50.

Lim, S. Y., Jarvenpaa, S. L., and Lanham, H. J. 2015. “Barriers to Interorganizational Knowledge Transfer in Post-Hospital Care Transitions: Review and Directions for Information Systems Research,” *Journal of Management Information Systems* (32:3), pp. 48-74.

List, J. A., and Shogren, J. F. 2002. “Calibration of Willingness to Accept,” *Journal of Environmental Economics and Management* (43), pp. 219-233.

Loch, K., Carr, H., and Warkentin, M. 1992. “Threats to Information Systems,” *MIS Quarterly* (16:2), pp. 173-186.

Madrigal, A.C. 2012 "How much is your data worth? Mmm somewhere between half a cent and $1,200", *The Atlantic*, published March 19, available online at: https://www.theatlantic.com/technology/archive/2012/03/how-much-is-your-data-worth-mmm-somewhere-between-half-a-cent-and-1-200/254730/

Malhotra, N., Kim, S., and Agarwal, J. 2004. "Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model," *Information Systems Research* (15:4), pp. 336-355.

McCallister, E., T. Grance, and K. Scarfone. 2010. "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)", Recommendations of the National Institute of Standards and Technology, NIST Special Publication 800-122.

McCullough, J. S., Casey, M., Moscovice, I., and Prasad, S. 2010. "The Effect of Health Information Technology on Quality in U.S. Hospitals," *Health Affairs* (29:4), pp. 647-659.

McCullough, J. S., Parente, S. T., and Town, T. 2016. "Health Information Technology and Patient Outcomes: The Role of Information and Labor Coordination," *RAND Journal of Economics* (48:1), pp. 207-236.

McDonald, J. F., and Moffitt, R. A. 1980. "The Uses of Tobit Analysis," *The Review of Economics and Statistics,* pp. 318-321.

McMillan, R. 2014. "The Hidden Privacy Threat of ... Flashlight Apps," *Wired,* October 20.

Mellott, M., Thatcher, J. B., and Roberts, N. 2013. "Electronic Medical Record Compliance and Continuity in Delivery of Care: An Empirical Investigation in a Combat Environment," *Health Systems* (2), pp. 147-161.

Meingast, M., Roosta, T., and Sastry, S. 2006. "Security and Privacy Issues with Healthcare Information Technology," *Engineering in Medicine and Biology Society, 2006. 28th Annual International Conference on the IEEE,* pp. 5453-5448.

Miller, A. R., and Tucker, C. 2009. "Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records," *Management Science* (55:7), pp. 1077-1093.

Mirhaydari, A. 2018. “Facebook stock recovers all $134B lost after Cambridge Analytica data scandal,” *CBS News,* May 10.

Nowak, G. J., and Phelps, J. 1992. “Understanding Privacy Concerns: An Assessment of Consumer’s Information-Related Knowledge and Beliefs,” *Journal of Direct Marketing* (6:4), pp. 28-39.

Ohm, P. 2010. “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,” *UCLA Law Review* (57), pp. 1701-1777.

Olenski, S. 2016. “For Consumers, Data Is a Matter of Trust, *Forbes,* April 18.

Pahnila, S., Siponen, M., and Mahmood, A. 2007. “Employees’ Behavior towards IS Security Policy Compliance,” *Proceedings of the 40th Hawaii International Conference on System Sciences,* pp. 156-166.

Peckham, C. 2016. “Medscape EHR Report 2016: Physicians Rate Top EHRs,” *Medscape,* August 25.

Phelps, J., Nowak, G., and Ferrell, E. 2000. “Privacy Concerns and Consumer Willingness to Provide Personal Information,” *Journal of Public Policy & Marketing* (19:1), pp. 27-41.

Pinsonneault, A., Addas, S., Qian, C., Dakshinamoorthy, V., and Tamblyn, R. 2017. “Integrated Health Information Technology and the Quality of Patient Care: A Natural Experiment,” *Journal of Management Information Systems* (34:2), pp. 457-486.

Plott, C. R., and Zeiler, K. 2005. “The Willingness to Pay-Willingness to Accept Gap, the ‘Endowment Effect,’ Subject Misconceptions, and Procedures for Eliciting Valuations Experimental,” *The American Economic Review* (95:3), pp. 530-545.

Ponemon Institute. 2017. “2017 Cost of Data Breach Study,” *Ponemon Institute*, https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=SEL03130WWEN.

Posner, R. A., 1981. “The Economics of Privacy,” *American Economic Review* (71), pp. 405-409.

Powell, J. L. 1986. “Symmetrically Trimmed Least Squares Estimation for Tobit Models,” *Econometrica* (54:6), pp. 1435-1460.

Preibusch, S. 2015. “How to Explore Consumer’s Privacy Choices with Behavioral Economics,” in *Privacy in a Digital, Networked World,* S. Zeadally and M. Badra (eds.), pp. 313-341.

Premkumar, G., Ramamurthy, K., and Saunders, C. S. 2005. “Information Processing View of Organizations: An Exploratory Examination of Fit in the Context of Interorganizational Relationships,” *Journal of Management Information Systems* (22:1), pp. 257-294.

Raman, R., and Bharadwaj, A. 2012. “Power Differentials and Performative Deviation Paths in Practice Transfer: The Case of Evidence-Based Medicine,” *Organization Science* (23:6), pp. 1593-1621.

Ribeiro, J. 2015. “Radioshack still plans to sell customer personal data despite state objections,” *PC World,* April 14.

Rivenbark, D. R. 2011. “Experimentally Elicited Beliefs Explain Privacy Behavior,” *Working Paper*.

Rivers, P. A., Woodard, B., and Munchus, G. 1997. “Organizational Power and Conflict Regarding the Hospital-Physician Relationship: Symbolic or Substantive?,” *Health Services Management Research: An Official Journal of the Association of University Programs in Health Administration / HSMC, AUPHA* (10:2), pp. 91-106.

Romanosky, S. 2016. “Examining the Costs and Causes of Cyber Incidents,” *Journal of Cybersecuirty,* pp. 1-15.

Romanosky, S., and Acquisti A. 2009. “Privacy Costs and Personal Data Protection: Economic and Legal Perspectives,” *Berkeley Technology Law Journal* (24), pp. 1061-1102.

Salane, D. E. 2009. “Are Large Scale Data Breaches Inevitable?,” *Cyber Infrastructure Protection ’09*.

Samuelson Law, Technology, and Public Policy Clinic. 2007. “Security Breach Notification Laws: Views from Chief Security Officers,” *University of California at Berkeley School of Law*, December.

Sarabi, A., Naghizadeh, P., Liu, Y., and Liu, M. 2016. “Risky Business: Fine-grained Data Breach Prediction Using Business Profiles,” *Journal of Cybersecurity* (2:1), pp. 15-28.

Schaffhauser, D. 2017. “Average Cost Per Record of U.S. Data Breach in Ed: $245,” *Campus Technology*, July 18.

Schneider, J. W. 2009. “Alternative Approaches to Deter Negligent Handling of Consumer Data,” *Boston University Journal of Science and Technology* (15), pp. 279.

Schwartz, P., and Janger, E. 2007. “Notification of data security breaches,” *Michigan Law Review*, pp. 913-984.

Sen, R., and Borle, S. 2015 “Estimating the Contextual Risk of Data Breach: An Empirical Approach,” *Journal of Management Information Systems*, (32:2), pp. 314-341.

Shah, R. 2015. “Do Privacy Concerns Really Change With The Internet of Things?,” *Forbes*, July 2.

Sheehan, K. B., and Hoy, M. G. 2000. “Dimensions of Privacy Concern among Online Consumers,” *Journal of Public Policy & Marketing* (19:1), pp. 62-73.

Shinal, J. 2017. "Faebook's revenue per user topped $5 for the first time," *CNBC.com*. Published November 2nd, available online at: https://www.cnbc.com/2017/11/02/facebooks-revenue-topped-5-per-user-for-the-first-time.html

Shogren, J. F., Shin, S. Y., Hayes, D. J., and Kliebenstein, J. B. 1994. "Resolving differences in willingness to pay and willingness to accept," *The American Economic Review*, pp. 255-270.

Smith, H. J., Dinev, T., and Xu, H. 2011. "Information Privacy Research: An Interdisciplinary Review," *MIS Quarterly* (35:4), pp. 989-1015.

Solove, D. J. 2006. "Taxonomy of Privacy," *Univ. of Pennsylvania Law Review* (154:3), pp. 477-564.

Soojian, C. 2015. "Content Personalization: It's What Consumers Want!," *Social Media Today*, April 4.

Spiekermann, S., Acquisti, A., Bohme, R., and Hui, K. 2015. "The challenges of personal data markets and privacy," *Electronic Markets* (25:2), pp. 161-167.

Staiano, J., Oliver, N., Lepri, B., de Oliveira, R., Caraviello, M., and Sebe, N. 2014. "Money walks: a human-centric Study on the Economics of Personal Mobile Data," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 583-594.

Straub, D. W. 1990. "Effective IS Security: An Empirical Study," *Information Systems Research* (1:3), pp. 124-133.

Straub, D. W., and Nance, W. D. 1990. "Discovering and Disciplining Computer Abuse in Organizations: A Field Study," *MIS Quarterly* (14:1), pp. 45-60.

Straub, D., W., and Welke, R., J. 1998. "Coping with Systems Risk: Security Planning Models for Management Decision Making," *MIS Quarterly* (22:4), pp. 441-469.

Sullivan, T., and Davis, J. 2017. “HIPAA Breach Fines: It’s Time to Rethink This Mess,” *Healthcare IT News*, https://www.healthcareitnews.com/news/hipaa-breach-fines-its-time-rethink-mess, May 10.

Sutanto, J., Palme, E., Tan, C., and Phang, C. W. 2013. “Addressing the Personalization-Privacy Paradox: An Empirical Assessment from a Field Experiment on Smartphone Users,” *MIS Quarterly* (37:4), pp. 1141-1164.

Taylor, S. E., and Fiske, S. T. 1978. “Salience, Attention, and Attribution: Top of the Head Phenomena,” *Advances in Experimental Social Psychology* (11), pp. 249-288.

Torrieri, M. 2013. “What to Do When Your Medical Practice Data Is Breached,” *Physicians Practice* (23), May 7.

Treisman, A. 1985. “Preattentive Processing in Vision,” *Computer, Vision, Graphics, and Image Processing* (31:2), pp. 156-177.

Tsai, J., Egelman, S., Cranor, L., and Acquisti, A. 2011. “The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study,” *Information Systems Research* (22:2), pp. 254-268.

Tufekci, Z. 2008. "Grooming, gossip, Facebook and MySpace: What can we learn about these sites from those who won't assimilate?". Information, Communication & Society. 11 (4): 544–64.

Tversky, A.; Kahneman, D. 1974. “Judgment under Uncertainty: Heuristics and Biases” *Science* (185:4157): 1124-1131.

Tyler, T. R., and Blader, S. L. 2005. “Can Businesses Effectively Regulate Employee Conduct? The Antecedents of Rule Following in Work Settings,” *Academy of Management Journal* (48:6), pp. 1143-1158.

Widup, S. 2010. “The leaking vault: Five years of data breaches,” *Digital Forensics Association*, July.

Williams, C., Asi, Y., Raffenaud, A., Bagwell, M., and Zeini, I. 2016. “The Effect of Information Technology on Hospital Performance,” *Health Care Management Science* (19), pp. 338-346.

Willison, R., and Warkentin, M. 2013. “Beyond Deterrence: An Expanded View of Employee Computer Abuse,” *MIS Quarterly* (37:1), pp. 1-20.

Woolridge, J. M. 2015. “Introductory Econometrics: A Modern Approach,” *Nelson Education*.

Xu, H., Teo, H. H., Tan, B. C. Y., and Agarwal, R. 2010. “The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services,” *Journal of Management Information Systems* (26:3), pp. 137-176.

Zetlin, Minda. 2018. “Despite all the #DeleteFacebook talk, data shows we’re using it more, not less,” Inc. Available online: https://www.inc.com/minda-zetlin/for-all-deletefacebook-talk-data-shows-were-using-it-more-not-less.html