Do Organizations Learn from a Data Breach?

Joseph Buckman
Management Information Systems, Eller College of Management, University of Arizona,
jbuckman@email.arizona.edu

Jesse C. Bockstedt
Information Systems and Operations Management, Goizueta Business School, Emory University,
bockstedt@emory.edu

Matthew J. Hashim
Management Information Systems, Eller College of Management, University of Arizona,
mhashim@email.arizona.edu

Tiemen Woutersen
Department of Economics, Eller College of Management, University of Arizona,
woutersen@email.arizona.edu

Abstract

This study investigates the duration between a data breach within an organization and subsequent data breaches at the same organization given certain organization, notification law, and data breach characteristics. We use a subsample of the publicly available Privacy Rights Clearinghouse data set to analyze data breaches between 2010 and 2016. We analyze the duration using a hazard model. The results from the hazard model provide that the duration between two data breaches at an organization increases when the organization is an educational institution, a medical organization, a non-retail business, or it resides in a state that requires notifying Consumer Reporting Agencies of the breach. The duration also increases when the breach type of the prior data breach was a system hack, an unintended disclosure of information, the mishandling of a portable device, or the mishandling of paper records. Furthermore, we find that the duration between two data breaches decreases if the organization must report to the state Attorney General's office, the state in which the organization resides has enacted data breach notification legislation, or the number of records compromised in the prior breach was greater than zero. These results provide a unique outlook on data breaches and supplements the current data breach literature by identifying how prior data breaches affect the duration until a subsequent breach.

Introduction

"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 potentially put at risk because of exposure."

Identity Theft Resource Center

In the United States, 2016 was a record setting year in which 1,093 breach incidents were recorded across five different industries (Identity Theft Resource Center). According to several data breach reports, such as those provided by Identity Theft Resource Center (ITRC) and Privacy Rights Clearinghouse (PRC), the industries that experience the most data breaches each year include the business and healthcare sectors; and often times, these data breaches are the result of malicious hacking or social engineering. Unfortunately, it is difficult to identify the underlying reason behind the significant increase in data breaches. However, scholars and industry leaders provide empirical evidence that data breaches are costly to both firms and consumers. In many cases, a breached firm is required to pay for notifying all potentially compromised consumers, provide those consumers with a minimum of twelve months of identity theft protection, and settle fines handed down by overseeing agencies (Romanosky and Acquisti 2009). Examples of these costs include the 2013 Target and 2014 Home Depot data breaches. Recent reports indicate that the sum of the cumulative costs for both companies have exceeded $550 million (Daly 2016). Further, more than twenty percent of breached firms reported losing a significant portion of their consumer base following a security incident (Burns 2017). Thus, we can expect such high costs to the organization to act as a deterrent for experiencing a single, let alone multiple, data breaches.

Yet, sources that track public data breaches, such as PRC, provide a different story. In fact, 11% of PRC's publicly available security breach data set, between 2010 and 2016, contains organizations that experience between two and ten data breaches. The latest of these is Yahoo Inc. who only recently publicized their 2013 data breach and then, months later, reported their 2014 data breach. In the wake of Yahoo Inc.'s breach notifications and impending Securities and Exchange Commission investigations, stock market specialists predict that breaches will cost the company hundreds of millions of dollars and significantly harm its company sale valuation (Snider 2017).

Using the 2010 through 2016 breach data by PRC, we extend the existing data breach literature by examining the duration between data breaches within an organization. Specifically, we analyze whether the type of organization, state data breach notification legislation characteristics, and the type of data breach a firm experiences affect the duration between a prior breach and a subsequent breach at the same firm. We find that when an organization experiences a system hack, unintentional disclosure of information, the mishandling of a portable device containing private information records (e.g., laptop, USB drive), the mishandling of paper records, or if the organization is required to notify a Consumer Reporting Agency (CRA) of the data breach there is a longer duration between the prior breach and a subsequent breach. We also discover that educational institutions, medical organizations, and non-retail businesses have longer durations between data breaches. Further, the results provide that when the organization resides in a state that has enacted data breach notification legislation, the organization must notify the state Attorney General's office, or the number of records compromised in the breach is greater than zero there is a shorter duration between the prior breach and a subsequent breach. The findings in this paper are preliminary as we continue to explore this phenomenon.

The remainder of this paper continues with Section 2 discussing the related data breach literature and our theoretical background. Section 3 provides a detailed description of our data and analysis. In Section 4, we discuss the results from our analysis, their theoretical contributions, and managerial implications. We end the paper with the limitations of our study, future research to carry forward, and conclusion.

Related Literature

Thus far, the data breach literature has focused on (1) the repercussions associated with and the effectiveness of data breach notification laws and (2) data breach modeling. Beginning with data breach notification laws, a majority of state governments within the United States have implemented data breach disclosure laws due to the rising number of data breaches affecting thousands and, at times, millions of Americans. The first instance of notification law was the California Civil Code Section 1798.29 in 2003, which required all firms, with business in California, to report a privacy or security breach to the individuals affected by the breach. The law also required firms report to local and state authorities when a breach affected five hundred or more records. 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).

The intent behind such 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 (Schwartz and Janger 2007). Specifically, data breach notification laws seek to increase consumer awareness and force firms to incur further financial costs by aiding consumers affected by the data breach through notification letters, customer

support call centers, and provision of identity theft protection services. Salane (2009) studied the largest breaches at the time and found evidence that a breach considered small, 200 consumers, cost the organization over $500,000 and a breach considered large, 100 million consumers, cost the organization over $12 million. Romanosky (2016) sought to examine the costs of data breach incidents. Romanosky found that the mean cost of a single data breach was roughly $4 million, but the mean cost doubles to almost $10 million for organizations experiencing multiple data breaches. While organizations cover a majority of the financial costs associated with a breach, consumers face time and effort inconveniences that leave them disgruntled with the organization; thereby leading to the chance of decreased sales (Salane 2009). Recently, the annual cyber security report by Cisco Systems provided 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. 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. On the other hand, 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, recent studies by Acquisti et al. (2006), Ko and Dorantes (2006), Gatzlaff and McCullough (2010), and Gay (2015) provide further support that there is 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 in part due to the use of differing methodologies, datasets, and market valuation metrics, but there is ample

evidence to suggest that a firm's stock market valuation is adversely affected by a data breach incident.

Ignoring that breached organizations front all monetary costs and lost revenue, studies have found that the requirement for breach disclosure also has a negative effect on an organization's brand name and it generates negative publicity (Leonard and Rubin 2006). Following a data breach, fully compliant organizations may still fall victim to reputation loss and decreased market value, 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). Specifically, 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 disclosure, Romanosky et al. (2011) sought to answer whether data breach notification laws successfully reduce the number of reported identity thefts and found only a nonsignificant, two percent decrease. In fact, privacy scholars have found that such laws are useful for informing consumers and eliciting transparency within organizations but are ultimately ineffective (Romanosky and Acquisti 2009).

The second area of focus in the data breach literature is data breach modeling with an interest in analyzing time series. Early studies such as Curtin and Ayres (2008) struggled to find conclusive evidence of significant trends within their breach data because of sample size limitations. Later, Widup (2010) conducted an in-depth time series analysis on yearly trends in data breaches between 2005 and 2009. Widup discovered that a majority of data breaches were caused by theft or loss of an employee laptop and that system hacking affected a greater number of breached records. Recently, Edwards et al. (2016) utilized a Bayesian approach to identify time series trends and attempt to predict future data breaches. Their model found that, for

breaches over 500,000 records, the size and frequency of the data breaches have remained stable. Further, Edwards et al.'s model predicted a 7.8% chance that another large-scale data breach exceeding 80 million records will occur between 2015 and 2018, which by the end of 2016 there were five public announcements of data breaches exceeding 80 million records. The research presented in this paper contributes to the data breach literature on modeling with time series analysis by predicting the likelihood of subsequent breaches at the same organization. Unlike prior work, we seek to understand the effects of characteristics surrounding an initial data breach on the probability that the same organization will experience a future data breach. In addition, we control for the organization's industry using NAICS codes because of differing state and federal data breach regulations across industries. In the next section of the paper, we discuss the theory motivating why firms want to take actions to mitigate their chances of falling victim to multiple data breaches and our expectation of a negative relationship between initial breach characteristics and the likelihood of a subsequent breach.

Theory: Deterrence Strategy

Although Romanosky and Acquisti (2009) concluded that data breach notification laws are ineffective and little evidence supports their reduction in identity theft crimes, organizations have responded to them by taking significant measures to improve their information and operational security (Samuelson Law 2007; Schneider 2009). Breach notification laws act as deterrence policies to incentivize firms. Applying Cohen (2000)'s investigation into environmental deterrence policies involving incident disclosures to data breach notifications, requiring organizations to disclose information about information security breaches serves as an informal penalization to be used in conjunction with any formal penalization handed down by an overseeing agency. Using the requirements of the notification laws and public reaction, we

define formal penalization for a breach incident as the total direct monetary costs stemming from (1) notifying consumers with compromised records, (2) providing compromised consumers with identity theft protection, and (3) settling lawsuits or fines placed against the organization for the breach. We define informal penalization as the indirect monetary loss organization's experience following a breach such as drops in stock value, revenue loss from negative publicity, and decrease in firm performance.

Naturally, deterrence policies lend themselves to an economic framework of analyzing costs and benefits. With this in mind, the penalizations businesses experience because of breach notification laws affect the incentives of firms to invest in preventive information security measures (Laube and Bohme 2016). Specifically, the collection of consumers' personal data offers organizations a significant competitive advantage and these organizations strive to maximize their data gathering capabilities (Spiekermann et al. 2015). However, consumer data is also a liability because it invites malicious activity to compromise the information, which can negate the benefits derived and result in tremendous monetary loss, at least temporarily. Thus, we assume organizations provide some form of system security and implement at least minor information security policies to secure their data assets and minimize their breach probability. In the event of a data breach at an organization, we then expect the organization to take further actions by enhancing their system security and implementing new information security policies to both publicly demonstrate their response to the breach and further diminish their breach probability.

Data Analysis and Results

Privacy Rights Clearinghouse

The data set we use for our analysis is publicly available from PRC, a non-profit organization that collects breach information from government agency and news media websites as soon as the breach becomes public. PRC has been gathering data 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 limit the data set to breaches that occurred between 2010 and 2016. We chose this subset of the data because data breach notification laws are currently at the state-level only and most states began implementing their own notification laws at varying times between 2005 and 2009. It was not until 2010 that forty-four out of the fifty states in the U.S. had implemented a form of data breach notification legislation, in which three of the remaining six still do not have notification laws. States used California's early notification legislation as a model for their own, thereby establishing similarity among the states and reducing the need to control for state-level effects. Thus, the subset of our data contains 2,488 data breach records.

Each breach record contains the following information: the specific date the breach was made public (in MM/DD/YYYY), a unique identifier, the name of the breached organization, the location in which the organization resides, the type of breach, the industry's organization, the number of records compromised, and a detailed description of the event. Specifically, the types of breaches include credit card fraud, system hacks, insider threats, mishandled paper documents, mishandled portable electronic devices, mishandled stationary electronic devices, unintended disclosures, and unspecified. Please refer to Table 1 for descriptions of the breach types found in the PRC dataset. The types of industries organizations include government, education, retail business, financial business, other non-retail/financial businesses, medical, and non-profit. For our analysis, we add one to the number of records compromised and then take the natural

logarithm of the number. The number of records compromised ranged from zero to one billion with a standard deviation of thirty million.

Table 1. Variable Descriptions	
Variable	**Description**
Industry	
Government	The organization is a governmental organization such as Department of Health and Human Services.
Education	The organization is an educational institution such as a state university.
Medical	The organization is a medical organization such as a hospital.
Financial	The organization is a financial institution such as a bank.
Business (Retail)	The organization is a retail business such as Target.
Business (Other)	The organization is not a retail business but has paying customers such as Netflix.
Non-profit	The organization is a non-profit organization.
Breach Notification Law	
Harm Threshold	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.
Notification Law Indicator	The state in which the breach occurred has enacted a data breach notification law.
Paper Records	The form of information covered for a data breach includes paper records.
Civil Penalties	The organization may incur civil penalties as a result of the breach.
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.
Other Notification	Organizations must provide notification of a breach to a state government agency other than the Attorney General's office.
No Notification Requirement	Organizations are not required to provide notification of a breach to any government or consumer agencies.
Breach Type	
Hacking	The organization's information systems are hacked by an outside party or infected by malware.
Unintended Disclosure	Unintended disclosure (not involving hacking, intentional breach or physical loss – for example: sensitive information posted publicly, mishandled or sent to the wrong party via publishing online or sending in an email).
Credit Card Fraud	Fraud involving debit and credit cards that is not accomplished via hacking. For example, skimming devices at point-of-service terminals.
Insider Threat	Insider (someone with legitimate access to intentionally breach information – such as an employee, contractor, or customer).
Mishandling Portable Device	Lost, discarded, or stolen laptop, PDA, smartphone, memory stick, CDs, hard drive, data tape, etc.
Mishandling Paper Records	Includes paper documents that are lost, discarded, or stolen (non electronic).
Mishandling Stationary Device	Stationary computer loss (lost, inappropriately access, discarded or stolen computer or server not designed for mobility).
Unspecified	A description of the event is not provided nor is a type assigned for the record. Generally, this indicates the organization is unsure of what happened to the records.

Other Variables	
log(Records Compromised)	The natural logarithm of the number of records compromised from a data breach.

In order to strengthen our hazard model, we made several additions to the PRC data set. The first addition we made was the assignment of each organization's North American Industry Classification System (NAICS) code. A NAICS code classifies businesses and organizations for ease of collecting, analyzing, and publishing statistical data related to the United States economy. A NAICS code is a six-digit number that distinguished sub categories within a broad industry. For instance, the medical industry contains thirty-eight subcategories that further specify the type of medical organization or facility. We added organizations' NAICS codes to control for heterogeneity within each industry sector because different rules and regulations apply to different industries and industry subcategories.

The second addition we made was to code data breach notification law characteristics associated with each state. Specifically, we created indicator variables for whether or not the state has enacted a data breach notification law, if the law contains a harm threshold, if notification applies to both electronic and paper records, the type of penalties an organization can be charged with, and who the organization must notify in the case of a data breach. Refer to Table 1 for descriptions of these variables. In the following sub-sections, we describe the hazard models and its analysis.

Subsequent Breach Analysis

Table 2. Summary Statistics				
Indicator Variables	**Total Records**		**Subsequent Breach Records**	
	Obs.	**Percent Yes**	**Obs.**	**Percent Yes**
Industry				
Government	2488	11.9	267	10.1
Education	2488	9.6	267	7.1

Medical	2488	32.5		267	24.7
Financial	2488	11.2		267	18.4
Business (Retail)	2488	12.9		267	16.1
Business (Other)	2488	19.7		267	17.6
Non-profit	2488	2.1		267	1.1
Breach Notification Law					
Harm Threshold	2488	53.2		267	46.1
Notification Law Indicator	2488	98.3		267	99.3
Paper Records	2488	15.2		267	11.2
Civil Penalties	2488	97.5		267	98.9
Criminal Penalties	2488	6.6		267	6.0
Attorney General Notification	2488	56.7		267	70.0
CRA Notification	2488	63.2		267	53.9
Other Notification	2488	21.7		267	19.5
No Notification Requirement	2488	8.0		267	7.1
Breach Type					
System Hack	2488	37.2		267	34.5
Unintended Disclosure	2488	18.0		267	24.7
Credit Card Fraud	2488	1.7		267	3.4
Insider Threat	2488	13.1		267	16.5
Portable Device	2488	13.9		267	7.1
Stationary Device	2488	2.1		267	1.1
Paper Records	2488	11.3		267	8.2
Unknown	2488	2.5		267	4.5
Continuous Variables					
	Obs.	**Mean**	**St. Dev**	**Min**	**Max**
log(Records Compromised)	2488	1.931	1.938	0	9
Dependent Variables					
	Obs.	**Number of Organizations with >1 Breach**			
Subsequent Breaches	2488	172			

The total number of records for analysis in the 2010-2016 subset was 2,488 and of that total 267 records were breach incidents at organizations that experienced two or more breaches. In other words, 11% of breaches in the subset of data were incidents involving organizations with multiple data breach records. We used a hazard model to analyze the effects of organization type, data breach legislation characteristics, data breach type, and the number of records compromised on the duration between a firm experiencing a prior breach and a subsequent breach. The model was right-censored due to the organizations continued risk of experiencing breaches and the ongoing collection of breaches that we did not include.

We established the data as censored survival data in the following ways. First, we created an *id* variable that assigned a unique identification number to each organization in the data set. The *id* allowed for assigning multiple breaches to the same organization. Next, we created a *t* (time) variable for the duration, in years, between a prior data breach and a subsequent data breach. Thus, we took the difference, in days, between the date the prior breach was made public and the date the subsequent breach was made public then divided it by 365.25. This value was then associated with the prior data breach record. Using the *t* variable, we incorporated censoring into the model with a binary indicator variable *event*. The *event* variable took a value of 1 if there was a subsequent data breach following the prior data breach. The *event* variable took a value of 0 if there was not a subsequent data breach.

For example, Apple Inc. experienced two data breach incidents, one on January 1, 2011 and another on January 1, 2014. We then assigned a unique identifier (*id*) to represent Apple Inc. Next, we use the difference between the breach dates and convert it to years for our *t* variable, which was 3.000384. Since there was a subsequent data breach, our *event* variable for the prior breach record (January 1, 2011) was set to 1. The *event* variable for the subsequent breach record (January 1, 2014) was set to 0 because Apple Inc. another data breach in our data set.

After preparing the data for survival analysis, we ran three hazard models with an exponential distribution. The independent variables in our first hazard model, Model (1), were the eight indicator variables for each organization type. The independent variables in our second hazard model, Model (2), built upon Model (1) by adding nine indicator variables for the state-level data breach notification legislation characteristics. The independent variables in our third hazard model, Model (3), built upon Model (2) by adding eight indicator variables for the breach types and a continuous variable for the natural logarithm of the number of records compromised

in the breach. We also estimated hazard models using Weibull and Gompertz distributions for comparison with the results of the exponential models. Each model was estimated using clustered standard errors on the NAICS codes. In addition, the three full models were estimated with the Unspecified data breach type as the baseline indicator variable. We used the Unspecified breach type as the baseline for our analysis because there was minimal information associated with the breach incident. Table 3 displays the hazard ratios from our models.

Table 3. Subsequent Breach Hazard Ratios

	Exponential (1)	Exponential (2)	Exponential (3)	Weibull	Gompertz
Variable					
Government	0.551 (0.223)	0.560 (0.230)	0.517 (0.222)	0.557* (0.193)	0.581 (0.199)
Education	0.455*** (0.048)	0.421*** (0.050)	0.457*** (0.073)	0.527*** (0.067)	0.628*** (0.087)
Medical	0.407* (0.194)	0.402* (0.194)	0.356** (0.173)	0.403** (0.165)	0.466* (0.189)
Financial	1.160 (0.548)	1.185 (0.557)	1.091 (0.514)	1.178 (0.466)	1.284 (0.523)
Business (Retail)	0.960 (0.518)	0.905 (0.499)	0.939 (0.556)	1.010 (0.512)	1.002 (0.518)
Business (Other)	0.370** (0.177)	0.342** (0.160)	0.403* (0.196)	0.455** (0.182)	0.520* (0.209)
Non-profit	0.325 (0.225)	0.308* (0.214)	0.335 (0.243)	0.436 (0.288)	0.556 (0.374)
Harm Threshold		0.905 (0.140)	0.942 (0.134)	0.919 (0.127)	0.851 (0.152)
Notification Law Indicator		3.674 (3.334)	4.952* (4.474)	3.794 (3.226)	3.135 (2.558)
Paper Records		0.670* (0.143)	0.646** (0.138)	0.700* (0.144)	0.798 (0.182)
Civil Penalties		0.543 (0.297)	0.474 (0.247)	0.526 (0.247)	0.554 (0.270)
Criminal Penalties		1.337 (0.373)	1.371 (0.370)	1.328 (0.352)	1.113 (0.390)
Attorney General Notification		1.542*** (0.222)	1.478*** (0.213)	1.427** (0.208)	1.478** (0.283)
CRA Notification		0.790 (0.137)	0.702** (0.115)	0.766 (0.130)	0.892 (0.187)
Other Notification		0.836 (0.156)	0.883 (0.174)	0.803 (0.147)	0.683 (0.164)
No Notification Requirement		0.850 (0.260)	0.743 (0.211)	0.800 (0.208)	0.883 (0.251)
System Hack			0.269*** (0.086)	0.295*** (0.086)	0.250*** (0.080)
Unintended Disclosure			0.450*** (0.133)	0.503** (0.147)	0.505** (0.158)
Credit Card Fraud			0.499	0.597	0.691

			(0.133)	(0.392)	(0.458)
Insider Threat			0.431 (0.127)	0.688 (0.197)	0.637 (0.201)
Portable Device			0.307*** (0.102)	0.333*** (0.108)	0.300*** (0.108)
Stationary Device			0.647 (0.312)	0.703 (0.317)	0.716 (0.333)
Paper Records			0.374** (0.142)	0.410** (0.158)	0.366** (0.155)
log(Records Compromised)			1.086*** (0.027)	1.066*** (0.026)	1.059** (0.030)
Constant	0.184*** (0.076)	0.095*** (0.084)	0.194* (0.182)	0.164** (0.147)	0.105** (0.097)
Distirbution Parameter (Weibull and Gompertz)				1.697*** (0.197)	0.736*** (0.072)
Obs.	2488	2488	2488	2488	2488
Organizations	2221	2221	2221	2221	2221
Number of Breach Events	267	267	267	267	267
Log Likelihood	-891.131	-876.403	-857.029	-808.235	-759.790
x^2	184.36***	180.31***	235.55***	174.45***	155.95***
Number of Clusters	394	394	394	394	394

* $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$
All models were estimated with clustered standard errors on the NAICS code.

The results from Model (3) provided a significant increase in the duration between a prior data breach and a subsequent data breach when the organization was an education institution, medical organization, or a non-retail business. Specifically, we found a 54.3%, 64.4%, and 59.7% lower hazard rate, respectively. Model (3) also demonstrated that when the prior data breach type was a system hack, unintended disclosure, the mishandling of a portable device, or the mishandling of paper records there was a significant increase in the duration between a prior data breach and a subsequent data breach. Specifically, we found a 73.1%, 55.0%, 69.3%, and 62.6% lower hazard rate, respectively. Furthermore, enacting a data breach notification law that requires organizations provide notification of paper record loss and submit all notifications to a Consumer Reporting Agency led to an increase in the duration between a prior data breach and a subsequent data breach. Specifically, we found a 35.4% and 29.8% decrease in the hazard rate, respectively. Interestingly, the presence of a data breach notification law, requiring organizations submit notifications to the state Attorney General, and the natural log of the number of records

compromised decreased the duration between a prior data breach and a subsequent data breach. Specifically, we found a 492%, 137%, and 108% increase in the hazard rate, respectively. The remaining variables did not significantly affect the duration between data breaches.

Discussion

Although the results from our analysis are preliminary, they offer intriguing insight into the characteristics affecting the duration between data breaches within an organization. The increase in the duration between data breaches when the organization is an education institution, medical organization, or non-retail business is interesting but difficult to pursue. In particular, the education and medical organizations have seen shifts in their rates of intrusions, as they become targets for malicious activity. These sectors have received considerable attention over recent years to safeguard their data. For instance, the HIPAA and the HITECH Act were beginning to take effect in 2010 for medical organizations and led to shifts in the needs among these organizations. Employees may have required extensive updating and training on proper data record handling as mistakes occurred. It is possible that there were adjustment periods at medical organizations in which employees made several early mistakes but eventually corrected them. Thus, educational institutions and medical organizations are difficult to extract a clean interpretation for the decrease in duration. However, it is possible that we can use firm size in accordance with the organization type to gather additional information for non-retail businesses.

The increase in the duration between a prior data breach and a subsequent data breach from requiring notification of data breach when paper records are included or submitting notifications to a Consumer Reporting Agency may be somewhat less cloudy. For instance, organizations in these states may take additional precautions for handling paper records. It could also be that notifying Consumer Reporting Agencies negatively affects the organization more

than notifying other groups or government entities. Unfortunately, it may be difficult to surmise the precautions taken at these organizations. However, in future extensions we can expand upon potential negative effects of reporting to Consumer Reporting Agencies by analyzing the losses associated with notification to a Reporting Agency and the losses associated with notifications other government entities.

Furthermore, the increase in the breach durations from experiencing a system hack, unintended disclosure of information, mishandling of a portable device, or mishandling of paper records may also be of use. The decrease in duration from a system hack may be the result of an organization's ability to identify and correct system flaws. Although the system was breached at some point, it may be easier to implement technical corrections than enforcing employee policies that oversee the remaining breach types.

Lastly, the decrease in the duration between data breaches at an organization from enacting a data breach notification law, requiring organizations notify the state Attorney General of a data breach, and the natural log of the records compromised in the prior breach offers expected results. It is natural to expect shorter durations between data breaches upon enacting notification legislation and requiring notification to government officials because firms are obligated to provide notification. Thus, firms will enter or reappear in a breach notification data more frequently in order to comply with the law. The same is true for notifying government officials. By requiring firms submit notice to government officials, the data becomes more accessible and increases the likelihood of being placed in the PRC data set. Regarding the number of records compromised, a follow up analysis, in which we control for individual firm characteristics, may shed light on this result.

Implications for Practice

The business world today relies heavily on the collection, dissemination, and dissection of consumer data. Thus, the large databases of consumer data an organization manages will remain at risk of a data breach. However, organizations can successfully mitigate this risk. For example, we briefly mentioned the costs of Target's data breach earlier. That particular breach resulted in stolen credit cards, damaged reputation, and loss of firm value, among other consequences. As a result, Target was an early adopter in the U.S. of EMV Chip-and-PIN information security technology for payment processing, which is a strong countermeasure against future credit card data threats.

From an information security professional's standpoint, the results in this paper offer evidence to upper-level management that investment in the organization's information security will likely yield longer durations between experiencing a data breach and significant losses following a breach. In addition, information security consultants and software vendors can leverage these findings to organizations (e.g., health care organizations and small businesses). That is, they can further demonstrate the usefulness of implementing stringent information security policies and training by supplementing an organization's knowledge of important information security investments.

The findings in this paper are also useful to legal scholars and policy makers as they provide evidence that data breach notification legislation is contributing to more firms disclosing data breach incidents. Such a result can strengthen the argument that organizations are complying with notification laws by disclosing breaches to the required entity. Unfortunately, ensuring organizations submit notifications of data breaches may also dull the impact of the notification. As data breaches become more public and consumers are increasingly aware of

them it may lead to a sense of inevitability. In other words, consumer reactions will become dulled to breach notifications and hold firms less accountable to the event.

Limitations and Future Research

As with all research, our study also has limitations. The first limitation is the usage of PRC's breach 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 2.5% of the data) has Unspecified 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. Second, we must rely 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 study's findings.

There are several possibilities for future research with this data. The findings from this study bring new research questions to light. In particular, privacy researchers can take these findings and apply additional firm-level characteristics such as firm size. Researchers can also investigate accounting and economic firm variables to aid with firm distinction.

Concluding Remarks

In conclusion, we investigate the effect of organization type, state data breach legislation characteristics, and prior data breach type on the duration between a prior data breach and a subsequent data breach. To conduct the analysis, we utilize the publicly available Privacy Rights Clearinghouse dataset for breach incidents between 2010 and 2016. Using a hazard model, we

find that there is an increase in the duration between data breaches at the same organization when the organization is an education institution, medical organization, or a non-retail business. We also discover an increase in the duration between breaches when the state requires notification of compromised paper records, the organization must report a data breach to a Consumer Reporting Agency, the data breach was a system hack, the breach was due to an unintended disclosure, the breach was from mishandling a portable device, or the breach was caused by mishandling paper records. Furthermore, we learn that the duration between a prior data breach and a subsequent data breach decreases when the state has enacted data breach legislation, the legislation requires organizations notify the state Attorney General's office of a breach, and the number of records compromised in the prior breach is greater than zero. To our knowledge, our study is the first to investigate changes in duration between data breaches. Our findings are preliminary but provide useful results for guiding future data breach research and policy regulations.

References

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.

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

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.

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.

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* (30), pp. 10245-10252.

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

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.

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.

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)*.

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.

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

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.

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.

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

Law, S. 2007. "Security Breach Notification Laws: Views from Chief Security Officers," *University of California-Berkeley School of Law*.

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

National Conference of State Legislatures. 2016. “Security Breach Notification Laws,” *National Conference of State Legislatures*, January 4.

Romanosky, S., and Acquisti, A. 2009. “Privacy Costs and Personal Data Protection: Economic and Legal Perspectives,” *Berkeley Technology Law Journal* (24:3), pp. 1062-1091.

Romanosky, S., Telang, R., and Acquisti, A. 2011. “Do data breach disclosure laws reduce identity theft?”, *Journal of Policy Analysis and Management* (30:2), pp. 256-286.

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

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.

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.

Snider, M. 2017. “SEC said to probe Yahoo data breaches,” *USA Today*, January 23.

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.

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