Cyber-Analytics: Identifying Discriminants of Data Breaches

by Diane Dolezel, EdD, RHIA, CHDA, and Alexander McLeod, PhD

Abstract

In this study, the relationship between data breach characteristics and the number of individuals affected by these violations was considered. Data were acquired from the Department of Health and Human Services breach reporting database and analyzed using SPSS. Regression analyses revealed that the hacking/IT incident breach type and network server breach location were the most significant predictors of the number of individuals affected; however, they were not predictive when combined. Moreover, network server location and unauthorized access/disclosure breach type were predictive when combined. Additional analyses of variance revealed that covered entity type and business associate presence were significant predictors, while the geographic region of a breach occurrence was insignificant. The results of this study revealed several associations between healthcare breach characteristics and the number of individuals affected, suggesting that more individuals are affected in hacking/IT incidents and network server breaches independently and that network server breach location and unauthorized access/disclosure breach type were predictive in combination.

Keywords: data breach; security; protected health information; breach portal; security modeling; cyber-analytics

Introduction

Healthcare data breaches continue to occur at extraordinary rates, leaving analysts challenged to identify the factors associated with data breach occurrences.[1] Between 2009 and 2017, 2,457 healthcare data breaches were reported to the Department of Health and Human Services (DHHS).[2] Breached facilities face heavy fines and litigation. For example, in 2018 Anthem health insurance paid $16 million dollars, the highest fine to date, to DHHS for a breach that exposed the data of approximately 79 million healthcare consumers.[3] Similarly, MD Anderson Cancer Center paid $4.3 million for failing to secure and encrypt patients' data.[4]

The laws on data breaches are strongly worded. According the Health Insurance Portability and Accountability Act (HIPAA) Breach Notification Rule, "A breach is, generally, an impermissible use or disclosure under the Privacy Rule that compromises the security or privacy of the protected health information."[5] Under the HIPAA Breach Notification Rule, healthcare providers must notify DHHS of breaches of unsecured personal health information affecting more than 500 individuals, and they must inform patients affected by the breach within 60 days of the breach discovery.[6] Correspondingly, the

Health Information Technology for Economic and Clinical Health (HITECH) Act states that breach notification is mandatory for vendors and third-party service providers.[7]

The reality is that breaches are often undetected for months, and breach reporting lags breach detection by prolonged periods.[8] In 2018, a breach left Blue Cross Blue Shield patient data publicly accessible for three months.[9] At UnityPoint Healthcare, cybercriminals had access to internal emails and hospital accounts from November 2017 to December 2018.[10] This type of stolen patient information is often used for medical identity theft, can be sold on the Dark Web, or may become part of ransomware threats.[11]

According to a recent Ponemon report, the annualized cost of healthcare cybercrime in the United States is $12.47 million dollars.[12] In addition to financial concerns, breached organizations experience a loss of reputation and potential loss of business.[13] Studies have attempted to model the factors related to predicting security breach occurrences by examining organizational factors.[14–17] Although healthcare analysts are challenged to identify the determinants of data breaches, few studies have examined the association between breach characteristics and the number of individuals affected by the breach.[18–20] Ronquillo et al. examined five years of healthcare breach data reports on 1,512 data breaches. They determined that hacking/IT incidents were the type of breach for 85 percent of the individuals affected by breaches during 2013 to 2017. However, their statistical association testing was limited to covered entity characteristics and the number of individuals affected.[21]

The purpose of this exploratory study was to examine the relationship between data breach characteristics and the number of individuals affected by the breaches. The number of individuals affected is important to consider because a single breach of 78 million records is 156,000 times greater than a single breach of 500 records. This study extends the Ronquillo et al. study by analyzing 2,021 reported breaches for the nine years between 2009 to 2018. It provides new information for the associations of breach type, breach location, business associate presence, and state and geographic region of the occurrence with the number of individuals affected. Additionally, the relationship between covered entity type and number of individuals affected was analyzed.

Research Questions

Understanding the factors associated with the size and scope of data breaches may assist in identifying discriminants of data breach occurrences. After a review of the literature on data breaches, the following research questions were developed:

1. What are the most common types and locations of data breach occurrences?
2. Is there a correlation between number of individuals affected and the type or the location of the breach?
3. Is there an association between data breach location and the type of the breach?
4. Can covered entity type affect the number of individual records acquired in a data breach?
5. Is there a relationship between the presence of business associates and the number of individuals affected by data breaches?
6. Do the numbers of individuals affected by data breaches vary by geographic region?

Methodology

To evaluate the relationship between the characteristics of entities who reported healthcare data breaches to DHHS and the number of individuals affected by those data breaches, the authors acquired data from the DHHS data breach reporting system for breaches affecting more than 500 individuals for the years 2009 to 2018. The data were cleaned and transformed in Excel and analyzed in SPSS 25.0. If the outcome variable—the number of individuals affected—was missing, that row of data was excluded.

Descriptive statistics were used to generate frequencies and percentages of data breach characteristics, and inferential testing was conducted with multiple regression testing and one-way analysis of variance (ANOVA).

The DHHS Office for Civil Rights (OCR) breach notifications are publicly available from the web portal at https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf. The breach database fields included the following: Name of Covered Entity, State, Covered Entity Type, Individuals Affected, Breach Submission Date, Type of Breach, Location of Breached Information, Business Associate Present, and Web Description. The categories of data breach type are Hacking/IT Incident, Improper Disposal, Loss, Theft, Unauthorized Access/Disclosure, Other, and Unknown. The DHHS portals' breach location options are Desktop Computer, Electronic Medical Record, Email, Laptop, Network Server, Paper/Film, Other, and Other Portable Electronic Device. Figure 1 displays the DHHS OCR portal for reporting breaches.

Data Analysis

The DHHS breach files for healthcare providers, healthcare clearinghouses, health plans, and business associates were downloaded from the web portal on October 8, 2018. There were 2,021 reported breaches for the years 2009 to 2018 in the downloaded data set. For our analysis, the dependent variable was the number of individuals affected. The independent categorical variables were breach type, breach location, type of covered entity, business associate presence (a binary variable), and the state where the facility reporting the breach is located. Individual facilities may report multiple types and locations of breach occurrence on a single DHHS report form; therefore, the breach types and locations were coded with 1 if the characteristic was reported for that entity and 0 if it was not reported for that entity. An extreme value analysis of the number of individuals affected indicated a need to remove five extreme cases that were significant outliers. For example, one case had 78,800,000 records, and it was also a duplicate case. Other cases trimmed from the analysis were those with numbers of individuals affected ranging from 4,900,000 to 11,000,000.

Individuals Affected by the Breach

The dependent variable was the number of individuals affected by the breach. After trimming, this variable ranged from 500 to 4,500,000 individuals (M = 36,573, SD = 269,218), with half of the facilities reporting data breaches affecting 2,201 or more individuals. The positive skewness confirms that the frequency trend is much greater than the median. Kurtosis is higher than three, indicating a more varied number of individuals affected by breaches. Table 1 shows the frequencies of individuals affected and the number of breaches by year. During 2009 to 2013, more than 73 million individuals were affected by 2,016 breaches. The largest number of individuals affected by breaches in a single year was 17,452,393 individuals (23.67 percent), in 2014. Moreover, in 2014 there were 314 reported breaches (15.58 percent of all breaches in the period), which is the largest number of breaches reported in one year. The year 2016 displayed the second highest number of individuals affected and number of breaches per year.

Data Breach Types and Locations

The most common data breach type was theft (n = 843), trailed by unauthorized access/disclosure (n = 588), hacking IT/incident (n = 337), and loss (n = 168). The most frequently occurring data breach location was paper/films (n = 509), followed by other (n = 434), laptop (n = 385), and network server (n = 371). Table 2 presents the data breach types and locations. It is important to remember that a facility may

report many breaches, and each breach may have one or more causes listed thus percentages may not add to one hundred percent.

Covered Entity Type and Business Associate Presence

Next, the type of covered entity was evaluated. Healthcare providers are defined by DHHS as doctors, clinics, psychologists, dentists, chiropractors, nursing homes, and pharmacies, if they transmit patient information electronically. Similarly, health plans are described as health insurance companies, HMOs, company health plans, and government programs that pay for healthcare, such as Medicare, Medicaid, and the military and veterans' healthcare programs. A healthcare clearing house is defined by DHHS as an entity that processes nonstandard health data received from another organization into a standard format. Business associates are defined as people or organizations performing activities involving protected health information on behalf of a covered entity.[22]

Of the 2,016 covered entities, 1,438 were healthcare providers, followed by 313 business associates, 261 health plans, and only four healthcare clearing houses. Regarding business associate presence, in 1,610 cases (79.9 percent) a business associate was not present, and in 406 cases (20.1 percent) business associates were present. The numbers of individuals affected and breaches reported by covered entity type by year for 2009 to 2018 are shown in Table 3. The largest number of breaches reported for healthcare providers was 196 (13.63 percent) in 2014, and for business associates it was 77 (24.60 percent) in the same year. During 2015, 62 health plans experienced breaches. Thus, the majority of organizations reporting a data breach were healthcare providers with business associates not present.

States with Breaches Reported

Figure 2 displays a dot plot of the number of individuals affected by state. Data breaches were reported in all 50 US states and in the District of Columbia and Puerto Rico. The four states with the largest number of individuals affected by breaches were Tennessee (10,757,170), California (9,193,740), New York (6,667,638), and Florida (6,090,108). The four states with the smallest number of individuals affected were Hawaii (14,336), Maine (9,403), Vermont (5,797), and Delaware (3,562).

To facilitate statistical analysis by region, the states were assigned region codes of 1 through 4 in accordance with the US Census listing of regions and states, shown in Table 4. For example, Florida and District of Columbia were assigned a region code of 3, since they are in Region 3 (South). Data from Puerto Rico were excluded from region analysis because Puerto Rico is a commonwealth not located in a US region.

Regression

Multiple regression tests were conducted to explore the relationships of data breach types (hacking/IT incident, etc.) and data breach locations (desktop computer, etc.) on the number of individuals affected. For these tests, the regression collinearity diagnostics indicated no tolerance less than 0.1 and no variance inflation factor (VIF) values greater than 10, signifying that no collinearity issues existed.[23] The type of breach was significantly related to the number of individuals affected by the breach, $r = .12$, $r^2 = .014$, $F(7, 2008) = 4.218$, $p < .001$. Table 5 shows the regression coefficients for the type of breach. However, when the type of breach was considered, only hacking/IT incident was marginally significant at predicting the number of individuals affected, with $p = .057$.

Similarly, location of the breach was very significantly related to the number of individuals affected by the breach, $r = .148$, $r^2 = .022$, $F(8,2008) = 5.615$, $p < .001$. Among predictors, only network

server breach location significantly predicted the number of individuals affected, $p < .001$, as shown in Table 6.

Location of Breach and Type of Breach

Hacking/IT incident and network server were significant individually; therefore, a combined linear regression with these two predictors was considered. This combined regression showed that only network server was significant in predicting the number of individuals affected, $r = .141$, $r^2 = .020$, $F(2, 2013) = 20.536$, $p < .001$. See Table 7 for the combined regression.

Further analysis considered data breaches involving network server as the location of breach, combined with all the types of breaches, in relation to the dependent variable, the number of individuals affected. Only unauthorized access/disclosure was associated with network server for the number of individuals affected by the breach, $p = .054$. Surprisingly, hacking/IT incident was not significant when combined with network server, as revealed in Table 8.

Covered Entities, Business Associates, and States

One-way ANOVA tests were conducted to examine the relationship of covered entity, business associate presence, and state where the breach occurred with the number of individuals affected by the breaches. The correlation between business associate presence (yes = 1, no = 2) and number of individuals affected was significant, $F(3,2012)$, $p = .006$. In this sample, breaches with business associates present affected the greatest number of individuals ($M = 69{,}564$, $SD = 386{,}117$) when compared to breaches with no business associate present ($M = 28{,}253$, $SD = 229{,}993$). The correlation between covered entity type and number of individuals affected was significant, $F(3,2012) = 5.318$, $p = .001$. Business associates had the highest number of individuals affected ($M = 90{,}827$, $SD = 500{,}069$). A post hoc Tukey test revealed that only the mean scores for the healthcare provider and business associate types differed significantly from the mean score for all covered entity types. Figure 3 shows the mean number of individuals affected by each type of covered entity.

States and Regions

Table 9 shows the total number of individuals affected by breaches in each US region. In the DHHS data set, more than 72 million individuals were affected by breaches. The South region had the largest number of individuals affected ($n = 30{,}551{,}910$), and the Midwest region had the lowest number ($n = 11{,}804{,}845$). The correlation between regions and the number of individuals affected was not significant, $F(3, 47) = 0.257$, $p = .856$.

Discussion

This study provides insight into the relationships between data breach characteristics and the number of individuals affected by breaches. After we determined that the hacking/IT incident breach type and the network server breach location were the most significant individual predictors, all breaches involving both a type and location were considered. The relationship between breach type and breach location was then analyzed to find the most significant combination. Unauthorized access/disclosure and network server, in combination, were predictive of the number of individuals affected. In hindsight, this finding makes sense. For example, while an organization may lose millions of records from a server, it is not likely that millions of paper records would be stolen. This form of record precludes the theft of extremely large data sets.

Another interesting point noted was that hacking/IT incident was only marginally significant in predicting the number of individual records stolen, and none of the other types of breaches were significant. This observation led us to consider the combination of breach type with location because the combination of these variables could have affected the number of individual records breached. Interestingly, we found that unauthorized access/disclosure was significant when controlling for the network server location. This finding introduces a problem: What is the difference between a hacking/IT incident and an occurrence of unauthorized access/disclosure? A search of the DHHS website provided descriptions of the covered entity types, but the website lacked definitions of breach types and locations. Being able to distinguish between breach types and locations is important in understanding and preventing data breaches.

We also wanted to evaluate differences between the various covered entity types and business associates. Our results indicate that covered entity type and business associate presence were significant predictors of the number of individual records affected. Of interest, the average number of individual records affected by breaches involving business associates was almost four times that of breaches involving healthcare providers. Business associates are therefore likely to have more records affected than other covered entity types.

Finally, we examined whether geographic region might predict the number of individuals affected. Our analysis indicated no significance; however, we found that the South region had the greatest number of individual records affected. The 17 states in the southern region had on average had almost twice as many people affected as the Midwest states. While region was a significant predictor of number of records affected, differences between regions are of interest and may point to areas for improvement.

Limitations

Like all research, this study had several limitations. First, most breaches were reported by covered entities who self-identified as healthcare providers, with very few reports coming from business associates, health plans, or healthcare clearing houses. The DHHS breach portal does provide examples of the covered entity types, but because it allows covered entities to self-identify, the potential for error exists. Second, while data breaches were reported in all 50 US states and in the District of Columbia and Puerto Rico, there is obviously a bias toward population centers. The four states with the largest total numbers of individuals affected by breaches were Tennessee, California, New York, and Florida. Thus, the results may not be generalizable, as facilities in different states may experience larger breaches. Accordingly, future extractions of data from the DHHS OCR reporting database may generate different results. Finally, the number of individuals affected is a self-reported estimate. In reality, many organizations that experience data breaches may be unable to identify the number of records breached and therefore may err by reporting the total number of records in the breached system. The use of self-reporting also means that some healthcare organizations could have underreported the number of records breached because of estimation errors.

Conclusion

The number of people affected by data breaches continues to increase alarmingly despite healthcare professionals' greater awareness of the risk factors associated with these breaches. Problems related to breaches pose serious risks, such medical identity theft and ransomware attacks. Thus far, little research has been done to associate data breach characteristics with the number of individuals affected by the breach. This type of research is important in understanding factors associated with data breaches. Knowing the relevant factors can help organizations determine the risk of future data breaches. Determining how breach characteristics are associated with the number of individuals affected can lead to greater understanding of how to strengthen defenses, reinforce risk management plans, and reduce collateral damage from data breaches.

Diane Dolezel, EdD, RHIA, CHDA, is an assistant professor in the Department of Health Information Management at Texas State University in San Marcos, TX.

Alexander McLeod, PhD, is an associate professor in the Department of Health Information Management at Texas State University in San Marcos, TX.

Notes

1. McLeod, Alexander, and D. Dolezel. "Cyber-Analytics: Modeling Factors Associated with Healthcare Data Breaches." *Decision Support Systems* 108 (April 2018): 57–68.
2. US Department of Health and Human Services. "Breach Portal." Available at https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf (accessed June 1, 2017).
3. US Department of Health and Human Services. "Anthem Pays OCR $16 Million in Record HIPAA Settlement Following Largest U.S. Health Data Breach in History." October 15, 2018. Available at https://www.hhs.gov/about/news/2018/10/15/anthem-pays-ocr-16-million-record-hipaa-settlement-following-largest-health-data-breach-history.html?language=en.
4. US Department of Health and Human Services. "Judge Rules in Favor of OCR and Requires a Texas Cancer Center to Pay $4.3 Million in Penalties for HIPAA Violations." June 18, 2018. Available at https://www.hhs.gov/about/news/2018/06/18/judge-rules-in-favor-of-ocr-and-requires-texas-cancer-center-to-pay-4.3-million-in-penalties-for-hipaa-violations.html.
5. US Department of Health and Human Services. "Breach Notification Rule." Available at https://www.hhs.gov/hipaa/for-professionals/breach-notification/.
6. Ibid.
7. "Health Information Technology for Economic and Clinical Health (HITECH) Act." Title XIII of the American Recovery and Reinvestment Act of 2009. February 17, 2009.
8. McLeod, Alexander, and D. Dolezel. "Cyber-Analytics: Modeling Factors Associated with Healthcare Data Breaches."
9. Davis, Jessica. "Employee Error Exposed Data of 16,000 Blue Cross Patients Online for 3 Months." *Healthcare IT News*, September 21, 2018. Available at https://www.healthcareitnews.com/news/employee-error-exposed-data-16000-blue-cross-patients-online-3-months.
10. Davis, Jessica. "1.4 Million Patient Records Breached in UnityPoint Health Phishing Attack." *Healthcare IT News*, July 31, 2018. Available at https://www.healthcareitnews.com/news/14-million-patient-records-breached-unitypoint-health-phishing-attack.
11. Gibbs, David, Karima Lalani, and Alexander McLeod. "Beware the Internet's Dark Side: What HIM Professionals and Patients Should Know About the.Dark.Web." *Journal of AHIMA* 88, no. 8 (2017): 30.
12. Ponemon Institute. *Cost of Cyber Crime Study*. 2017. Available at https://cdn2.hubspot.net/hubfs/85462/2018/2018_VUENUE/2018_BLACK%20HAT/Accenture-2017CostCybercrime-US-FINAL.pdf?t=1521831469946.
13. Spence, Nikki, Niharika Bhardwaj, David P. Paul III, and Alberto Coustasse. "Ransomware in Healthcare Facilities: A Harbinger of the Future?" *Perspectives in Health Information Management* (Summer 2018): 1–22.
14. McLeod, Alexander, and D. Dolezel. "Cyber-Analytics: Modeling Factors Associated with Healthcare Data Breaches."
15. Stachel, Richard D., and Marilyn DeLaHaye. "Security Breaches in Healthcare Data: An Application of the Actor-Network Theory." *Issues in Information Systems* 16, no. 2 (2015): 185–94.

16. Angst, Corey, Emily Block, and Ken Kelley. “When Do IT Security Investments Matter? Accounting for the Influence of Institutional Factors in the Context of Healthcare Data Breaches.” *MIS Quarterly* 41, no. 3 (2017): 893–916.
17. Venkatesh, J., P. Cherurveettil, and P. Sivashanmugam, P. “Using a Prediction Model to Manage Cyber Security Threats.” *The Scientific World Journal* (2015): 1–5. doi:10.1155/2015/703713.
18. Ayyagari, Ramakrishna. “An Exploratory Analysis of Data Breaches from 2005–2011: Trends and Insights.” *Journal of Information Privacy and Security* 8, no. 2 (2012) :33–56.
19. Ponemon Institute. *Sixth Annual Benchmark Study on Privacy & Security of Healthcare Data*. 2016. Available at https://www.ponemon.org/local/upload/file/Sixth%20Annual%20Patient%20Privacy%20%26%20Data%20Security%20Report%20FINAL%206.pdf.
20. Wikina, Suanu. “What Caused the Breach? An Examination of Use of Information Technology and Health Data Breaches.” *Perspectives in Health Information Management* (Fall 2014): 1–16.
21. Ronquillo, J., E Winterholler, K. Cwikla, R. Szymanski, and C. Levy. “Health IT, Hacking, and Cybersecurity: National Trends in Data Breaches of Protected Health Information.” *JAMIA Open* 1, no. 1 (2018): 15–19. doi:10.1093/jamiaopen/ooy019.
22. US Department of Health and Human Services. “Health Information Privacy: Covered Entities and Business Associates.” Available at https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html.
23. Field, A. *Discovering Statistics Using IBM SPSS Statistics*. 4th ed. Los Angeles: Sage, 2013.

Figure 1

Department of Health and Human Services Breach Reporting Portal

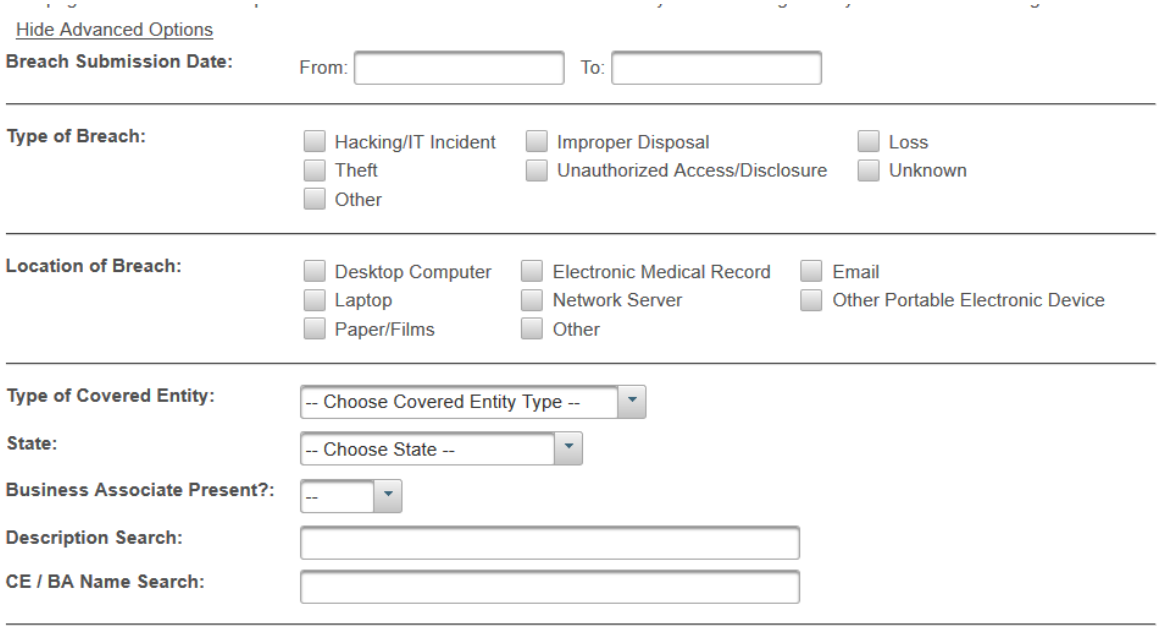

Figure 2

Number of Individuals Affected by State

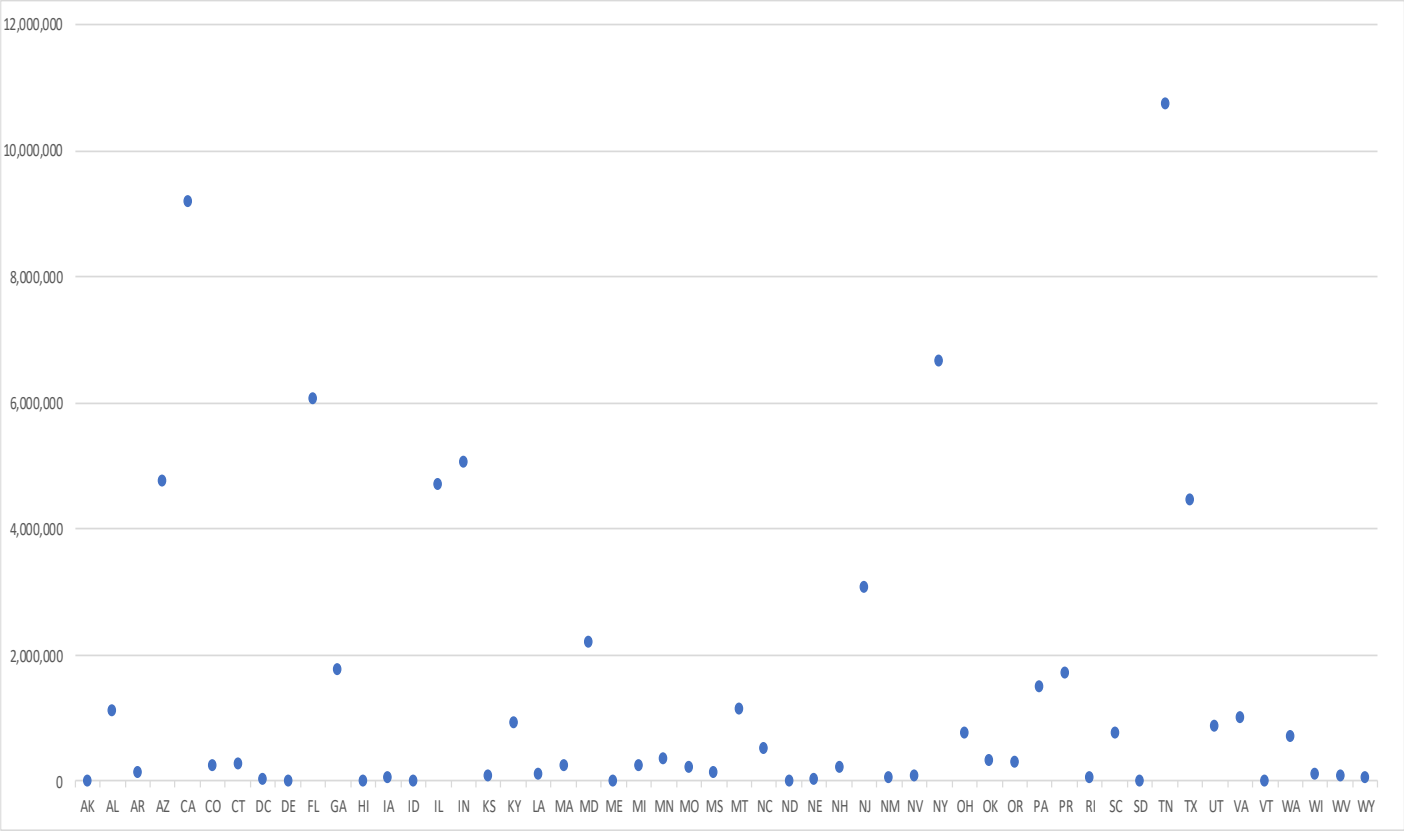

Figure 3

Mean Number of Individuals Affected by Covered Entity Type

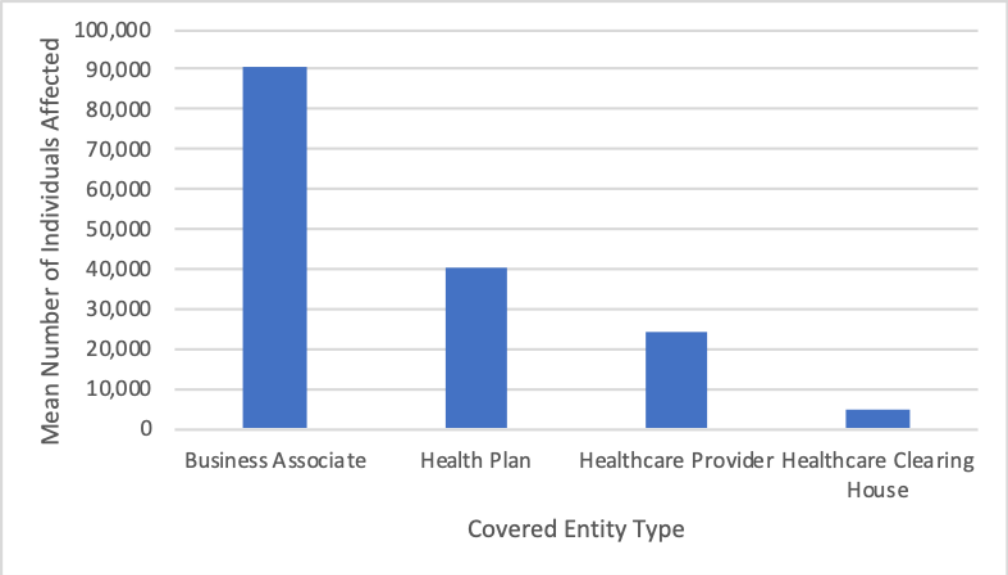

Table 1

Number of Individuals Affected and Breaches Reported by Year, 2009–2013

Year	Number of Individuals Affected	Percentage of Individuals Affected	Number of Breaches Reported	Percentage of Reported Breaches
2009	134,773	0.18	18	0.89
2010	5,932,276	8.05	199	9.87
2011	8,262,158	11.21	199	9.87
2012	2,854,525	3.87	218	10.81
2013	7,022,786	9.52	277	13.74
2014	17,452,393	23.67	314	15.58
2015	13,469,510	18.27	265	13.14
2016	15,842,512	21.49	303	15.03
2017	2,584,790	3.51	198	9.82
2018	175,849	0.24	25	1.24
Totals	73,731,572	100.00	2016	100.00

Table 2

Data Breaches by Types and Locations

Data Breach Locations	**Number**	**Percentage**
Desktop computer	240	11.9
Electronic medical record	136	6.7
Email	230	11.4
Laptop	385	19.1
Network server	371	18.4
Other portable electronic device	221	11.0
Paper/films	509	25.2
Other	434	21.5
Total	2,019	
Data Breach Type	**Number**	**Percentage**
Hacking/IT incident	337	15.9
Theft	843	39.8
Improper disposal	71	3.3
Loss	168	7.9
Unauthorized access/disclosure	588	27.7
Other	97	4.6
Unknown	16	0.8
Total	2,120	

Table 3

Individuals Affected and Breaches Reported by Covered Entity Type by Year, 2009–2018

Year	Breaches Reported by Covered Entity Type, *n* (%)			
	Business Associate	**Health Plan**	**Healthcare Clearing House**	**Healthcare Provider**
2009	3 (0.96)	1 (0.38)	0	14 (0.97)
2010	44 (14.06)	21 (8.05)	0	134 (9.32)
2011	44 (14.06)	19 (7.28)	1 (25.00)	135 (9.39)
2012	40 (12.78)	23 (8.81)	1 (25.00)	154 (10.71)
2013	64 (20.45)	18 (6.90)	2 (50.00)	193 (13.42)
2014	77 (24.60)	41 (15.71)	0	196 (13.63)
2015	12 (3.83)	62 (22.22)	0	195 (13.56)
2016	18 (5.75)	48 (18.39)	0	237 (16.48)
2017	9 (2.88)	27 (10.34)	0	162 (11.27)
2018	2 (0.64)	5 (1.92)	0	18 (1.25)
Totals	313	261	4	1,438
Year	**Individuals Affected by Covered Entity Type, *n* (%)**			
	Business Associate	**Health Plan**	**Healthcare Clearing House**	**Healthcare Provider**
2009	91,400 (0.32)	3,800 (0.04)	0	39,573 (0.11)
2010	1,529,729 (5.38)	3,564,344 (34.00)	0	838,203 (2.41)
2011	4,036,804 (14.20)	89,977 (0.86)	1,250 (7.04)	4,134,127 (11.88)
2012	1,146,711 (4.03)	336,265 (3.21)	10,000 (56.33)	1,361,549 (3.91)
2013	1,058,760 (3.72)	97,555 (0.93)	6,504 (36.63)	5,859,967 (16.84)
2014	12,988,487 (45.69)	2,247,146 (21.44)	0	2,216,760 (6.37)
2015	3,954,463 (13.91)	3,119,905 (29.76)	0	6,395,142 (18.38)
2016	3,552,724 (12.50)	817,847 (7.80)	0	11,471,941 (32.96)
2017	66,227 (0.23)	166,636 (1.59)	0	2,351,927 (6.76)
2018	3,656 (0.01)	39,418 (0.38)	0	132,775 (0.38)
Totals	28,428,961	10,482,893	17,754	34,801,964

Table 4

US Census Regions and States

Regions	**States in Each Region**
Region 1 (Northeast)	Division 1 (New England)): Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont Division 2 (Middle Atlantic): New Jersey, New York, Pennsylvania
Region 2 (Midwest)	Division 3 (East North Central): Illinois, Indiana, Michigan, Ohio, Wisconsin Division 4 (West North Central): Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
Region 3 (South)	Division 5 (South Atlantic): Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia Division 6 (East South Central): Alabama, Kentucky, Mississippi, Tennessee Division 7 (West South Central): Arkansas, Louisiana, Oklahoma, Texas
Region 4 (West)	Division 8 (Mountain): Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming Division 9 (Pacific): Alaska, California, Hawaii, Oregon, Washington

Table 5

Regression Coefficients for Types of Breach

Model	***B***	***SE***	***t***	***p***	**Tol.**	**VIF**
(Constant)	50,869	24,452	2.08	.04		
Hacking/IT incident	52,049	27,279	1.91	.06	.34	2.91
Improper disposal	−32,906	37,106	−0.89	.38	.76	1.32
Loss	−28,988	28,373	−1.02	.31	.58	1.73
Other	−34,263	32,797	−1.05	.30	.72	1.39
Theft	−19,585	24,532	−0.80	.43	.24	4.12
Unauthorized access/disclosure	−35,884	246,995	−1.45	.15	.28	3.55
Unknown	111,161	67,689	1.64	.10	.99	1.02