

# Timing and Scale of U.S. Data Breaches: Organizational, Breach-Type, and Information-Type Determinants

Shizhen (Jasper) Jia  
Quinnipiac University  
[shizhen.jia@quinnipiac.edu](mailto:shizhen.jia@quinnipiac.edu)

Tianyu (Bell) Pan  
University of Florida  
[tpan1@ufl.edu](mailto:tpan1@ufl.edu)

Guohou Shan  
Northeastern University  
[g.shan@northeastern.edu](mailto:g.shan@northeastern.edu)

## Abstract

*This study examines the factors that influence the scale and timeliness of data breaches in an era of escalating digital threats, where the average breach cost reached \$4.88 million in 2024. Analyzing 19,255 incidents from the Privacy Rights Clearinghouse using regression models, the findings reveal that insider-led breaches result in a reporting lag of approximately 79 days longer than external breaches, although affecting fewer records. Additionally, organizations in the healthcare, government, and education sectors experience breaches that are over 100% larger in scale than those in other sectors. The analysis also demonstrates that data sensitivity has a significant impact on breach severity and disclosure dynamics. These insights offer valuable contributions to crisis communication theories, inform policymakers in developing nuanced data breach laws, and equip organizational leaders to tailor incident response strategies effectively.*

**Keywords:** Data breach, cybersecurity, reporting delay, breach notification, incident response, information security, risk management, organizational determinants

## 1. Introduction

The digital transformation of the global economy has introduced significantly opportunities for innovation and efficiency, but it has also increased the threat of data breaches. In recent years, the scale, frequency, and financial impact of these breaches have risen sharply, making cybersecurity a top concern for organizations, governments, and individuals (Alawida et al., 2022; Toledano, 2024). In 2024, the average cost of a data breach soared to a record of \$4.88 million, a 10% increase from the prior year (Mohsin, 2025). Globally, cybercrime costs are projected to reach \$10.5 trillion by 2025 (Dalei & Kandpal, 2025), highlighting its threat to economic stability.

In 2023, there was a 78% surge in publicly reported data breaches compared to 2022, affecting over 353 million individuals in the U.S. alone (Kapko, 2024).

Notable incidents, such as the ransomware attack on Change Healthcare (Neprash et al., 2024), highlight the tangible harm caused to individuals and organizations alike. This attack alone disrupted national pharmacy operations, compromised the data of about one-third of Americas, and incurred direct costs of at least \$872 million for UnitedHealth Group in early 2024 (Alder, 2025). Similarly, breaches at AT&T and loanDepot further exemplify vulnerabilities across critical sectors (Nygard, 2025).

While increases in breach frequency and cost are well-documented, a nuanced understanding of the factors influencing breach scale and disclosure timing remains limited. Academic literature presents conflicting findings, with some suggesting consistent upward trends in breach frequency, while others dispute this view due to biases in data collection, reporting laws, and media attention (Molitor et al., 2024; Edwards et al., 2016). This study aims to clarify these inconsistencies by utilizing a large-scale, standardized dataset.

Additionally, there is a lack of research that comprehensively examines the determinants of breach characteristics. Most studies focus on either reporting delays or breach scale independently (Avanzi et al., 2025; Nikkiah & Grover, 2022; Mitra & Ransbotham, 2015), leaving key questions unanswered. This paper aims to investigate how breach origin (external vs. insider), industry sector, and data sensitivity impact both the severity of incidents and firms' responses. **Three core research questions guide this study:**

**RO1:** *How do different breach types (e.g., insider threats (INSID), external hacking (HACK), and physical theft (PHYS)) affect the time lag between the date of the breach and the date it is reported?*

**RO2:** *To what extent does the type of organization (e.g., healthcare (MED), financial services (BSF), government (GOV)) explain variations in the total number of records impacted by a breach?*

**RO3:** *Does the nature of compromised data (e.g., sensitive health information, financial credentials, government identifiers) influence both the extent of the impact (number of records exposed) and the promptness of public disclosure?*

This study makes several novel contributions to the fields of information systems, cybersecurity, public policy, and risk management. Firstly, this research provides strong empirical evidence to enhance theories of crisis communication, reputation management, and deterrence. It demonstrates that insider-led breaches lead to longer reporting delays, challenging simplistic disclosure models that focus solely on external pressures. The findings emphasize the importance of internal dynamics, detection capabilities, and investigative complexities in shaping a firm's response, offering a better understanding of behavior during security crises.

Secondly, the insights are valuable for both industry practitioners and government regulators. For organizational leaders, such as Chief Information Security Officers (CISOs), the results highlight the need for tailored incident response playbooks for different threat types, as the optimal response for insider threats differs significantly from that for ransomware attacks. For policymakers, the study provides important evidence for the debate on a potential federal data breach notification standard. The variability in reporting patterns by industry sector and breach type suggests that a one-size-fits-all regulation may be less effective than a more nuanced approach, particularly in evaluating and refining existing laws, such as the Health Insurance Portability and Accountability Act (HIPAA).

## 2. Literature Review

### 2.1. The Evolving Nature and Cost of Data Breaches

An increasing financial impact and growing complexity characterize the landscape of data breaches. According to the IBM Cost of a Data Breach Report, the global average cost in 2024 reached \$4.88 million, with the U.S. experiencing the highest average at \$9.36 million per incident (IBM, 2024; Mohsin, 2025). These costs include directly expenses, such as forensic investigation, legal fees, regulatory fines, and the costs of notifying affected individuals, which averaged \$370,000 in 2023 (IBM, 2024). However, the bulk of these expenses arise from indirect costs associated with customer turnover, reputational damage, and operational downtime during recovery (Kuipers & Schonheit, 2022).

Furthermore, the financial burden varies significantly across industries (Sen & Borle, 2015; Richardson et al., 2019; Haislip et al., 2019). For fourteen consecutive years, the healthcare sector has had the highest average breach cost at \$9.77 million in 2024 (Alder, 2025). This is primarily due to stringent HIPAA regulations, the high value of protected health

information (PHI) on the dark web, and the complex IT environments within healthcare organizations. The financial services sector follows with an average cost of \$6.08 million per breach (IBM, 2024), reflecting the direct monetary value of its protected data. This variation among sectors highlights the study's focus on whether organization type is a significant predictor of breach scale.

### 2.2. Regulatory Compliance and Crisis Management

The U.S. legal framework for data breach notification is complex, consisting of both federal and state laws with varying compliance requirements (Park, 2019; Almeida et al., 2022). At the federal level, the HIPAA Breach Notification Rule requires covered entities to notify affected individuals and the Department of Health and Human Services (HHS) within 60 days of discovering a breach of unsecured PHI (Subramanian et al., 2024). However, regulators warn that any unjustifiable delay within this timeframe can still lead to non-compliance (Burgess, 2022). In contrast, the European Union's General Data Protection Regulation (GDPR) imposes stricter requirements, mandating a 72-hour notification to supervisory authorities (Marovic & Curcin, 2020).

At the state level, breach notification laws vary significantly across all 50 states, with deadlines ranging from as short as 5 days in Iowa to 30 to 45 days in states, such as Florida and Ohio (Nikkhah & Grover, 2022). Many states use the vague standard of "without unreasonable delay." This variation complicates incident response, as legal obligations change based on the residency of affected individuals, providing a baseline for analyzing corporate disclosure behavior.

Companies' decisions during data breaches are influenced by theories related to crisis and reputation management (Gwebu et al., 2018; Wang et al., 2025). Organizations often delay notifications due to fears of reputational damage, declines in stock prices, and loss of customer trust (Kuipers & Schonheit, 2022). Research indicates that stock prices can drop by 2% to 7% following breach disclosures (Johnson et al., 2017; Rosati et al., 2019), and many consumers tend to lose trust in organizations after a breach, choose to take their business elsewhere (Martin et al., 2017; Ou et al., 2022).

Nonetheless, recent studies suggest that the reputational impact of breaches is not uniform (Confente et al., 2019; Makridis, 2021). While large breaches typically result in significant reputational loss, smaller breaches may enhance brand visibility and familiarity due to increased media attention, creating a complex incentive structure (Syed, 2019). For many non-catastrophic incidents, the reputational costs may

not deter disclosure and could even be viewed as a positive (Guha & Kandula, 2012). This perspective emphasizes that the decisions to delay or expedite notifications are strategic and influenced by the perceived severity of the breach.

### 2.3. Determinants of Breach Reporting Timeliness

The literature highlights several factors influencing the delay between breach occurrence and public reporting, forming the basis for this study's RQ1.

**2.3.1. Breach Origin (Insider vs. Hacker).** The source of a breach is a crucial determinant of reporting timeliness. Insider attacks, executed by employees or trusted individuals, are generally more difficult to detect and take longer to resolve than external attacks (Dearden et al., 2023). For instance, one study indicated that the median detection time for employee theft schemes exceeds a year. This extended timeline is due to insiders using their legitimate access to evade security alarms designed for external threats, with detection often relying on manual audits or whistleblowers (Homoliak et al., 2019). This evidence supports the hypothesis that breaches categorized as insider will experience significantly longer reporting lags than those classified as hacker.

**2.3.2. Organizational Sector.** The industry in which an organization operates also significantly affects reporting delays (Nikkhah & Grover, 2022). Recent research shows that the education sector has the longest reporting lags, averaging 4.8 months for ransomware-related breaches (Halikias, 2024). This may stem from limited cybersecurity resources, complex decentralized networks, and reduced regulatory pressure compared to other sectors. Conversely, industries that are heavily regulated, such as finance and healthcare, tend to have quicker response times due to stringent laws like the GLBA and HIPAA (Isibor, 2024). However, even in these sectors, the complexity of investigating significant breaches can result in extended breach lifecycles from intrusion to resolution.

**2.3.3. The Investigation Process.** A legitimate reason for reporting delays is the time required for thorough forensic investigations (Brown, 2015; Chernyshev et al., 2019). Upon discovering a potential incident, organizations must determine its scope, identify the root cause, assess compromised data, and notify affected individuals (Sen & Borle, 2015). This process can be lengthy, particularly for sophisticated attacks. Many organizations prefer to wait for confirmation of data theft, often dependent on whether the attackers publicly release the data on the dark web, before notifying regulators or the public (Chertoff, 2017). Although regulations like HIPAA allow for a "reasonable"

investigation period, they also caution against undue delays.

### 2.4. Determinants of Breach Scale and Severity

The literature identifies key factors influencing the scale of data breaches, which is central to RQ2 of this study.

**2.4.1. Influence of Organization Type.** The type of organization significantly affects its vulnerability to breaches. The healthcare sector remains a primary target due to the high value of PHI, the extensive data held by providers and insurers, and vulnerabilities in complex IT systems often reliant on outdated technology and insufficient cybersecurity funding (Ira Bedenbaugh & Brown, 2024). Similarly, the financial sector is attractive to attackers due to the direct monetary value of its data, making it vulnerable to large-scale fraud and identity theft. Government and educational institutions, with their vast amounts of sensitive personal data, are also at high risk (Pomerleau & Lowery, 2020; George et al., 2024).

**2.4.2. Heavy-Tailed Nature of Breach Sizes.** Research indicates that data breaches show a heavy-tailed distribution, such as a log-normal distribution, rather than a normal distribution (Edwards et al., 2016). This suggests that while most breaches are small, significant events involving millions of records are more common than expected. This has crucial implications for risk management. For risk managers, it means that "black swan" events are foreseeable threats.

### 2.5. The Criticality of Compromised Information

The literature also establishes a link between data sensitivity and both regulatory scrutiny and attacker motivation, forming the basis for this study's RQ3.

**2.5.1. Defining and Valuing Data Sensitivity.** Data is diverse, with varying values and exposure risks. Research indicates that sensitive data types of command significantly higher prices on the dark web (Wang et al., 2025). For example, healthcare records can be valued at up to 50 times that of credit card numbers. Furthermore, personally identifiable information (PII) of customers and employees remains the most frequently compromised, highlighting its appeal to attackers.

**2.5.2. The Link Between Data Sensitivity and Disclosure.** The sensitive of the compromised data is crucial for breach notification requirements (Gibson & Harfield, 2023). Breaches involving highly sensitive information, such as Social Security numbers or health records, almost always necessitate notification under state and federal laws. Regulations impose stringent requirements specifically for these data categories

(Herath et al., 2024), suggesting that organizations in sectors handling sensitive information face intense regulatory pressure to disclose information promptly and accurately. However, larger breaches in these sectors often lead to more protracted investigations (Cheng et al., 2017), creating a complex dynamic worthy of exploration in this study.

### 3. Methodology

#### 3.1. Dataset

This study employs a quantitative approach using Privacy Rights Clearinghouse's Data Breach Chronology as the data source. The dataset comprises 19,255 publicly reported breach incidents in the United States, covering breaches from 2005 through 2024. This large-scale, standardized dataset mitigates many sampling biases present in earlier studies (e.g. media-based or self-reported data), thereby offering a more reliable basis for regression analysis (Kim et al., 2024).

#### 3.2. Dependent Variable

Two dependent variables are analyzed. *Breach disclosure timing* is measured as the number of days between the breach's occurrence (or discovery) date and the date of public reporting ("days to report"). This captures the reporting lag which addresses RQ1. *Breach severity* is measured by the maximum number of records impacted in the incident, a proxy for breach scale (RQ2). Because the distribution of breach sizes is highly skewed (with most incidents being small and a few extremely large), we analyze severity in both absolute and logarithmic terms. In the raw count models, the Max Records variable represents the total number of individual records or identities compromised. In parallel, a Log(Records) transformation is used to normalize the heavy-tailed distribution, reducing the influence of outliers (e.g., mega-breaches involving millions of records) on the regression estimates. This dual-specification approach provides robustness: significant findings in the log-scale model indicate proportional effects on breach size, complementing the count model which retains absolute magnitudes.

#### 3.3. Independent Variable

The key predictors correspond to breach origin, organizational sector, and data sensitivity, aligned with RQ1–RQ3. *Breach origin* is operationalized via an Insider Attack dummy, indicating whether the breach was caused by an internal actor (e.g., a malicious employee) as opposed to an external attack. In the

Privacy Rights Clearinghouse data, breaches are classified by type (e.g., HACK for external cyberattacks, INSD for insider incidents, PHYS for physical loss, etc.). We created a binary indicator INSIDER that equals 1 for incidents involving internal perpetrators (INSD) and 0 for all external breaches (with external hacks, theft of devices, and other outside attacks as the reference category). This captures the theorized detection and response differences between insider and outsider breaches.

To examine *industry effects* (RQ2), we included Organization-Type categorical variables denoting the sector of the breached entity. The dataset classifies organizations into sectors such as Financial Services (BSF), Retail (BSR), Other Business (BSO, including technology and professional services), Healthcare/Medical (MED), Government/Military (GOV), Education (EDU), and Non-Profit (NGO). We incorporated these as dummy variables in the regression (with one category omitted as the baseline). This enables an assessment of how breach severity differs, for example, in healthcare, government, or education relative to the baseline sector. The choice of baseline category was based on the largest and most heterogeneous group ("Other Business") to serve as a reference point encompassing a broad range of private-sector firms. This way, positive coefficients on sector dummies indicate a larger breach size (or longer delay) compared to a typical private business, while negative coefficients indicate smaller breaches (or faster reporting) than the baseline. We note that sector dummies also indirectly capture differences in regulatory environment across industries (such as HIPAA regulations for healthcare or GLBA for finance), which can influence both breach scale and response speed.

Finally, to address RQ3 regarding *data sensitivity*, we considered the types of information compromised in each incident. The PRC data lists whether sensitive personal data were involved (e.g., Social Security numbers, financial account details, medical records, etc.). Breaches involving highly sensitive information (like SSNs or health records) almost always trigger mandatory notification under state and federal laws. We initially intended to include dummy variables for the presence of certain sensitive data (such as a Health Information indicator, Financial Information indicator, etc.). However, these attributes showed high collinearity with industry (for instance, nearly all breaches in healthcare involve PHI, and financial-sector breaches often involve account data), making it difficult to disentangle industry effects from data-type effects in the regression. Instead, we controlled for regulatory context by introducing state fixed effects in our models, and we interpret the influence of data sensitivity through the

lens of industry and regulatory variables. The state fixed effects (a dummy for each U.S. state where the breached entity is located or where notifications were filed) account for disparate state-level notification requirements and any region-specific factors. In practice, states impose varying deadlines and legal thresholds (e.g., Florida's 30-day rule vs. "without unreasonable delay" in other states), which could systematically affect disclosure timing. Including state dummies absorbs this variation, ensuring that our coefficients for breach type and industry are not biased by geographic regulatory differences. Year fixed effects were not included, as our focus is cross-sectional determinants; moreover, the inclusion of state laws (many of which evolved over time) partly captures temporal changes in the breach disclosure landscape.

### 3.4. Estimation Approach

We applied ordinary least squares (OLS) regression models for each dependent variable. Four models were specified for the two primary outcomes (reporting lag in days, and records lost): a base model with the key predictor of interest and a full model with all controls. Specifically, for *Reporting Lag*, Model 1 includes only the INSIDER breach dummy (testing RQ1), while Model 2 adds the full set of industry dummies and state fixed effects. Similarly, for *Records Impacted*, Model 3 includes only the INSIDER dummy (testing whether insiders differ in breach size) and Model 4 adds all industry dummies and state controls (addressing RQ2). In addition, we ran parallel models for *Log(Records)*: Model 5 with INSIDER only, and Model 6 with industry and state controls. All regressions use robust standard errors to account for heteroskedasticity in breach sizes. The variance inflation factors for the full models were inspected and did not indicate severe multicollinearity (the highest VIF was under 2.5, as the categorical variables were well-balanced in the sample).

## 4. Results

### 4.1. Descriptive Overview

Before turning to the regressions, it is noteworthy that the average public reporting delay in this dataset is about 59 days (median 30 days), reflecting the common practice of taking several weeks to investigate breaches before disclosure. However, the range is wide: many breaches were disclosed within a week, while some extreme cases took over a year to come to light. Breach sizes are even more skewed. The median number of records exposed is approximately 2,300, but the mean is

around 58,000, inflated by a long tail of mega-breaches. This heavy-tailed distribution of breach severities is consistent with prior research, reinforcing our use of a log transformation to capture proportional effects. Against this backdrop, we discuss the regression findings for each research question.

### 4.2. Regression Results

The regression results provide strong evidence that insider-caused breaches are associated with substantially longer disclosure delays. As shown in Table 1, the coefficient on the Insider breach dummy in the reporting lag models is positive and significant. In the baseline Model 1 (with no other controls), insider incidents average 70.2 days longer to report than external breaches ( $p < 0.05$ ). In the full Model 2, which controls industry and state, the effect size increases slightly to 78.9 days ( $p < 0.01$ ). This confirms RQ1 that breaches originating internally tend to be reported much later. In substantive terms, an organization experiences an extra delay of roughly two and a half months when the breach is perpetrated by an insider versus an external hacker. This result aligns with the notion that insider attacks are harder to detect and often require lengthy investigations before public disclosure (Wang et al., 2019). Importantly, the insider effect remains robust even after accounting for sector and state influences, suggesting it is a general phenomenon not confined to any one industry or region (Wang et al., 2015).

We find considerable sectoral variation in the number of records compromised, addressing RQ2. Several industry dummy variables are significant predictors of breach size in the full models. Because Model 4 (raw record counts) has very low explanatory power due to extreme outliers, we focus interpretation on Model 6, which uses the log-transformed records and has a higher adjusted  $R^2$  (~5.0%). In Model 6, the coefficients for Healthcare (MED), Education (EDU), Retail (BSR), and Nonprofit (NGO) are all positive and statistically significant ( $p < 0.001$ ), indicating that breaches in those sectors tend to expose substantially more records than the baseline category (Other Business). For example, the coefficient for healthcare is 0.745 ( $SE = 0.09$ ,  $p < 0.001$ ), which implies that holding other factors constant, a breach in a healthcare organization involves about 110% more records than a breach at a baseline business. Education breaches show an even larger effect: the coefficient 0.972 ( $p < 0.001$ ) corresponds to a 164% increase in records affected relative to the baseline. Government breaches also exhibit above-average sizes, though with a smaller magnitude that is marginally significant ( $p < 0.10$ ). Retail business breaches have a coefficient of 0.807 ( $p < 0.001$ ), roughly 125% larger in scope than baseline,

consistent with the high-profile mega-breaches in the retail sector (e.g., Target, Home Depot) (Wolf et al., 2023). Nonprofit organizations see breaches about 97% larger than baseline. By contrast, Financial Services (BSF) firms do not show a statistically significant difference from the baseline in our models. This suggests that while financial institutions suffer costly breaches, the sheer scale (in terms of individuals impacted) is often greater in sectors like healthcare, education, and retail, which may hold larger volumes of personal data in more centralized systems.

Our results indirectly address RQ3 through the combined effects of industry and the state fixed effects (which proxy regulatory stringency). The regression evidence suggests two main insights: First, sectors dealing in sensitive personal data under strict regulation (healthcare with PHI, finance with financial identifiers) do not report breaches significantly faster than others in practice. In fact, healthcare firms appear to have some of the largest breaches and also do not show significantly shorter delays than baseline businesses. For instance, the healthcare sector dummy in the lag Model 2 is positive (20.2 days,  $p < 0.05$ ), indicating slightly longer reporting times than baseline, even though HIPAA imposes a 60-day notification deadline. Financial sector breaches have a small and non-significant negative coefficient in Model 2, hinting they might be reported a bit faster on average (consistent with stricter oversight and perhaps more frequent detection by regulators or customers), but this effect was not statistically robust. Second, breaches involving large numbers of records tend to undergo protracted investigation and containment efforts before. This is evidenced by our earlier finding that industry sectors like education and government (which store extensive sensitive personal data) have some of the longest reporting delays *and* the largest breaches. The education sector, for example, not only had the largest average lag in prior studies (4.8 months on average for ransomware) but our data also shows very large breaches (education breaches are 2.64 times larger than baseline on average). Taken together, these patterns support RQ3's premise that the nature of compromised information influences breach outcomes: organizations handling highly sensitive information face intense regulatory and public scrutiny to notify promptly, yet when a massive breach occurs, the complexity of the incident can slow down the disclosure process.

**Table 1. Regression Results**

Variable / Model	Model 1: Days	Model 2: Days	Model 3: Records	Model 4: Records	Model 5: Log(Records)	Model 6: Log(Records)
Insider breach (INSD)	70.21* (29.41)	78.86** (29.53)	— 313,564 .5** (120,000)	— 466,664 .1** (170,000)	— 1.626*** (0.16)	— 1.249*** (0.16)

Financial services (BSF)		1.584 (5.76)		44,122.5 (180,000)		0.406*** (0.07)
Retail business (BSR)		43.69*** (6.87)		675,004.2 (830,000)		0.807*** (0.09)
Education (EDU)		46.69*** (12.39)		— 162,586.8 (150,000)		0.972*** (0.11)
Government (GOV)		79.73** (24.38)		— 127,552.8 (130,000)		0.304+ (0.16)
Healthcare (MED)		20.24* (8.41)		— 80,458.9 (110,000)		0.745*** (0.09)
Non-profit (NGO)		8.950 (9.90)		— 276,504.0* (120,000)		0.678*** (0.15)
Other/Unknown sector		— 29.02** (10.15)		— 218,154.5 (140,000)		0.323+ (0.19)
Constant	151.9*** (2.09)	236.0** (87.61)	316,912.8** (120,000)	46,152.1 (140,000)	5.080*** (0.02)	4.491*** (0.47)
State Fixed Effects	No	Yes	No	Yes	No	Yes
Observations	19,255	19,255	19,255	19,255	19,255	19,255
R <sup>2</sup>	0.0011	0.0138	0.0000	0.0014	0.0043	0.0570

## 5. Discussion and Implications

### 5.1. Slower Disclosure, Smaller Scale

This study reveals a notable paradox in insider breaches: they are often reported publicly much later yet involve significantly fewer records than external attacks. Specifically, insider breaches have a 79-day longer reporting lag and encompass 71% fewer records. This reflects the distinct nature of insider threats. The reporting delay can be divided into two phases: pre-discovery and post-discovery. Research indicates that the pre-discovery phase is considerably longer for insider threats (Bakker et al., 2017). Unlike external hacking attempts, which typically trigger immediate alerts within an organization's security information and event management (SIEM) systems, insiders use their actions more challenging to detect.

Consequently, insider breaches are often uncovered through manual audits, employee tips, or the eventual public emergence of stolen data, which can take months or even years. Once an insider threat is identified, the subsequent investigation becomes more complex. It involves not only IT and security teams but also Human Resources, the legal department, and senior management, introducing various procedural, political, and legal challenges. Factors such as employee privacy

rights and the need to build a case for termination or legal action can significantly delay public disclosure.

Moreover, the limited scale of insider breaches aligns with the motivations behind them. While external hackers typically aim for large-scale data exfiltration for financial gain, insiders usually pursue specific objectives, such as acquiring customer lists or proprietary information. This targeted approach results in fewer compromised records compared to the broad, indiscriminate nature of many external hacks. These findings highlight the inadequacy of a one-size-fits-all approach to incident response and regulation. Organizations must prioritize investment in advanced internal threat detection programs to identify anomalous behavior among trusted users rather than solely focusing on external threat mitigation.

## 5.2. Sectoral Variation in Breach Severity

Our analysis reveals significant variation in breach severity across sectors, with healthcare, education, and government experiencing breaches that are, on average, over twice as large as those in other industries. This aligns with prior findings that these sectors maintain vast amounts of sensitive personal data and often lack robust cybersecurity infrastructure (Ira Bedenbaugh & Brown, 2024; George et al., 2024). Regulatory frameworks like HIPAA also mandate disclosure of any PHI breaches affecting over 500 individuals, ensuring greater visibility of such incidents (Subramanian et al., 2024).

Education and government sectors similarly manage expansive datasets, such as student, employee, and citizen records, but often operate under budgetary and technological constraints that slow detection and response (Nikkhah & Grover, 2022). Recent news show education has among the longest average reporting delays for ransomware events, exacerbating breach scale (Merod, 2025). Government agencies, while better funded, are frequent targets of state-sponsored attacks and often maintain centralized repositories that, when compromised, lead to outsized impacts (Holt et al., 2023).

Retail breaches are also large in scope, particularly due to high transaction volumes and frequent exposure of payment data. Financial services, by contrast, did not show significantly greater breach sizes in our regression, possibly reflecting stronger compliance with GLBA and PCI-DSS standards, and more mature incident response capabilities (Greig, 2024).

Non-profits showed larger-than-expected breach sizes despite lacking regulatory mandates, reinforcing the idea that limited security resources and high-value data repositories can lead to greater breach exposure. While regulation in sectors like healthcare and finance

improves breach detection and disclosure, it does not prevent large-scale incidents. Conversely, under-regulated sectors face not only greater risk but also weaker response mechanisms.

## 5.3. Implications

**5.3.1. Implications for Crisis Response and Organizational Practice.** The results underscore the need for differentiated incident response plans that account for the nature of the breach. As evidenced by the stark contrast between insider and external breaches, a “one-size-fits-all” response strategy is. Organizations should develop specialized playbooks for insider threats, which involve unique challenges such as internal investigations, potential involvement of law enforcement for employee misconduct, and legal/privacy considerations when dealing with an internal perpetrator. Our finding that insiders are associated with 79-day longer disclosure lags suggests that many firms struggle with swift response in these scenarios. To reduce this lag, firms could invest in advanced internal threat detection and monitoring systems (e.g., user behavior analytics to flag anomalous insider activity). They should also conduct regular audits and institute whistleblower channels, as literature indicates insider incidents are often uncovered via manual oversight or tips. By proactively identifying insider misuse, organizations can shorten the *pre-discovery phase* of the breach lifecycle, thereby enabling quicker notification. Our findings support theories of crisis communication that emphasize tailoring the response to the situation. In practice, a breach caused by an employee’s data theft might warrant a different communication strategy (perhaps a more apologetic tone and internal accountability measures) compared to a breach caused by an external hacker (where the focus might be on technical remediation and assurances of improved defenses). This aligns with insights from Gwebu *et al.* (2018) that an organization’s initial response strategy and reputation management approach can mitigate or exacerbate stakeholder backlash.

Furthermore, the sectoral differences in breach severity call for sector-specific risk assessment and resource allocation. Firms in data-rich industries like healthcare, education, and retail should recognize that their breach impact surface is inherently larger. This means higher potential losses, more legal liability, and greater harm to individuals when incidents occur. These organizations should invest in proportional cybersecurity measures: encryption of large databases, network segmentation to prevent broad access, and robust backup and recovery systems (particularly in education and government, to mitigate ransomware).

Our results showing education and government lagging in response suggest that additional support and oversight may be needed in sectors with limited cybersecurity budgets. Boards and regulators could push for minimum security standards in higher education and local government (similar to how financial institutions comply with FFIEC guidelines or how healthcare complies with HIPAA security rule), to ensure baseline capabilities for breach prevention and response.

### 5.3.2. Implications for Public Policy and Regulation.

Our results challenge the effectiveness of uniform breach notification laws. Given the extended timelines required to investigate insider breaches, overly rigid deadlines may result in premature or incomplete disclosures (Subramanian et al., 2024). A more flexible federal standard—modeled after the GDPR’s 72-hour initial notice with allowance for ongoing updates—could improve timeliness while accommodating investigative realities. Education and local government, where breach delays are most pronounced (Halikias, 2024), may benefit from sector-specific regulatory guidance and oversight, akin to what the Department of Health and Human Services provides for healthcare breaches under HIPAA. Targeted support could include federal funding for cybersecurity upgrades, mandatory breach drills, and participation in threat information-sharing initiatives. High breach variability within baseline sectors also suggests that complacency in less-regulated industries is risky. Even companies outside traditional high-risk sectors can face catastrophic breaches, reinforcing the need for broad regulatory preparedness frameworks (Makridis, 2021).

### 5.3.3 Implications for Theory and Future Research.

Our findings support theories of information asymmetry and deterrence: insider breaches go undetected longer because perpetrators operate within authorized systems (Homoliak et al., 2019). Delays in breach reporting also reflect strategic disclosure choices influenced by anticipated reputational and financial consequences (Confente et al., 2019). Future research could explore why sectors like education lag in response, and how breach dynamics differ across organizational structures. Given the low explanatory power of breach size models, further studies should integrate firm-level characteristics, attacker motivations, and qualitative case data to better understand breach outcomes. Content analysis of disclosure statements (Nikkhah & Grover, 2022) could also reveal how firms’ communication approaches align with breach severity and response timing.

## 6. Conclusion

This study examined how breach origin, organizational sector, and data sensitivity influence the

severity and disclosure timing of data breaches in the U.S. Using a dataset of 19,255 incidents, we find that insider breaches take significantly longer to report—on average nearly 80 days more than external breaches—yet affect fewer individuals. Sectoral differences are substantial: breaches in healthcare, education, and government sectors are consistently larger in scale, highlighting the elevated risk faced by data-rich and often resource-constrained organizations.

These findings emphasize the need for differentiated incident response strategies and sector-specific regulatory attention. While regulation can help improve transparency and accountability, our results suggest that detection complexity and organizational capacity also shape disclosure behavior. Policymakers and practitioners should consider these contextual factors when designing breach notification standards and incident response frameworks. More broadly, the study reinforces that breach outcomes are shaped not only by external threats but also by internal dynamics and sectoral vulnerabilities, calling for targeted approaches to cybersecurity resilience.

## 7. References

- Alawida, M., Omolara, A. E., Abiodun, O. I., & Al-Rajab, M. (2022). A deeper look into cybersecurity issues in the wake of Covid-19: A survey. *Journal of King Saud University-Computer and Information Sciences*, 34(10), 8176-8206.
- Alder, S. (2025). UnitedHealth Adopts Aggressive Approach to Recover Ransomware Attack Loans. *The HIPAA Journal*. Available from: <https://www.hipaajournal.com/change-healthcare-responding-to-cyberattack/>.
- Almeida, D., Shmarko, K., & Lomas, E. (2022). The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. *AI and Ethics*, 2(3), 377-387.
- Avanzi, B., Tan, X., Taylor, G., & Wong, B. (2025). On the evolution of data breach reporting patterns and frequency in the United States: a cross-state analysis. *North American Actuarial Journal*, 1-32.
- Bakker, R. M., & Shepherd, D. A. (2017). Pull the plug or take the plunge: Multiple opportunities and the speed of venturing decisions in the Australian mining industry. *Academy of Management Journal*, 60(1), 130-155.
- Burgess, T. (2022). Understanding breach notification delays. *Barracuda*. Available from: <https://blog.barracuda.com/2022/12/20/breach-notification-delays>.
- Brown, C. S. (2015). Investigating and prosecuting cyber crime: Forensic dependencies and barriers to justice. *International Journal of Cyber Criminology*, 9(1), 55.

- Cheng, L., Liu, F., & Yao, D. (2017). Enterprise data breach: causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(5), e1211.
- Chernyshev, M., Zeadally, S., & Baig, Z. (2019). Healthcare data breaches: Implications for digital forensic readiness. *Journal of medical systems*, 43, 1-12.
- Chertoff, M. (2017). A public policy perspective of the Dark Web. *Journal of Cyber Policy*, 2(1), 26-38.
- Confente, I., Siciliano, G. G., Gaudenzi, B., & Eickhoff, M. (2019). Effects of data breaches from user-generated content: A corporate reputation analysis. *European Management Journal*, 37(4), 492-504.
- Dalei, N. N., & Kandpal, V. (2025). Understanding the Global Landscape of Cybersecurity Risks, Economic Impacts, and Challenges: Lesson for India. In *Cybersecurity, Law, and Economics* (pp. 17-37). Routledge.
- Dearden, T. E., Parti, K., Hawdon, J., Gainey, R., Vandecar-Burdin, T., & Albanese, J. (2023). Differentiating Insider and Outsider Cyberattacks on Businesses. *American Journal of Criminal Justice*, 48(4), 871-886.
- Edwards, B., Hofmeyr, S., & Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, 2(1), 3-14.
- George, A. S., Baskar, T., & Srikanth, P. B. (2024). Cyber threats to critical infrastructure: assessing vulnerabilities across key sectors. *Partners Universal International Innovation Journal*, 2(1), 51-75.
- Gibson, D., & Harfield, C. (2023). Amplifying victim vulnerability: Unanticipated harm and consequence in data breach notification policy. *International Review of Victimology*, 29(3), 341-365.
- Greig, J. (2024). SEC to require financial firms to have data breach incident plans. *The Record*. Retrieved from <https://therecord.media/sec-to-require-data-breach-plans-financial>
- Guha, S., & Kandula, S. (2012, October). Act for affordable data care. In *Proceedings of the 11th ACM workshop on hot topics in networks* (pp. 103-108).
- Gwebu, K. L., Wang, J., & Wang, L. (2018). The role of corporate reputation and crisis response strategies in data breach management. *Journal of management information systems*, 35(2), 683-714.
- Haislip, J., Kolev, K., Pinsker, R., & Steffen, T. (2019, June). The economic cost of cybersecurity breaches: A broad-based analysis. In *Workshop on the economics of information security (WEIS)* (Vol. 1, p. 37).
- Halikias, H. (2024). Business Impacts of Ransomware. In *Digital Shakedown: The Complete Guide to Understanding and Combating Ransomware* (pp. 25-47). Cham: Springer Nature Switzerland.
- Herath, H. M. S. S., Herath, H. M. K. K. M. B., Madhusanka, B. G. D. A., & Guruge, L. G. P. K. (2024). Data protection challenges in the processing of sensitive data. In *Data Protection: The Wake of AI and Machine Learning* (pp. 155-179). Cham: Springer Nature Switzerland.
- Holt, T. J., Griffith, M., Turner, N., Greene-Colozzi, E., Chermak, S., & Freilich, J. D. (2023). Assessing nation-state-sponsored cyberattacks using aspects of Situational Crime Prevention. *Criminology & Public Policy*, 22(4), 825-848.
- Homoliak, I., Toffalini, F., Guarnizo, J., Elovici, Y., & Ochoa, M. (2019). Insight into insiders and it: A survey of insider threat taxonomies, analysis, modeling, and countermeasures. *ACM Computing Surveys (CSUR)*, 52(2), 1-40.
- IBM. (2024). Cost of a Data Breach Report 2024. Available from: <https://www.ibm.com/reports/data-breach>.
- Ira Bedenbaugh, J., Brown, E. (2024). The rising cost of a healthcare data breach. *Elliott davis*. Available from: <https://www.elliottdavis.com/insights/the-rising-cost-of-a-healthcare-data-breach>.
- Isibor, E. (2024). Regulation of Healthcare Data Security: Legal Obligations in A Digital Age. Available at SSRN 4957244.
- Johnson, M., Kang, M. J., & Lawson, T. (2017). Stock price reaction to data breaches. *Journal of Finance Issues*, 16(2), 1-13.
- Kapko, M. (2024). US data comprises surged to record high in 2023. *CYBERSECURITY DIVE*. Available from: <https://www.cybersecuritydive.com/news/data-compromises-surge/705549/>.
- Kuipers, S., & Schönheit, M. (2022). Data breaches and effective crisis communication: a comparative analysis of corporate reputational crises. *Corporate Reputation Review*, 25(3), 176-197.
- Makridis, C. A. (2021). Do data breaches damage reputation? Evidence from 45 companies between 2002 and 2018. *Journal of Cybersecurity*, 7(1).
- Marovic, B., & Curcin, V. (2020). Impact of the European general data protection regulation (GDPR) on health data management in a European Union candidate country: a case study of Serbia. *JMIR Medical informatics*, 8(4), e14604.
- Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. *Journal of marketing*, 81(1), 36-58.
- Merod, A. (2025). Data breach reporting lags in education sector, study finds. *K-12 Dive*. Retrieved from <https://www.k12dive.com/news/data-breach-reporting-lags-in-education-sector-study-finds/748273/>
- Mitra, S., & Ransbotham, S. (2015). Information disclosure and the diffusion of information security attacks. *Information Systems Research*, 26(3), 565-584.
- Mohsin, D. K. (2025). Cybercrime and Privacy in the Digital Age: Legal Frameworks, Emerging Challenges, and Future Trends. *Emerging Challenges, and Future Trends (March 07, 2025)*.
- Molitor, D., Saharia, A., Raghupathi, V., & Raghupathi, W. (2024). Exploring the Characteristics of Data Breaches: A Descriptive Analytic Study. *Journal of Information Security*, 15(2), 168-195.
- Neprash, H. T., Dameff, C., & Tully, J. (2024). Cybersecurity Lessons From the Change Healthcare Attack. *JAMA internal medicine*, 184(11), 1283-1284.
- Nikkhah, H. R., and Grover, V. (2022). "An Empirical Investigation of Company Response to Data Breaches," *MIS Quarterly*, 46(4), 2163-2196. <https://doi.org/10.25300/MISQ/2022/16609>

- Nygaard, L. (2025). This troubling AT&T data breach has put 86 million customers at risk of identity theft. *Yahoo! News*. Available from: <https://www.yahoo.com/news/troubling-t-data-breach-put-184336178.html>.
- Ou, C. X., Zhang, X., Angelopoulos, S., Davison, R. M., & Janse, N. (2022). Security breaches and organization response strategy: Exploring consumers' threat and coping appraisals. *International Journal of Information Management*, 65, 102498.
- Park, S. (2019). Why information security law has been ineffective in addressing security vulnerabilities: Evidence from California data breach notifications and relevant court and government records. *International Review of Law and Economics*, 58, 132-145.
- Pomerleau, P. L., & Lowery, D. L. (2020). Countering cyber threats to financial institutions. *A private and public partnership approach to critical infrastructure protection*.
- Richardson, V. J., Smith, R. E., & Watson, M. W. (2019). Much ado about nothing: The (lack of) economic impact of data privacy breaches. *Journal of Information Systems*, 33(3), 227-265.
- Rosati, P., Deeney, P., Cummins, M., Van der Werff, L., & Lynn, T. (2019). Social media and stock price reaction to data breach announcements: Evidence from US listed companies. *Research in International Business and Finance*, 47, 458-469.
- Sen, R., & Borle, S. (2015). Estimating the contextual risk of data breach: An empirical approach. *Journal of Management Information Systems*, 32(2), 314-341.
- Subramanian, H., Sengupta, A., & Xu, Y. (2024). Patient Health Record Protection Beyond the Health Insurance Portability and Accountability Act: Mixed Methods Study. *Journal of Medical Internet Research*, 26, e59674.
- Syed, R. (2019). Enterprise reputation threats on social media: A case of data breach framing. *The Journal of Strategic Information Systems*, 28(3), 257-274.
- Toledano, S. A. (2024). *Critical Infrastructure Security: Cybersecurity lessons learned from real-world breaches*. Packt Publishing Ltd.
- Wang, J., Gupta, M., & Rao, H. R. (2015). Insider threats in a financial institution. *MIS Quarterly*, 39(1), 91-112.
- Wang, J., Shan, Z., Gupta, M., & Rao, H. R. (2019). A longitudinal study of unauthorized access attempts on information systems. *MIS Quarterly*, 43(2), 601-A8.
- Wang, X., Yan, J., Munyon, T. P., & Crook, T. R. (2025). Breached but not broken: How attributional information shapes shareholder reactions to firms following data breaches. *Corporate Reputation Review*, 28(1), 71-92.
- Wang, Y., Arief, B., & Hernandez-Castro, J. (2025). Secure in the dark? An in-depth analysis of dark web markets security. *International Journal of Information Security*, 24(3), 1-15.
- Wolf, A. (2023). 10 Major Retail Industry Cyber Attacks. *Arctic Wolf*. Available from: <https://arcticwolf.com/resources/blog/10-major-retail-industry-cyber-attacks/>