Original Research Article

DIGITAL HEALTH

DIGITAL HEALTH
Volume 10: 1–11

DOI: 10.1177/20552076231224164
journals.sagepub.com/home/dhj

Sage

Do hospital data breaches affect health information technology investment?

Jinhyung Lee[1], Hyeyeong Kim[2] and Sung J Choi[3]

Abstract

Objectives: Data breaches are a financial and operational threat to hospitals. In this study, we examine the association between a data breach and information technology capital and labor investment.

Methods: In this retrospective cohort study, we used American Hospital Association data from 2017 to 2019 and an unbalanced panel of hospitals with 6751 unique hospital-year observations. The breached group had 482 hospital-years, and the control group had 6269 hospital-years. We estimated the association between data breaches, information technology capital, and labor investment using the average treatment effect with propensity-score matching.

Results: From 2017 to 2019, hospitals experienced more hacking and information technology incidents but fewer thefts and losses. We found that hospital data breaches were associated with a 66% increase in employed information technology staff and a 57% increase in outsourced information technology staff. Breaches were not associated with information technology operating expenses and information technology capital expenses.

Conclusion: Higher information technology labor investment due to the remediation of data breaches is an added cost to the healthcare system. Hospitals and policymakers should consider initiatives to improve cybersecurity and protect patient data.

Keywords

Data breach, information technology investment, privacy, cybersecurity

Submission date: 31 August 2023; Acceptance date: 15 December 2023

Introduction

Data breaches are a devastating threat to the users of information technology. The healthcare sector has seen frequent and costly data breaches relative to other sectors. A health data breach is an impermissible use or disclosure of protected health information.[1] It is broadly classified into theft, loss, unauthorized access/disclosure, improper disposal, and hacking with non-mutually exclusive categories.[1] Healthcare organizations in the US had the highest average costs related to data breaches for the past decade.[2] In 2021, 714 healthcare data breaches, which affected over 40 million individuals' records, were reported by healthcare providers and health plans to the Department of Health and Human Services Office for Civil Rights (OCR).[1,3] For organizations, a healthcare data breach cost $10.1 million on average between 2001 and 2022.[2]

The Health Information Technology for Economic and Clinical Health (HITECH) Act (2009), and specifically the Medicare and Medicaid Electronic Health Record (EHR) Incentive Program[4] promoted the adoption of health information technology. By 2019, virtually all non-federal acute care hospitals adopted EHRs, with 96% of

[1]Department of Economics, Sungkyunkwan University, Seoul, Republic of Korea
[2]Department of Education, Incheon National University, Incheon, Republic of Korea
[3]School of Global of Health Management and Informatics, University of Central Florida, Orlando, FL, USA

Corresponding author:
Sung J Choi, School of Global of Health Management and Informatics, University of Central Florida, 528 West Livingston St, Orlando, FL 32801, USA.
Email: sung.choi@ucf.edu

hospitals reporting any EHR usage and 86% of hospitals reporting 2015 edition certified EHR usage in the Office of the National Coordinator for Health Information Technology (ONC) survey 2019–2021.[5] HITECH also included provisions to improve the privacy and security of personal health records[3] The notifications of healthcare data breaches were strengthened by requiring healthcare providers, health plans, and business associates to report a breach of protected health information that affects more than 500 individuals to those individuals and the Department of Health & Human Services (HHS).[1] The reported data breaches since 2009 are published in a public web database to inform consumers and patients.[6]

Despite efforts from industry and regulators, healthcare organizations continue to experience data breaches. According to the Healthcare Information and Management Systems Society (HIMSS) survey 2021, more than 67% of surveyed organizations had experienced a significant data breach in the preceding 12 months.[7] Specifically, hospitals have been involved in some of the largest data breaches in the past decade, which exposed millions of individual records.[8,9]

Prior studies investigating the relationship between security investments and data breaches explored an organization's motivation for investing in security. Organizations have data and information assets that are critical to their operations. Especially healthcare organizations deal with protected health information, which is valuable to malicious actors.[10] Criminology and deterrence theory suggests that data breaches by malicious actors, assumed to be rational, can be deterred with preventive security investment and penalties for malicious actions.[11–13] This literature extends to the routine activity theory (RAT), which emphasizes the organization's role as a guardian of data and information assets to defend against rational malicious actors.[11]

The RAT provides a framework for understanding a healthcare organization's motivation to guard its assets and avoid regulatory penalties. Healthcare organizations are subject to HIPAA regulations, which enforce corrective actions against breached organizations.[14] The Office for Civil Rights may monitor the breached organization for three years to oversee the implementation of corrective actions, including penalties, and revision of systems, policies, and procedures.

Empirical studies examining the effect of information technology (IT) security investments on data breaches have shown that IT security investment is related to data breaches by mediating and moderating organizational factors.[11,15–17] These studies have focused on the implications of IT security investments on data breaches. However, few studies have examined the effect of data breaches on IT security investments. The reductions in hospital care quality after a data breach suggest it is critical to examine the financial implications of hospital data breaches.[10]

Data breach remediation efforts were negatively associated with hospital quality, that is, with a slower time-to-electrocardiogram rate and with a higher 30-day acute myocardial infarction death rate.[18]

Disruptions in health IT systems caused by a data breach may disrupt or delay the workflow of healthcare providers, therefore decreasing hospital efficiency. Moreover, breached hospitals potentially face investigation, fines, and several years of monitoring by the Office for Civil Rights.[3] Thus, breach remediation (changes to health IT system and staff training) as required by OCR may take some time to complete, and such oversight by OCR may itself degrade hospital efficiency.[19]

The significant disruptions and costs associated with data breaches have implications for hospital IT investments. Whereas health IT is intended to improve hospital quality by reducing errors,[20–22] hospitals need to consider the costs associated with remediating data breaches when making continued IT investments that are critical to the management of hospital services.

We address this gap in the literature by analyzing the relationship between data breaches and hospital IT security investment using a quasi-experimental design. Estimating the effect of data breaches on hospital IT security investments is a novel contribution to the literature. Using the RAT framework, we hypothesize that breached healthcare organizations will make IT security investments to protect patient data, deter malicious actors, and prevent security failures.

Prior studies on hospital data breaches and IT expenditures studied years before 2017 using state-level data when breaches due to hacking and IT incidents were less common.[10,15,16] The recent rise in hacking and IT incidents is a serious concern for hospitals because they tend to be more disruptive.[17,18] This paper contributes to the literature by exploring the implications of data breaches for hospitals using recent nationally representative hospital data, which in turn offers timely insights for hospital managers and policymakers.

Materials and methods

We employed the reported information on breaches as collected by HHS to create a pooled sample of hospital-year observations from 2017 to 2019. In accordance with the ICMJE guidelines, patient consent was not applicable to this study. This study did not involve human subjects, and all data were obtained from publicly available sources. We estimated the changes in IT investment associated with breached hospitals by using the average treatment effect (ATE) model[23] with propensity-score matching (PSM), controlling for hospital characteristics and financial variables.

To evaluate the effect of a policy change on two groups, simple or regression-adjusted comparisons may lead to

biased estimates because of omitted variables arising from unobserved and uncontrolled differences between the two groups. This is a critical concern in estimating the treatment effect. Thus, randomized trials are used to address the omitted variable bias when assessing treatment effects. However, without an experiment or randomized trials, the link between omitted variables bias and treatment effects could be seen using the potential-outcomes framework that could be observed in alternative states. Thus, the potential outcomes framework has become the concept for non-experimental as well as experimental studies in many fields.[24–26]

One commonly used statistical technique using the potential outcomes framework is matching and regression.[27] Matching is similar to regression in that it assumes the only source of omitted variables is the set of observed covariates. However, unlike regression, the treatment effects are constructed by matching subjects with the same covariates instead of using a linear model for the effect of covariates. Moreover, matching has a weaker assumption than regression; the effect of covariates on dependent variables need not be linear.

Thus, the key (conditional independence) assumption becomes

$$E[Y_{ji}|X_i, D_i] = E[Y_{ji}|X_i] \text{ for } j = 0(control) \,/\, 1(treatment) \text{ and hospital } i$$

where Y is the dependent variable, X is a vector of covariates, and D is a dummy variable for treatment ($D=1$) and control ($D=0$) groups.

Then, ATE of data breach (D) on health IT investment (Y) could be calculated by

$$E[Y_{1i}\text{-}Y_{0i}|D_i = 1] = E\{E[Y_{1i}|X_i, D_i = 1]\text{-}[Y_{0i}|X_i, D_i = 1]| D_i = 1\} = E\{E[Y_{1i}|X_i, D_i = 1]\text{-}[Y_{0i}|X_i, D_i = 0]|D_i = 1\}$$

Then

$$E[Y_{1i}\text{-}Y_{0i}] = E\{E[Y_{1i}|X_i, D_i = 1]\text{-}[Y_{0i}|X_i, D_i = 0]\}$$

In sum, ATE can be constructed by averaging covariate-specific treatment-control contrasts, and then rearranging these covariate-specific contrasts using the marginal distribution of covariate. Since these equations involve observable quantity, consistent estimators could be achieved.[27] Therefore, the ATE is generally used to compare treatment and control in the evaluation of policy interventions and medical trials.

However, since the assignment of subjects to the treatment or control groups is not random, the estimation of the treatment effect may be biased because of confounding factors. PSM can reduce the estimation bias of treatment effects controlling for the confounding factors, by using treated and control subjects who are as similar as possible.[28] The estimated equation is followed by:

$$y_{it} = \alpha_i + \theta\, Breach_{it} + \gamma_1 HC_{it} + \gamma_2 HF_{it} + yr + \epsilon_{it} \quad (1)$$

For a hospital (i) at a given year (t), the dependent variable y is IT investment measured as IT operating expense, IT capital expense, number of employed IT staff, and number of outsourced IT staff. Specifically, these variables were log-transformed to address modeling issues with a skewed dependent variable.[29] *Breach* is a dummy indicator that represents a data breach. θ is the coefficient that captures the relationship between a data breach and IT investment.

AHA provided a rich set of variables to control for potential confounding variables. *HC* represents hospital characteristics, such as teaching status, case-mix index (CMI), disproportionate share hospital (DSH) payments, wage index, Census Bureau Division (Metro, Micro, and Rural hospitals), and the number of licensed beds. *HF* represents hospital financial variables, such as total gross patient revenue, total expense, and total assets. γ is the vector of coefficients on the control variables. *yr* is the year-fixed effect. Standard errors accounted for clustering at the hospital level. Analyses were performed using Stata Statistical Software 17.[30]

The relationship between data breaches and IT investment may be confounded by hospital characteristics. For example, large teaching hospitals are more likely to experience data breaches because they have more staff who are handling devices and data.[31] PSM adjusted for potential sample selection bias due to observable differences between the breached and control hospitals.[32–34] Also, financial, legal, and reputational costs associated with data breaches should deter hospitals from making intentional data breaches,[2,3] which mitigates concerns for endogeneity.

Data

The AHA Annual Survey profiles more than 6500 member hospitals. The AHA database has been widely used by government agencies, decision-makers, media, and industry for timely analysis, and it is widely cited in the literature on health IT.[35] The AHA database provides hospital characteristics data, including location, size, structure, and personnel. In addition, the AHA data provides financial and health IT data from income statements and balance sheets, including revenue, assets, liability, health IT investment, and labor expenses.

The HHS breach data provided information on breached hospitals, breach date, and breach type, including theft, unauthorized access/disclosure, hacking and IT incident, improper disposal, and loss.[6] The HHS breach data only

provides data on large breaches affecting more than 500 individuals. This data limitation is further discussed below.

In accordance with the ICMJE guidelines, patient consent was not applicable to this study. This study did not involve human subjects, and all data were obtained from publicly available sources.

This study merged the HHS breach data and AHA survey data using the hospital name, location, and year. Our study sample pooled together short-term acute general hospitals from 2017 to 2019, which were the latest data available to the public that included IT-related expenditures such as IT operating expense, IT capital expense, number of employed IT staff, and number of outsourced IT staff.

Our analysis primarily focused on community hospitals, which we defined according to the AHA.[35] In the United States, the vast majority of hospitals, accounting for over 80%, are categorized as community hospitals.[35] This focus is particularly pertinent to patients and policymakers. Conversely, hospitals falling into categories such as psychiatric, rehabilitation, and government hospitals exhibit distinctive operating characteristics, financial reporting procedures, and financing mechanisms, setting them apart from community hospitals. By concentrating our efforts on community hospitals, we aimed to maintain a consistent framework for our financial analysis. Furthermore, for sample consistency, hospitals with less than a year of reporting period were excluded. The study sample included 6751 hospital-year observations, with 6269 hospitals that were not breached and 482 hospitals that were breached. While the sample size was small enough to allow for manual matching, there may be false matches because standard identifiers were not available. Finally, all financial variables were trimmed at the top 1% to exclude outliers.

Dependent variables

The dependent variable is measured as four variables: IT operating expense, IT capital expense, number of employed IT staff, and number of outsourced IT staff. IT operating expenses include expenses related to information technology and cybersecurity, excluding department depreciation and operating dollars paid against capital leases. IT capital expense includes expenses related to the capital budget of IT and cybersecurity for the current year. The number of employed IT staff is the number of full-time equivalent (FTE) staff employed in the IT department/organization and on the hospital payroll. The number of outsourced IT staff is for contracted staff related to IT.

Independent variable

Breach measured whether a hospital experienced a data breach in a given year. The model assumed that a breach was a one-time event. Even though multiple breaches within a year are possible, we observed no hospitals that experienced multiple breaches in our sample.

Control variables

Teaching status is indicated if the hospital is a Member of the Council of Teaching Hospitals (COTH) of the Association of American Medical Colleges (AAMC).[36] CMI reflects the diversity, complexity, and severity of patient illnesses treated at a given hospital. A higher CMI indicates a more severe and resource-intense patient mix.[37] DSHs serve a significantly disproportionate number of low-income patients and receive payments from the Centers for Medicaid and Medicare Services to cover the costs of providing care to uninsured patients.[38] The wage index reflects the hospital wage level in the geographic area of the hospital relative to the national average hospital wage level. Census Bureau Division reflects the location of Metro, Micro, and Rural hospitals. The number of licensed beds is the number authorized by a state licensing (certifying) agency. Total expense sum is the expenses incurred during the ordinary course of operating the hospitals. Total gross patient revenue sums total gross inpatient and total gross outpatient revenue. Finally, total assets are the sum of current, fixed, and other assets.

Results

Descriptive statistics

Figure 1 plots the trends in hospital data breaches by type from 2017 to 2019. There were 53 breaches in 2017, 41 breaches in 2018, and 61 breaches in 2019, showing an increasing trend overall. Unauthorized access was the most common type of breach for all years, with 28 breaches in 2017, 21 breaches in 2018, and 30 in 2019. Notably, hacking and IT incidents increased from 13 breaches in 2017 and 2018 to 27 breaches in 2019, doubling in

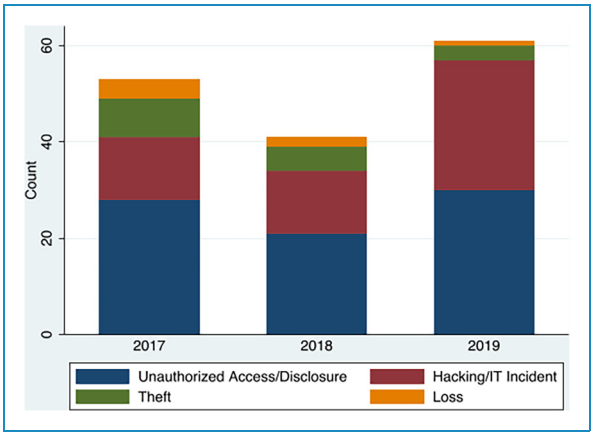

Figure 1. Trends of hospital data breaches by type 2017–2019.

number from 2017 to 2019. Theft and loss were less common among the breaches. Thefts were responsible for eight breaches in 2017, five breaches in 2018, and three breaches in 2019. Losses were responsible for four breaches in 2017, two breaches in 2018, and one breach in 2019. In 2017, theft and loss combined were 12 breaches, but they were reduced by a third to four in 2019. Thus, despite the reduction in theft and losses, the growth in hacking and IT incidents drove the overall increase in data breaches.

Table 1 provides the descriptive statistics comparing the breached hospitals and not-breached hospitals. The pooled hospital-year observations were classified by breach status into the breached group or control (not-breached) group. The full sample size was 6751 unique hospital-year observations. The breached group included 482 hospitals, and the control group included 6269 hospitals. The breached group had more than twice the IT operating expense ($34.54 million vs. $14.42 million) and more than twice the IT capital expense ($10.18 million vs. $4.82 million) than the control group. The breached group hospitals had more than twice the number of employed IT staff (147.84 vs. 67.01) and almost twice the number of outsourced IT staff (15.45 vs. 8.44) than the control group. The breached group was likely to have more beds (383.8 vs. 184), a higher CMI (1.75 vs. 1.59), and a higher wage index (1.00 vs. 0.97). The breached group had almost three

Table 1. Descriptive summary of breached and control hospitals between 2017 and 2019 (6751 hospital years). All the variables except ownership were statistically different between no breach group and breach group.

		No breach group ($n = 6269$)		Breach group ($n = 482$)		
Variable		Mean	Std. dev.	Mean	Std. dev.	p-value
IT operating expense ($1000)		14,421.17	44,776.04	34,538.31	56,497.2	<0.01
IT capital expense ($1000)		4815.37	17,188.02	10,181.76	21,928.81	<0.01
Number of employed IT staff		67.01	856.76	147.84	220.69	<0.01
Number of outsourced IT staff		8.44	60.96	15.45	39.08	<0.01
Number of beds		184	224	383	310	<0.01
Case-mix index		1.585	0.300	1.748	0.318	<0.01
Wage index		0.974	0.180	1.002	0.153	<0.01
Total gross patient revenue ($ million)		863.8	1,436,913	2344.7	2,917,547	<0.01
Total expense ($ million)		237.9	385,509	697.3	895,708	<0.01
Total assets ($ million)		372.5	1,215,501	963.7	1,543,454	<0.01
Metro (%)	Metro	58.24		83.2		<0.01
	Micro	18.57		13.28		
	Rural	23.19		3.53		
Ownership (%)	Government	25		17.01		0.203
	Not-for-profit	64.6		78.4		
	Investor-owned	10.4		4.6		
Teaching (%)		6.57		26.56		<0.01
DSH payment		6.25		84.2		<0.01

DSH: disproportionate share hospital; IT: information technology.

times the total gross patient revenue ($2.34 billion vs. $863.80 million), almost twice the total expense ($697.30 million vs. $237.90 million), and more than twice the total assets ($963.7 million vs. $372.5 million) that the control group had. The breached group was more likely to be located in a metropolitan area (83.20% vs. 58.24%). The percentage of hospitals in a micropolitan area was 13.28% for the breached group, and 18.57% for the control group. The percentage of hospitals in a rural area was 3.53% for the breached group, and 23.19% for the control group. The breached group was more likely to be a not-for-profit hospital (78.40% vs. 64.60%). The percentage of government hospitals was lower among the breached group (17.01% vs. 25%), and the percentage of investor-owned hospitals was lower among the breached group (4.6% vs. 10.4%). The breached group was more likely to be a teaching hospital (26.56% vs. 6.57%) and more likely to be DSH payment recipients (84.2% vs. 6.25%). All variables except ownership were statistically different between no breach group and breach group.

Regression results

Standardized mean difference and variance ratios were adopted to evaluate the validity of the matches. A standardized mean difference of less than 0.1 for each covariate indicates negligible imbalance in the matches. Moreover, variance ratios close to one also suggested a similar distribution of respective covariates. Table 2 shows that the before-match column had large differences over the variables. However, differences over the variables in the after-matched column are negligible, and variance ratios are all near one, suggesting that appropriate matches have been made.

After matching by using the propensity-score matched sample, we regressed data breach on IT investment (IT operating expense, IT capital expense, number of employed IT staff, and number of outsourced IT staff), controlling for hospital characteristics (ownership, teaching status, CMI, DSH payment status, wage index, urban status, and number of licensed beds), and hospital financial variables (total gross patient revenue, total expense, and total assets).

Table 3 shows the regression coefficient estimates of the ATE. Data breaches were associated with a statistically significant increase in staffing. Since the dependent variables were log-transformed to address skewness, the coefficients were exponentiated to be interpreted as multiplicative changes.

The breached group was associated with a 66.3% ($\exp(0.509) = 1.663$, $p < 0.001$) increase in the number of employed IT staff, all other things being equal. Also, the breached group was associated with a 56.8% ($\exp(0.450) = 1.568$, $p < 0.001$) increase in the number of outsourced IT staff, all other things being equal. However, the breached group was not significantly associated with either IT operating expenses ($p = 0.242$) or IT capital expenses ($p = 0.563$).

Table 2. Standardized means differences and variance ratios of the before-matched and after-matched samples.

	Standardized differences		Variance ratio	
Variables	Before matched	After matched	Before matched	After matched
Number of beds	0.6014	−0.0572	0.9689	1.0912
Case-mix index	0.4713	−0.0535	1.1403	0.9151
Wage index	0.1673	0.0065	0.8021	0.7717
Total gross patient revenue	0.5956	−0.0571	0.9897	1.0831
Total expense	0.6717	−0.0407	1.1404	1.1590
Total assets	0.4909	−0.0318	1.5004	1.0287
Metro	−0.3828	−0.0736	0.3708	0.6556
Ownership	−0.2341	−0.0566	0.7550	0.9284
Teaching	−0.4977	0.0118	2.5265	0.9690
DSH payment	0.0248	−0.0610	0.9466	1.1395

DSH: disproportionate share hospital.

Table 3. ATE with PSM.[a]

ATE	Breach group	Coefficient	Std.Err.	$p > z$	[95% confidence interval]	
IT operating expense	(breach vs. no breach)	−0.124	0.106	0.242	−0.332	0.084
IT capital expense	(breach vs. no breach)	−0.069	0.120	0.563	−0.306	0.166
Number of employed IT staff	(breach vs. no breach)	0.509	0.114	0.000	0.287	0.732
Number of outsourced IT staff	(breach vs. no breach)	0.450	0.114	0.000	0.227	0.673

ATE: average treatment effect; IT: information technology; PSM: propensity-score matching.
aAll ATE estimation was controlled by teaching status, case-mix index, dish payment, wage index, Census Bureau Division, number of licensed beds, total gross patient revenue, total expense, and total assets.

Discussion

Our study examined the effects of data breaches on IT security investments. We hypothesized that the breached healthcare organizations would make IT security investments to better guard their data. Our results advance the theoretical understanding of the effects of data breaches on hospital investments.

There was an increase in hospital data breaches from 2017 to 2019. The increase in data breaches was driven by the growth in hacking and IT incidents, which more than offset the reduction in theft and losses.

Reduction in theft and losses is a welcome change as hospitals are improving the physical security of their devices. However, the surge of hacking and IT incidents, which coincides with the emergence of hospital ransomware attacks, which is a more severe form of hacking, in 2016[39,40] is a serious concern to hospitals. Hospitals may be reluctant to share details regarding hacking or IT incidents because of legal liabilities and ongoing legal cases. Thus, few details are available publicly for researchers and patients. The attack on Hancock Regional Hospital is a rare example in which the organization shared its experience to inform the public.[41] Hancock Regional's experience highlights the challenges of hospitals in remediating and recovering from an attack.

In our descriptive analysis, the breached group was significantly larger in size, measured in terms of beds, IT expenses, and assets. Larger organizations have more users and devices that form an attack surface. Hence, these organizations are more vulnerable to data breaches. The significant differences in hospital and financial characteristics between the breached group and the control group indicated that propensity-score matching was the appropriate strategy to adjust for observable differences.

The empirical model for estimating the ATE of data breaches on IT security investments found that breached hospitals were associated with higher IT labor investments, which demonstrates that breached hospitals take remedial action to repair the damages after a breach. Regression results showed that data breaches were associated with statistically significant increases in the number of employed IT staff and the number of outsourced IT staff. Breached hospitals may remedy IT disruptions by hiring more IT staff.

In the context of existing knowledge, this short-run association between breaches and IT labor investment in 2017–2019 is in contrast with a previous study that only found a long-run association using 2012–2016 data.[42] Data breaches are a shock to hospitals; hence, the response efforts to a breach may require IT staffing that is above what is required for normal operations. Health IT labor shortages may have been a barrier in the past.[43–45] However, the findings from this study suggest that the IT labor market may have expanded in recent years to meet demand shocks from breaches.

Setting our study apart from prior research (Lee, Choi 2021), the dataset used in our investigation presents notable differences. Firstly, we utilized a broader dataset encompassing 6751 hospital years of American Hospital Association (AHA) data, in contrast to earlier studies which primarily focused on California hospital data, limited to 2610 hospital years spanning the years 2012 to 2016. Furthermore, our dataset featured more recent years, ranging from 2017 to 2019, compared to the relatively earlier time frame used in previous studies. Our analytical approach also differed from prior research. We adopted a methodological framework incorporating ATE and PSM to mitigate the influence of unobserved variables. In contrast, the prior study relied on the Difference-in-Difference (DID) design to analyze the data.

Our findings provide practical application for hospital managers and policymakers seeking to improve IT security. Given the growth in hacking and IT incidents, healthcare organizations should seriously consider investing in IT security. The short-run response to a data breach requires higher IT labor investments. The Cybersecurity Act (CSA) of 2015 introduced the 405(d) program to enhance cybersecurity in the healthcare and public health sectors.

The program offers guidance on cybersecurity best practices for awareness and mitigation. Cybersecurity practices include developing an incident response plan, personnel training, and the development and implementation of protection processes and technology.[46] A comprehensive incident response plan is crucial for quickly and effectively addressing security threats and breaches.

It is worth noting that healthcare organizations have traditionally taken a reactive approach to cybersecurity in that cybersecurity investments are made in response to breaches or regulatory mandates. Hospitals have to consider political or regulatory decisions as well as economic decisions when investing in security investments.[47–50] Lack of leadership or resources often leaves organizations to take a reactive approach to developing cybersecurity.[51]

Healthcare organizations should work towards strengthening their proactive cybersecurity strategies. Proactive investments aim to prevent breaches and attacks. Proactive investments are found to be more cost-effective in healthcare security than reactive investments.[16] Compared to other sectors, healthcare has been lagging in cybersecurity ratings.[52] Given the increasing frequency and sophistication of breaches and cyberattacks in the healthcare sector, proactive approaches are essential for a comprehensive strategy.

There is heterogeneity in the type of breach, as hacking can be classified as a breach by external agents, whereas unauthorized access is a breach by internal agents.[53] Such variation in breaches may require varying levels of effort for remediation. Hacking and IT incidents are less likely to be intentional acts by agents inside hospitals. The higher proportion of hacking and IT incidents among breaches mitigates concern for endogeneity problems in the model estimates.

However, data breaches were not associated with IT operating and IT capital expenses. Breached hospitals may not make short-term changes in IT capital expenditure in response to a breach because capital expenditures require long-term planning, are costly to secure financing, and are difficult to reverse. Moreover, IT implementation costs, such as system and software design, coding, installation to hardware, and testing, are typically funded as a capital expense rather than an operating expense. That is, the costs can be spread over several years by capitalizing on the IT system.[19] Thus, we may not observe an increase in IT capital spending in the short term. Operating expenses in health IT may include leasing or renting of technology, such as cloud computing services. Switching to cloud computing may alleviate the burden of cybersecurity from hospitals to the cloud provider. However, transitioning to cloud computing is resource and time-intensive. The lack of association between breaches and IT operating expense and IT capital expense suggests that hospitals are not making short-run changes to physical hospital IT equipment after a breach.

The AHA data provided a rich timely dataset on health IT capital and labor. However, our study was limited to analyzing data breaches that affected more than 500 individuals published by HHS. This is likely an undercount of actual data breaches. However, severe data breaches, especially those by hacking or IT incidents, tend to affect more than 500 individuals. Small breaches with fewer than 500 affected individuals are less likely to cause a significant impact on hospitals and thus are not published by HHS.[54]

Additionally, we conducted a comparison between the treatment group (those who experienced a breach) and the control group (those who did not experience a breach), but we did not assess the data before and after the breach occurred. This limitation could introduce potential bias into the analysis. We recognize that assessing the temporal variation would have provided a more comprehensive perspective.

Conclusions

In this study, we examined the association between data breaches, hospital IT capital, and labor investment. The composition of hospital data breaches changed from 2017 to 2019, as hospitals saw more hacking and IT incidents but fewer theft and losses. Breached hospitals were associated with more resources allocated to IT labor investments. Hospital IT operating expenses and capital expenses were not sensitive to breaches. Higher IT labor investment due to the remediation of data breaches is an added burden to the healthcare system. Therefore, hospital managers and policymakers should consider initiatives to improve cybersecurity and protect patient data.

Our study has several limitations. First, the landscape of hospital data breaches and cybersecurity is changing rapidly with emerging threats. Our study provides an analysis of timely data available to researchers, but the pace of hospital survey data release lags behind current events in cybersecurity. Thus, future research could build on this study by incorporating more years of data, which could lead to more accurate forecasting and estimation. Moreover, other statistical techniques beyond regression analysis, like machine learning algorithms, could be employed to provide further insights into the phenomenon being studied. Second, our analysis of US healthcare data may not generalize globally due to differences in healthcare systems and data availability in other countries. Therefore, it may be necessary to conduct separate studies in other countries to better understand the nature of data breaches in their respective healthcare systems. Moreover, it may be important to consider how differences in healthcare policies and regulations could impact the frequency and severity of data breaches in other countries.

Contributorship: Theoretical model design, Jinhyung Lee; data collection, Sung Choi; validation, Jinhyung Lee, Hey Yeong

Kim, and Sung Choi; methodology, Jinhyung Lee and Sung Choi; software, Jinhyung Lee; writing–original draft, Jinhyung Lee and Hey Yeong Kim; writing–review and editing, Hey Yeong Kim and Sung Choi. All authors have reviewed and agreed to the final version of the manuscript.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: No human or animal participants were involved and no patient data was collected. No human-identifiable data is reported in this study.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5A2A03061909).

Guarantor: SC.

ORCID iD: Sung J Choi https://orcid.org/0000-0003-0299-5382

Patient consent: In accordance with the ICMJE guidelines, we confirm that patient consent was not applicable to this study. This study did not involve human subjects, and all data were obtained from publicly available sources.

References

1. Office for Civil Rights. Breach notification rule [Internet]. HHS.gov. 2009. Available at: https://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html.
2. IBM. Cost of a data breach 2022. 2022 [cited 2022 November 7]; Available at: https://www.ibm.com/reports/data-breach.
3. Office for Civil Rights (OCR). HIPAA compliance and enforcement [Internet]. HHS.gov. 2008. Available at: https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/index.html.
4. Centers for Medicare & Medicaid Services. Promoting interoperability programs | CMS [Internet]. Available at: https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms (2023, accessed 18 January 2023).
5. Adoption of electronic health records by hospital service type 2019–2021 | HealthIT.gov [Internet]. Available at: https://www.healthit.gov/data/quickstats/adoption-electronic-health-records-hospital-service-type-2019-2021 (accessed 10 January 2023).
6. U.S. Department of Health, for Civil Rights HSO. Breach portal: notice to the secretary of HHS breach of unsecured protected health Information [Internet]. Available at: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf
7. HIMSS (Healthcare Information, Society) MS. 2021 HIMSS healthcare cybersecurity survey report [Internet]. Available at: https://www.himss.org/resources/2021-himss-healthcare-cybersecurity-survey-report.
8. CHS settles patient data breach for $5M [Internet]. Available at: https://www.beckershospitalreview.com/cybersecurity/chs-settles-patient-data-breach-for-5m.html (accessed 7 November 2022).
9. UCLA Health reports cyberattack affecting 4.5M [Internet]. Available at: https://www.beckershospitalreview.com/healthcare-information-technology/ucla-health-reports-breach-affecting-4-5m.html (accessed 7 November 2022).
10. Koppel R and Kuziemsky C. Healthcare data are remarkably vulnerable to hacking: connected healthcare delivery increases the risks. *Stud Health Technol Inform* [Internet] 2019 Jan 1; 257: 218–222. Available at: https://europepmc.org/article/MED/30741199 (accessed 15 March 2023).
11. Li H, Yoo S and Kettinger WJ. The roles of IT strategies and security investments in reducing organizational security breaches. *J Manage Inf Syst* 2021; 38: 222–245.
12. D'Arcy J, Hovav A and Galletta D. User awareness of security countermeasures and its impact on information systems misuse: a deterrence approach. *Inf Syst Res* [Internet]. 2009 Mar 1; 20: 79–98. Available at: https://pubsonline.informs.org/doi/abs/10.1287/isre.1070.0160 (accessed 15 March 2023).
13. Wang J, Gupta M and Rao HR. Insider threats in a financial institution: analysis of attack-proneness of information systems applications. *MIS Q* 2015 Mar 1; 39: 91–112.
14. Office for Civil Rights. Enforcement process [Internet]. Available at: https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/enforcement-process/index.html (2021, accessed 28 February 2023).
15. Kwon J and Eric Johnson M. Health-care security strategies for data protection and regulatory compliance. *J Manag Inf Syst* [Internet]. 2014 Oct 1; 30: 41–66. Available at: https://www.tandfonline.com/doi/abs/10.2753/MIS0742-1222300202 (accessed 15 March 2023).
16. Kwon J and Johnson ME. Proactive versus reactive security investments in the healthcare sector. *MIS Q* 2014 Jun 1; 38: 451–471.
17. Angst CM, Block ES, D'Arcy J, et al. When do it security investments matter? Accounting for the influence of institutional factors in the context of healthcare data breaches. *MIS Q*. [Internet] 2017 Sep 1; 41: 893–916. Available at: https://dl.acm.org/doi/10.25300/MISQ/2017/41.3.10 (accessed 15 March 2023).
18. Choi SJ, Johnson ME and Lehmann CU. Data breach remediation efforts and their implications for hospital quality. *Health Serv Res* [Internet] 2019; 54: 971–980. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-6773.13203.
19. Lee J and Choi SJ. Hospital productivity after data breaches: difference-in-differences analysis. *J Med Internet Res* [Internet] 2021 Jul; 23: e26157. Available at: https://www.jmir.org/2021/7/e26157.
20. Parente ST and McCullough JS. Health information technology and patient safety: evidence from panel data. *Health Aff (Millwood)*. 2017 Aug 2; 28: 357–360.
21. Bates DW and Singh H. Two decades since to err is human: an assessment of progress and emerging priorities in patient safety. *Health Aff*. 2018 Nov 5; 37: 1736–1743.

22. Wachter RM and Howell MD. Resolving the productivity paradox of health information technology: a time for optimism. *JAMA* [Internet] 2018 Jul 3; 320: 25–26. Available at: https://jamanetwork.com/journals/jama/fullarticle/2683126 (accessed 7 November 2022).
23. Wooldridge J. *Econometric analysis of cross section and panel data*. 2nd ed. Cambridge, MA: MIT Press, 2010. Available at: https://www.stata.com/bookstore/econometric-analysis-cross-section-panel-data/ (accessed 10 January 2023).
24. Holland PW. Statistics and causal inference. *J Am Stat Assoc* 1986; 81: 945–960.
25. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* [Internet] 1974 Oct; 66: 688–701. Available at: https://dash.harvard.edu/handle/1/3408692 (accessed 15 March 2023).
26. Rubin DB. Assignment to treatment group on the basis of a covariate. *J Educ Stat* [Internet]. 1977 Mar 1;2:1–26. Available at: https://journals.sagepub.com/doi/10.3102/10769986002001001 (accessed 15 March 2023).
27. Angrist JD. Treatment effect. *Microeconometrics* [Internet] 2010: 329–338. Available at: https://link.springer.com/chapter/10.1057/9780230280816_36 (accessed 19 March 2023).
28. Becker SO and Ichino A. Estimation of average treatment effects based on propensity scores. *Stata J: Promot Commun Stat Stata* [Internet]. 2002 Dec 1; 2: 358–377. Available at: https://journals.sagepub.com/doi/10.1177/1536867X0200200403 (accessed 28 January 2023).
29. Deb P and Norton EC. Modeling health care expenditures and use. *Annu Rev Public Health* [Internet]. 2018 Apr 2; 39: 489–505. Available at: https://www.annualreviews.org/doi/abs/10.1146/annurev-publhealth-040617-013517 (accessed 18 January 2023).
30. StataCorp. Stata: software for statistics and data science [Internet]. 2017. Available at: https://www.stata.com/.
31. Gabriel MH, Noblin A, Rutherford A, et al. Data breach locations, types, and associated characteristics among US hospitals. *Am J Manag Care* [Internet] 2018 Feb 1; 24: 78–84. Available at: https://europepmc.org/article/med/29461854 (accessed 28 November 2022).
32. Rosenbaum PR and Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* [Internet] 1983 Apr 1; 70: 41–55. Available at: https://academic.oup.com/biomet/article/70/1/41/240879 (accessed 29 November 2022).
33. Rosenbaum PR and Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 1985 Feb; 39: 33.
34. Choi SJ and Johnson ME. Understanding the relationship between data breaches and hospital advertising expenditures. *Am J Manag Care* [Internet] 2019 Jan; 25: e14–e20. Available at: https://pubmed.ncbi.nlm.nih.gov/30667613/.
35. American Hospital Association. AHA Annual Survey Database™ | AHA Data [Internet]. Available at: https://www.ahadata.com/aha-annual-survey-database (2022, accessed 28 November 2022).
36. AAMC. Council of Teaching Hospitals and Health Systems (COTH) | AAMC [Internet]. Available at: https://www.aamc.org/career-development/affinity-groups/coth (2023, accessed 19 January 2023).
37. Case Mix Index | HealthData.gov [Internet]. https://healthdata.gov/State/Case-Mix-Index/gi5x-y67j/data (2023, accessed 19 January 2023).
38. Centers for Medicare & Medicaid Services. Case Mix Index | CMS [Internet]. Available at: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Acute-Inpatient-Files-for-Download-Items/CMS022630 (2022, accessed 28 November 2022).
39. Becker's Health IT, Review BH. Hospitals are hit with 88% of all ransomware attacks [Internet]. Becker's Hospital Review. 2016. Available at: https://www.beckershospitalreview.com/healthcare-information-technology/hospitals-are-hit-with-88-of-all-ransomware-attacks.html.
40. Goodin D. Ransomware forces 3 hospitals to turn away all but the most critical patients [Internet]. *Ars Technica* 2019. Available at: https://arstechnica.com/information-technology/2019/10/hamstrung-by-ransomware-10-hospitals-are-turning-away-some-patients/.
41. Garrity M. "We did what a hospital does every day"—How Hancock Regional responded to a ransomware attack [Internet]. Becker's Hospital Review June, 2019. Available at: https://www.beckershospitalreview.com/cybersecurity/we-did-what-a-hospital-does-every-day-how-hancock-regional-responded-to-a-ransomware-attack.html.
42. Choi SJ, Johnson ME and Lee J. An event study of data breaches and hospital IT spending. *Health Policy Technol* 2020; 9. doi:10.1016/j.hlpt.2020.04.008
43. Terry K. IT talent shortage hitting healthcare hardest | CIO [Internet]. CIO. Available at: https://www.cio.com/article/244363/it-talent-shortage-hitting-healthcare-hardest.html (2015, accessed 15 March 2023).
44. HIMSS. 2013 HIMSS workforce survey: trends and barriers [Internet]. Available at: https://www.himss.org/2013-himss-workforce-survey-trends-and-barriers (accessed 15 March 2022).
45. Perna G. Report: health IT facing significant worker shortage | Healthcare Innovation [Internet]. Available at: https://www.hcinnovationgroup.com/finance-revenue-cycle/health-it-market/news/13020950/report-health-it-facing-significant-worker-shortage (2013, accessed 15 March 2023).
46. Office of the Assistant Secretary for Preparedness HHS. Health industry cybersecurity practices: managing threats and protecting patients [Internet]. Available at: https://www.phe.gov/Preparedness/planning/405d/Documents/HICP-Main-508.pdf (2017, accessed 9 February 2023).
47. Anderson R. Why information security is hard – an economic perspective. Proceedings – annual computer security applications conference. ACSAC 2001, 2001-January, pp. 358–365.
48. Bhuyan SS, Kabir UY, Escareno JM, et al. Transforming healthcare cybersecurity from reactive to proactive: current status and future recommendations. *J Med Syst* [Internet] 2020 May 1; 44: 1–9. Available from: https://link.springer.com/article/10.1007/s10916-019-1507-y (accessed 24 October 2023).
49. Bodin LD, Gordon LA and Loeb MP. Evaluating information security investments using the analytic hierarchy process.

Commun ACM [Internet] 2005 Feb [cited 2023 Oct 24]; 48: 78–83.
50. Rowe BR and Gallaher M. Private sector cyber security investment: an empirical analysis. Workshop on the economics of information security 2006.
51. Jalali MS and Kaiser JP. Cybersecurity in hospitals: a systematic, organizational perspective. *J Med Internet Res* [Internet] 2018 May 1; 20: e10059. Available at: https://www.jmir.org/2018/5/e10059 (accessed 24 October 2023).
52. Choi SJ and Johnson ME. The relationship between cybersecurity ratings and the risk of hospital data breaches. *J Am Med Inform Assoc* 2021; 28: 2085–2092.
53. Kwon J and Johnson ME. Meaningful healthcare security. *MIS Q*. [Internet] 2018 Dec 1 [cited 2022 Nov 27]; 42: 1043–1067. Available at: https://dl.acm.org/doi/10.25300/MISQ/2018/13580.
54. Heubusch K. Little breaches: OCR releases first "small breach" data. *J AHIMA* [Internet] 2011 Oct; 82: 56–57. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22029215.