Chitra Marti
May 30th, 2025

Competition and Cybercrime: A Theoretical Assessment and Empirical Application

Abstract

This paper examines how the structure of digital technology markets influences cybersecurity outcomes, focusing on how market concentration in critical sectors shapes the security landscape. Cyberattacks pose a unique supply risk, driven by attackers' strategic behavior and the scalability of attacks when shared technology products have common vulnerabilities.

First, I present the outcome of a theoretical model comparing monopoly (full concentration) with competitive (full decentralization) market structures, identifying four externalities that shape the relationship between market structure and cybersecurity outcomes. Two forces in concentrated markets can improve cybersecurity, while two others can worsen it. I evaluate the effectiveness of various cybersecurity policies within this model, under certain conditions – particularly when security costs are high for users and firms – antitrust policy may be the most effective solution for creating a stable and secure digital economy.

Then, I evaluate my theoretical findings in the empirical context of U.S. healthcare digitization in the twenty-first century. I explore how regulation and the nature of digital technology have led to highly concentrated markets, leaving users with limited choices. Hospitals, driven by regulation and positive network effects, often prioritize product compatibility over security, reducing the incentive for technology vendors to focus on designing secure products. The resulting regulatory and market environment has left hospitals vulnerable to large-scale cyberattacks, as they often rely on the same insecure technologies.

I investigate the trend of hospital-level cyberattacks since 2010 and shows that digitization has unilaterally worsened cybersecurity outcomes. Using a revealed-security method, I find heterogeneity in technology product and hospital risk. With the theoretical model, I find the identified negative externalities are evident in the healthcare technology market, while positive externalities are weak or absent. A negative network effect occurs when hospitals use the same technology as those that have already been breached, making everyone more vulnerable to scaled attacks. Using the model, I find that overall, concentration has had a negative effect on cybersecurity outcomes, and that antitrust policy may be the key to preventing future cyberattacks in the U.S. healthcare sector.

1 Introduction

Do you remember the 2021 holiday season? As the world recovered from a global pandemic, only to be hit once more with variants that waved into new lockdowns, those of us here in New York had to deal with a unique extra crisis: The Great Christmas Cream Cheese Catastrophe of December 2021.

With no cream cheese to be found, panicked bagel shop and bakery owners called any and every supplier in the tri-state area; customers debated whether tofu-based vegan cream "cheese" was worth trying; and bakeries wondered what could possibly replace a holiday cheesecake (*New York Times* (2021)).

"I was like, 'What am I going to do this weekend?" the owner of Tompkins Square Bagels, Christopher Pugliese – close to New York University's Stern School of Business, where I have

been a PhD student since 2019 – told the *Times*. "Four people just told me they can't get me cream cheese."

'Sunday bagels are sacred," Mr. Pugliese said. "I hate feeling like I've let people down."

The *Times* described cream cheese as another casualty of "supply chain issues," that ever-present post-pandemic problem. But Christmas cream cheese might have actually been derailed by another newly prevalent problem: a cyberattack.

In mid-October – about a month before the *Times* article, and around the time I first began questioning how market concentration might be linked with poor cybersecurity – Schreiber Foods, the second-largest producer of cream cheese, was hit by a cyberattack that halted production completely at the company's plants and distribution centers for at least five days (*Wisconsin State Farmer* (2021)). Reports of a ransom of over one million dollars were not officially confirmed by the company. Because dairy is perishable, production was necessarily cut – and, a few weeks later, the full effect was felt at the end of the supply chain: by New York City bagel shops (*Bloomberg* (2021)).

The Great Christmas Cream Cheese Catastrophe of 2021 in New York City is perhaps the only lighthearted recent example of a cyberattack. Everywhere else, more dire examples abound.

In February 2024, a ransomware attack hit Change Healthcare, which operates the largest clearinghouse for medical payments (Rundle and Stupp (2024)). The result was that almost *no insurance claims* were processed, leaving literally millions of patients, providers, and pharmacists in the lurch. Payments systems were taken offline, leaving providers to either take out loans or shut down operations entirely as the flow of income was stopped.

Not coincidentally, Change is owned by the largest health insurer, UnitedHealth Group, after it was rolled up into the company via vertical merger in February 2022 – exactly two years before the crippling attack. That merger was challenged on the grounds that the data UnitedHealth handled via Change would naturally come from its competitors and give the insurer an unfair advantage, but the challenge was dismissed (Bartz (2022)).

The hacking group behind the attack allegedly received a $22 million ransomware payment (Greenberg (2024)) – one of the largest ever made. UnitedHealth confirmed a "substantial portion" of the American population had their sensitive personal health information – not to mention identity and insurance information – breached in the attack. An American Hospital Associated (AHA) survey found *ninety-four percent* of AHA hospitals reported a financial impact from that single attack, with more than 33% reporting that half their revenue was impacted (American Hospital Association (2024)). The Change attack, without exaggeration, has affected *five percent of U.S. GDP* (Goldsmith (2024)). Hackers made off with millions, UnitedHealth is still the market leader, but nursing homes and small practices are devastated.

How did we get here? The Change attack was large but neither an exception nor the largest ever: since 2010 hackers have conducted at least five thousand unique attacks and stolen at least 200 million sensitive records from just the U.S. healthcare system, typically for resale and follow-up fraud. Over the last decade, healthcare providers have embraced digital technologies and cybercriminals have embraced the sensitive health and financial information now at risk.

In the Change healthcare attack, the attack compromised credentials of a high-level administrator through social engineering. The compromised application did not have multifactor authentication (MFA), considered one of the most basic security provisions (Rundle (2024)). Two fundamentally economic questions arise: why didn't Change install MFA? And why was that one mistake so devastating for the *entire U.S. healthcare system*?[1]

[1]The Wall Street Journal later covered the issue of concentrated markets in the article "Healthcare Sector Maps

The reliance on digital technology, the extremely concentrated market with few alternative options, a fragile and brittle supply chain with no room for error, the systems that malicious attackers find both vulnerable and vulnerable – these seem to be basic facts of the economy in 2024. Digital technologies have unlocked a new world economy, and even more so in healthcare; to try to capture the effect of digitization in a few sentences is futile. However, what digital technology provides, cybercrime threatens to sabotage. How do we design secure, stable systems?

In this paper, I investigate, both theoretically and empirically, how market forces shape cybersecurity outcomes for firms and users of digital technologies. Cyberattacks are a unique type of risk in two specific ways. First, attacks are not random but rather *strategic* choices of malicious attackers, who gather information about their targets and make choices to maximize their payoffs – usually, to maximize destruction. Second, just like the digital technologies they target, cyberattacks show *scale*: the same attack can be deployed repeatedly against multiple targets, as long as the targets have the same vulnerability. Both factors suggest that the structure of the market for technology products that facilitate cyberattacks – both in terms of market concentration on just a few firms and in terms of networks – is crucial in determining the final security of the overall industry.

The study of the economics of cybersecurity rests naturally at the intersection of the literature on the *economics of digitization* (low costs of replication and scale, the death of distance) and the *economics of crime* (strategic criminal agents weighing costs and benefits). This paper examines an *unexpected effect of market concentration* in the technology sector: the provision of cybersecurity and the ultimate outcomes of data breaches on consumers. I also contribute to the literature on *the use and protection of data within healthcare*.

In this paper, I first present a theoretical model illustrating four externalities that arise from the relationship between market competition and cybercrime outcomes. Using the basic facts of technology and cyberattacks, I describe two positive externalities that lead concentration to *improve* and two negative externalities that *worsen* the impact of cybercrime on users. I then show, through the model, how various commonly proposed cybersecurity policies would alter final outcomes, evaluating each for efficiency and efficacy. I show that, if the cost of security to users and firms is high enough, antitrust policy – i.e. moving to a decentralized system – may be the most desirable path to a stable and secure economy.

I next apply the findings of the model to a key industry experiencing significant cyberattacks: the U.S. healthcare sector. In the U.S., healthcare spending accounts for about one-sixth of GDP. It is also increasingly the first stop for cybercriminals seeking to breach significant amounts of important data. I study the role the digitization of healthcare has played in ensuing cyberattacks, and in particular investigate how the concentrated market for Electronic Medical Records (EMR) software has affected outcomes. I find that not only are some vendors simply worse than others, but also that there exists a *negative network effect* at the hospital-level of using the same technology as another breached hospital, precisely because scaled attacks put everyone at risk. Using the model, I find that overall, software market concentration has

Cyber Risk Posed by 'Single Points of Failure." I am quoted in the article as highlighting the issue of concentration across many points in the healthcare system, including electronic medical records software:

Potential trouble spots include electronic medical records, said Chitra Marti, a doctoral student at New York University's Stern School of Business who is studying the economics of cybersecurity in healthcare. The dominance of a handful of suppliers of those records may be understandable—but it is still a risk point, she said.

"Just like in the payments processing world, doctors and nurses don't necessarily want to be learning new software every time they change jobs, but that's the software that is responsible for holding and protecting our most sensitive data," Marti said.

had a negative effect on cybersecurity outcomes, and that there is scope for antitrust policy in preventing future cyberattacks in the U.S. healthcare sector.

2 The State of the Literature

In this section, I briefly and non-comprehensively review several related strands of literature to contextualize the findings of this dissertation.

The general question of how market power negatively impacts economic outcomes, especially those that are outside typical price or quantity effects, has spurred a growing literature on, broadly, the "non-market effects of market power." One such non-market effect is the proliferation of cyberattacks and general security concerns that can also be classified as "supply chain resilience." The direct question of how economic incentives shape cybersecurity outcomes itself rests on a literature spanning game theory, macroeconomics, and information systems. Further, due to the importance of the healthcare sector and the data available, a few have studied specifically how healthcare cyberattacks proliferate.

2.1 Non-Market Effects of Market Power

Market power, or the ability of firms to set positive markups and extract surplus from the consumer, is generally associated with higher prices and lower quantities, an inefficient market outcome. A growing literature studies the effects of market power outside of the market, concerned with outcomes like productivity, innovation, and resilience against shocks.

While it seems clear that markets have become more concentrated, the effects have not always aligned with the theoretically-predicted negative effects. Ganapati (2021) shows via industry-level estimates that concentration increases were actually positively correlated with productivity and real output growth, which suggests **increases** in consumer welfare thanks to concentration. The literature there argues that barriers to entry essentially posit a selection effect that weeds out unproductive firms, raising average productivity. Other papers, such as Autor et al. (2020), suggest that "superstar" firms increase productivity and welfare despite concentration. Others have shown that market concentration at the national level might actually decrease concentration a household experiences at the local level (Benkard et al. (2021), Rossi-Hansberg et al. (2021)). Many tech firms experience economies of scale thanks to growing networks, resulting in natural monopolies that are highly productive. In particular, Farboodi and Veldkamp (2021) show the not necessarily adverse consequences of market power accrued through a data-heavy production process.

This paper contributes to the literature on unexpected effects of concentration by examining a different and perhaps under-studied mechanism: the market concentration may decrease resilience to negative shocks (or equivalently amplify the effect of shocks), particularly those of a malicious actor. I show how the accrual of market power by firms providing healthcare technologies may lead to an underinvestment in product quality along the dimension of security. In addition to the usual effects of low quality, low security is unusual because of the ability of attackers to engage in *scaled* attacks or to take advantage of *contagion* so that the attack has follow-on effects on unrelated entities. That is, market power and underinvestment in security then amplify negative shocks and decrease the resilience of the entire system. Jamilov et al. (2021) confirm that firm-level cyber risk is positively correlated with firm size, but do not further investigate the *impact* the attacks can have when they target larger firms.

Other papers have explored shock amplification in the context of the COVID-19 pandemic and monetary policy, which are exogenous, non-strategic shocks (Hyun et al. (2020), Wang and Werning (2020), Mongey (2021)), and find mixed effects: market power can give the firm room to absorb a shock, but downstream firms may suffer from a lack of options. Geer et al. (2020) discuss non-technically what one might expect the effect of market concentration to be on cybersecurity risk, splitting the effect into threats, vulnerabilities, and impacts and finding qualitatively ambiguous effects of market concentration on each stage of the cyberattack.

2.2 Supply Chain Resilience & Shock Transmission

Relatedly, the paper thus aims to speak to the growing literature on the *role of firm networks in transmitting shocks* throughout the economy. Oberfield (2018) looks broadly at how network linkages can increase productivity. Hulten (1978) argued networks do not have first-order effects on efficient economies, a result that depended on the exogenous structure of the network. Baqaee and Farhi (2019) and Lim (2018) examine lower-order terms and show that with endogenous networks, shocks to an individual sector aggregate non-trivially to illustrate the role of networks in the business cycle. Further, Duffie and Younger (2019) implicitly illustrate the need for a federal backstop to contain the contagion of a financial cyber shock when intermediaries are not able to fully absorb shocks

As evidenced by the Change Healthcare attack, an attack on a crucial intermediary can have downstream effects that touch firms at every level down: pharmacies, hospitals, healthcare systems, insurers, and, at the end of the day, everyday patients. That is, network structure is not a trivial artifact of a macroeconomy, but a crucial input into understanding the effects of any kind of negative shock.

Many papers characterize *underinvestment in security* when transmission can occur through a network, but for computational reasons typically take either nodes (firms) as identical or the network as exogenous and symmetric (meaning nodes all have the same number of edges). Galeotti et al. (2020), Goyal and Vigier (2014), and Larson (2011) endogenize network formation but with identical nodes focus on symmetric networks to characterize the level of underinvestment in security (the papers differ in the role of the attacker and specific payoff structure). Yet in practice, firms are not symmetric, and thus the network has asymmetries that must influence a strategic attacker in nontrivial ways. For example, the small firm SolarWinds was targeted in December 2020 not just for its dramatic underinvestment in security but specifically for its connections to Microsoft and the Pentagon, two large institutions from which the attackers sought to extract value (Chesney (2021)).

Acemoglu et al. (2016) focus on asymmetric networks, but still take firms as identical and thus must take the network as exogenous. Their model introduces the possibility of *over-investment* by some firms while others underinvest. If a strategic attacker observes security investment, then some firms over-invest in an effort to transfer risk onto other firms. In the context of healthcare technology, that would suggest large vendors might be *over*-investing in security as a deterrent, while small vendors would remain vulnerable.

However, it is clear that attacks target large or well-connected firms make more lucrative targets, and thus firm size and market concentration should play a role in determining attack patterns. Large firms have more value to protect, and thus have greater incentive to shift risk to other firms if the attacker is constrained. O'Donnell (2008) shows how such size imbalances can induce risk transference to illustrate how Apple products initially appeared to be totally secure: Microsoft was the dominant platform and thus likely attracted more threats.

The Acemoglu model only incorporates network centrality, not firm size or market share, and thus would not be able to explain the phenomenon in O'Donnell (2008).

2.3 The Economics of Cybersecurity

The question of how to best secure systems against malicious actors is no longer considered a purely technical question (Soo Hoo (2000), Anderson (2001), Schneier (2008)) but rather an economic one about incentives and trade-offs. Basic economic theory suggests that positive externalities result in suboptimal investment, and the case of cybersecurity is no different. The final victim of a cyberattack is not necessarily the technology provider whose product facilitated the attack, but the consumer whose data has been breached. Various papers have theoretically explored how software firms may or may not be incentivized to invest in security along with how users may be incentivized to add in their own security in the context of direct breaches, ransomware, vulnerability disclosures, and other simple polices (Herley and Florêncio (2008), Florencio and Herley (2011), Choi et al. (2007), August et al. (2014), August et al. (2021), Niculescu et al. (2012), Arce (2020), Arce (2020)).

Empirical studies of cybercrime are few and far between; these tend to focus on very specific cases that limit their possibility for policy analysis. Crosignani et al. (2021) find that the indirect effects (in terms of stock prices) of the NotPetya attack were driven by customers of the victim who have few alternative suppliers – meaning upstream concentration may worsen an attack. They also find that affected firms made persistent adjustments to their supply chain network after their attack, which I test in the healthcare setting. In a "pre-mortem" analysis Eisenbach et al. (2021) find that an attack on any one of the five largest financial institutions' ability to make payments via FedWire would result in a bank-run-style scenario with over 38% of total market bank assets affected, suggesting concentration and network centrality matter jointly.

2.4 Cyberattacks Against Healthcare Systems

Hospitals and healthcare providers over the last two decades have incorporated digital technologies and completely transformed the healthcare experience.

The effect of a data breach in a hospital – once it has actually occurred – is grave (Choi et al. (2019)). In particular, McGlave et al. (2023) find that ransomware attacks on hospitals – which can completely debilitate a hospital's ability to access its digital systems – decrease hospital volume of patients admitted and increase mortality for patients admitted.

One might expect a mechanical increase in data breaches and cybersecurity incidents after hospitals digitize, simply because information is exposed to remote attacks. However, many data breaches are the consequence of human error and can be either physical or digital. As I show in Section 4.2, many breaches are not actual crimes but simply mistakes due to provider errors (e.g., lost papers). McLeod and Dolezel (2018) find using simple logit regressions that, broadly, increased digitization and connectivity are positively correlated with cyber-driven data breaches. However, they take digitization as given and do not distinguish between types or levels of digitization. Clement (2023) finds breaches are likely to increase around hospital M&A activity, possibly due to increases in human error as technology systems are synchronized. Finally, Kwon and Johnson (2015) find a data breach has no immediate impact on patient choice of hospital, as the healthcare market – like the healthcare technology market – is often concentrated with little room for patient choice.

This paper is perhaps most closely related to Kim and Kwon (2019), who study how EHR adoption affects hospital breaches. They find specifically that implementation of EHRs, particularly if hospitals aim for the "meaningful use" objectives outlined in the HITECH Act, increases data breach incidents, particularly accidents. In general, they examine the "extensive margin" of EHR choice: implementation or not. However, they do not specify the mechanism through which adoption affects breaches, and, importantly, focus on EHRs as a monolith rather than the specific products adopted by the hospitals. This dissertation builds and expands their analysis in multiple ways. Using a theoretical model as a foundation, I examine specific externalities that result from hospital choice not just of whether or not to use an EHR but also which EHR product and its security characteristics. With more detailed data on hospital characteristics and technology choices, I can distinguish between types of breaches and isolate the effect of vendor concentration. To my knowledge, no other study has examined specifically examined the impact of the competitive market structure of healthcare IT services on data breach outcomes, nor conclusively discussed policy solutions such as security minimums and antitrust activity.

3 Theoretical Model

As described in the Introduction, cybersecurity risk is distinguished from other economics shocks in two ways:

1. **Strategy**: the choice to perpetuate a cyberattack is not random but a *strategic* choice made by a malicious attacker who observes a system and makes choices in order to maximize an objective.
2. **Scale**: just like the digital technologies they seek to compromise, cyberattacks benefit from near-zero costs of replication and communication, meaning a common vulnerability may be repeatedly exploited by an attacker across targets and over time.

Largely due to scale and network effects (Katz and Shapiro (1985)), many markets for digital technology seem to have either one or two large firms: Google in Search; Meta (Facebook, Instagram) and X (Twitter) in Social Media; Microsoft and Apple in desktop operating systems; iOS and Android in mobile operating systems; Epic and Cerner in Electronic Medical Records systems; and more. Users find themselves with fewer and fewer true choices about the technologies they interact with daily. Most such concentration is primarily due to the physical facts of digital technology and the regulatory landscape: digital technologies experience both positive network effects (where a key product characteristic is the ability to interact with other users) and economies of scale (low marginal costs of adding users). At the same time, facing the exact same bugs and vulnerabilities, cyberattacks can use the same attack on many users all at once, creating massive destruction with just one investment.

How does such concentration affect cybersecurity outcomes? In this section, I describe four forces that shape the relationship between market structure and cybersecurity outcomes. For a more comprehensive elucidation of these forces, please see Marti (2024).

In this setting, the firm makes an investment in security to protect its own assets; its security investment also benefits users in that they are less likely to lose their own value following a cyberattack. Users are able to invest their own security. Attackers make threat investments against both the firm and then, if the firm is breached, against the users of the firm's product.

3.1 The "Floodgates Externality"

The firm's choice of defense is decreasing in its expected losses, but the more the firm invests, the less each user will experience in losses. Because the firm does not internalize the users' losses fully, firms will underinvest in security in both competitive and monopoly market structures. One can imagine that such underinvestment may be lower under competition, a a competitive firm may stand to lose more under a breach as users can switch products, leading to knock-on losses in addition to direct losses from the breach that the firm can internalize.

On the other hand, if users do not factor security into their product choices then competitive firms may also act as conduits for attacks. In that case, both the monopolist and the competitive firm would have very little incentive to provide security.

3.2 The "Gatekeeper Externality"

The cost structure the firm faces for providing security will determine the security level. If security costs have zero marginal cost in the number of users, there is no place if the product is popular, it does not automatically become more or less secure.

Under zero marginal cost for securing additional users, the monopolist can exert a *positive externality* on its users that I term the "Gatekeeper Externality": all users are hidden behind one potentially very large wall, rather than every user behind their own short wall, and adding more users costs nothing. Therefore, the monopolist's choice of security for any one of its users protects other users for free.

Put simply, if both the competitive firms and the monopolist each seek a level of security, then to achieve that level the monopolist spends only once while the competitive firms must *each* spend the same amount, duplicating costs. In that way, if the model's assumption of economies of scale for the firm holds, then there can actually be a positive externality from the firm to every user. Users may take into account the differences in security when choosing their products in the pre-model choice setting, allowing for a transfer via price that accounts for the externality.

On the other hand, if security costs differ by competitive structure in some meaningful way, the gatekeeper effect need not hold. In particular, if the cost function of the monopolist does *not* show economies of scale, one could even see diseconomies of scale depending on the relationship. The gatekeeper effect therefore only exists with the assumption of a scaling cost function.

A Weak Gatekeeper Can Worsen Attacks Do we expect a strong gatekeeper effect? In Geer et al. (2020), large firms are described as likely to be more internally complex, with software meant for a variety of use cases naturally likely to contain more vulnerabilities. As the popularity of the product grows, and it becomes "general purpose," it may grow labyrinthian, with new vulnerabilities. In the model, such a growth in complexity can be modeled as *dis*economies of scale in the number of users. Large firms may then find it harder to find and patch vulnerabilities than small firms.

On the other hand, small firms may face level high costs to implementing security, e.g. hiring their first security specialist can be expensive. Under competition, small firms would then simply not be able to provide the same level of security. Bouveret (2018), for example, finds that small and medium-sized banks are more likely to fold under a cyberattack than larger firms. In its App Store Monopoly antitrust case, Apple has argued it can provide better security as the monopolist, acting quite literally as a gatekeeper (*Bloomberg* (2022)).

The question is then whether a vendor's cost of security is *concave* in the number of users it serves or *convex*. In the concave case, which is also the one where the firm has zero marginal cost for adding users to its secure software and is the case I examine in this section, the gatekeeper effect is positive: large firms find it easier to secure each of its users. If the cost function is convex, however, then the gatekeeper effect is negative: large firms find it harder to secure each of its users.

3.3 The "Magnet Externality"

I also assume the attacker experiences economies of scale when attacking the monopolist: as they are only seeking to breach the firm, their only variable of concern is the single cost investment in finding breaches. Once the attacker has developed the exploit for the firm's vulnerability, she can deploy it costlessly to every user:

This costless deployment drives what I term the "Magnet Externality": although the attacker's costs do not change in the number or value of users, her expected utility increases, driving higher investments. A new user of the monopolist's product, then, has a *negative externality* on all other users, by driving up attacker investment.

The externality is akin to the "threat attraction" effect described in Geer et al. (2020), namely that larger firms are going to attract more threats because of the potential value at the end of the attack. An attacker's basic cost-benefit analysis results in her investing in more threats against larger firms.

In the competitive case, user-to-user externalities are minimal, especially under the extreme where each user hides behind their own firm. The "Magnet Externality" captures the negative user-to-user externality that occurs exclusively in the monopoly case. The user-level externality that cannot necessarily be easily corrected through transfers (so the Coase Theorem does not apply). To maintain the same level of security, a firm that serves more, or larger, users would need to correspondingly invest more in security.

The effect on overall welfare of an increase in the value of any single user – e.g., a large hospital digitizes its operations – is ambiguous and will depend on how the expected losses to all other users and additional defensive and attacker spending outweigh the user's value increase from digitization.

Quality Differences Exacerbate the Magnet Effect It may either be that the competitive firm or the monopolist firm provides greater quality, and therefore more potential value loss by users depending on exactly what we mean here by quality.

In general, competitive firms are thought to provide higher quality than the monopolist, as long as quality is a dimension that affects consumers' choices (Tirole (1988)). On the other hand, digital technologies specifically benefit from *positive network effects*, where using the same product as other users leads to greater per-user value for the product. Then, we would see wider adoption and usage under a monopoly case than a competitive case, as discussed in Katz and Shapiro (1985).

Here, positive network effects would show up in the "User-Level Negative Externalities": users who see greater value in the monopolist case may also increase the value the attacker can capture. Positive network effects would therefore *exacerbate* the magnet effect, making the monopolist case even more lucrative to the attacker than the competitive case.

3.4 The Sum-of-Efforts Externality under Monopoly

Finally, I examine the role the user's own choice of investment–on top of what is provided by the firm–plays in determining outcomes. The user's security investment function essentially to counteract the user-specific values, by allowing the user to add in their own security and decrease the overall impact of the attack. The attacker, when she encounters such user-level security, will find her attack failed.

We see again that *any* user-level security helps decrease the investment of the attacker, and therefore the expense of the firm. Less trivially, though, is what I call the "Sum-of-Efforts Effect": under the monopolist, the relevant parameter is not the individual's security, but rather the sum total of security (multiplied by the user's value).

As such, under the monopolist case, users' investments in security have positive externalities by decreasing the attacker's overall value to be captured, and therefore her overall investment – which translates into lower overall attack probabilities.

In the competitive case, users only rely on the firm to protect them – not each other – and so there is no change in one user's probability of being breached if another adds security.[2]

Security Bundles Amplify the Sum-of-Efforts Effect A monopolist could more easily mandate that its users implement certain security practices, e.g., multifactor authentication, while competitive firms may leave users with more choices. For example, Microsoft announced a rollout of MFA and a redesign of the software to force all users to use MFA (WSJ). It is certainly possible for the software developer to bundle in other security practices, so the user need not shop around and implement their security completely alone. In that case, not only would users see a direct boost to their own breach safety thanks to the new security, but they will boost other users' security posture too by deterring attacks at the firm level. Therefore, the sum-of-efforts effect can be amplified if the technology provider bundles in good security practices.

In a competitive environment, however, firms may differ in their ability to bundle in security practices. Firms may make different choices based on perceived competitive impact. The GDPR privacy protections, through a number of mechanisms, were found to decrease website engagement and traffic (Congiu et al. (2022)). They may also, later on, serve to decrease a firm's liability if the user's security was insufficient, as in the Change attack where lack of MFA was blamed for the attack – rather than the product design where an administrative user could shut down a third of U.S. insurance payments. Firms may even compete on the amount of security they force users to take on: users who dislike having to pay the [time] cost of using MFA or securing a password could, in a fully competitive and informed environment, switch vendors to those that have lower security consciously. In competition, then, we could see a wide variety of user security choices. Users may be more free to make their own privacy and security choices, but will also not receive any positive externalities from other users' investments.

[2]Note that the lack of effect is due to the attacker's lack of a budget constraint. The attacker simply chooses each threat separately to maximize expected value. The existence of a budget constraint might create tradeoffs between types of threats and result in a risk-redistribution effect. I direct the reader to Acemoglu et al. (2016) for a model specifically highlighting the risk redistribution phenomenon.

4 Empirical Application

In this section, I describe the data I use to test the magnitude of the four forces described in Section 3 in the context of data breaches in U.S. hospitals and healthcare providers. I omit from this version of the paper a discussion of the regulatory landscape in favor of a focus on the industry and data used in this specific empirical application.

4.1 Data

In this section I describe the datasets I use throughout the rest of this paper to analyze competition, cybercrime, and hospital digitization activity. A major part of the work of this paper has been constructing crosswalks between the three datasets, which, subject to Data Use Agreements, I aim to make available to other researchers. I describe how the matching process shapes the sample I analyze and compare the quality of the datasets when topics overlap.

The study of data breaches in healthcare is facilitated by the extensive data available. Empirical studies of cybercrime typically suffer from at least three issues:

1. **Disclosure Bias**: Targets are often disincentivized from disclosing the existence of a data breach, either due to the anticipated reputational impact (Ford et al. (2021)) or for fear of follow-on or copycat attacks (Choi et al. (2007)). In settings where firms are not required to disclose, it is unlikely that all firms would disclose; only large, public firms with shareholder obligations would be expected to disclose attacks. Under the HITECH Act of 2009, however, any data breach that affects more than 500 individual records must be disclosed. Although disclosure is still voluntary and audits do not occur, the mandate certainly goes farther than any other sector did at the time.

2. **Assigning Fault**: While the target may through its own investigation understand why a breach took place, they are often not required to disclose such information. By pairing the data breach information with a verbal description provided by the entity and data on the technology vendors used by the breached entity, I can look for latent commonalities between attacks and move towards assigning fault – key for my counterfactual analyses. The key assumption here is that the causes of attacks are common across hospitals, and not that each breach is entirely unique and special.

3. **Success Bias**: targets may only be aware of attacks that have actually taken place; it is not always possible to know if *attempts* were made (or are ongoing). The healthcare data I use in this study will not address the success bias issue. Instead, I will distinguish between *mistakes* and *crimes* to detect when a third-party is involved in the breach versus when the software may simply be confusing or humans may use it incorrectly.

Empirical studies of cybersecurity are therefore frequently limited by the lack of detailed, definitive data on the incidence of cyberattacks and victim's security landscapes. That lack of data is at least somewhat by design: most victims are companies and institutions, who worry that "any such information may be used to criticize their security posture or, even worse, as evidence for a government investigation or class-action lawsuit" (Schneier (2024)). The challenge of this paper is then to draw inferences on how and why hospitals are experiencing data breaches when explicit data is not available.

To do so, I combine datasets that each provide difference sides of the equation.

In Section 4.2, I introduce the U.S. Department of Health and Human Services' administrative dataset containing information on every cyberattack affecting five hundred or more

healthcare records that was reported to the HHS under the mandate of the HITECH Act.

The HIMSS and AHA data, introduce in Section 4.3 and 4.5, provide information data on hospital technologies I use to form the other side of the equation: what was the technology landscape of the hospital at the time of its breach? Here, thanks to the other incentives of the HITECH Act, hospitals are able to report on their new technologies that lead them to the Medicare incentive payments. The data are quite detailed, and will allow us to get closer to *Assigning Fault*: what technologies are and aren't in use in a hospital at the time of its breach?

There continues to be a deep need for comprehensive data on cyberattacks and the security landscape of *all* possible victims, so empirical work can be done to evaluate which security practices are most effective. New policies, such as the Cyber Incident Reporting for Critical Infrastructure Act of 2022, will hopefully provide promising data sources that can be studied in the same manner as this paper for a wider set of industries beyond healthcare and possible victims (CISA).

4.2 U.S. Department of Health and Human Services (HHS)

Comprehensive data on breaches of healthcare information comes from the U.S. Department of Health and Human Services' Office of Civil Rights (OCR).

The HITECH Act created one of the first broad cyberattack disclosure laws in the world. Any covered entity (CE) was required, from 2010 onwards, to report to the OCR whenever a data breach that affected more than 500 individual records took place. Each incident is published by the OCR in a publicly searchable and downloadable portal (Breach Portal). Each report contains the following:

- Name of Covered Entity
- State
- Breach Submission Date
- Number of Individuals Affected
- Covered Entity Type: Healthcare Provider, Health Plan, Healthcare Clearing House, or Business Associate
- Type of Breach: Hacking/IT Incident, Improper Disposal, Loss, Theft, Unauthorized Access/Disclosure, Unknown, Other (may be more than one)
- Location of Breach: Desktop Computer, Electronic Medical Record, Email, Laptop, Network Server, Other Portable Electronic Device, Paper/Films, Other (may be more than one)
- Web Description: a short description of the breach, available for all but very recent cases

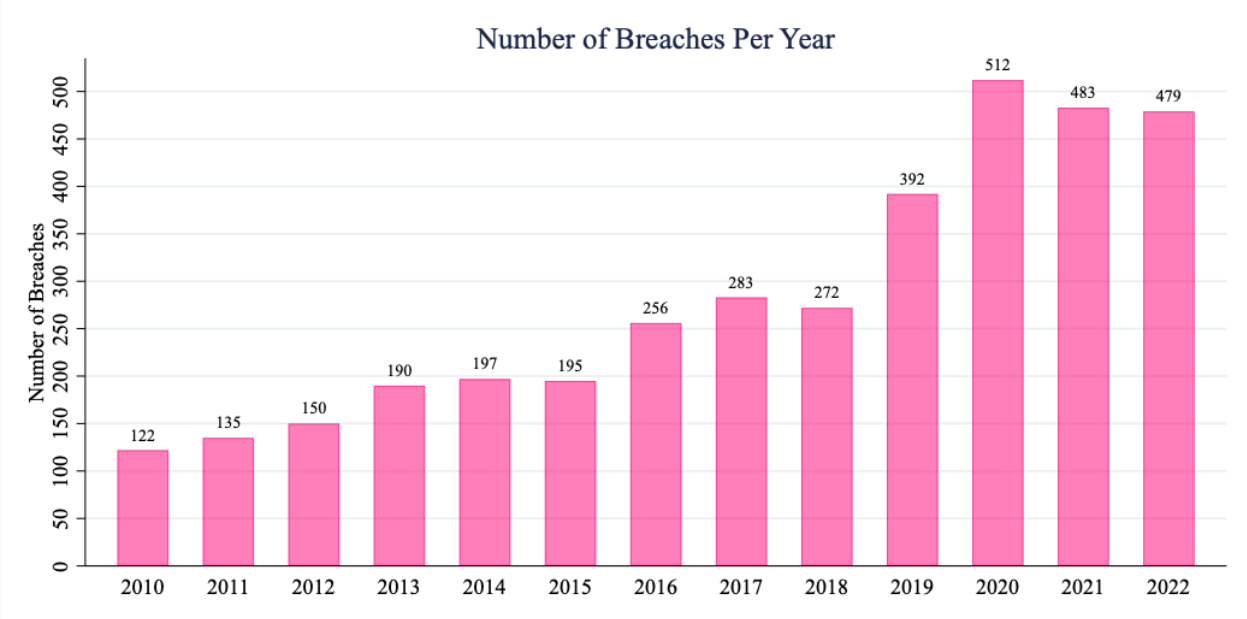
CEs are mandated to report any breach they become aware of; there is nothing to be done about breaches they remain unaware of. Indeed, the breach submission date can be interpreted as the breach awareness date. CEs who do not know or attempt to hide a breach will not be in the dataset. Individuals may file complaints with the OCR if they believe their data has been compromised, in which case the CE may not be able to hide the breach. Furthermore, attackers themselves have been known to report cyberattacks publicly or to authorities to bolster extortion attempts (Dark Reading). OCR will then investigate the breach and assign appropriate fines. Whether the breach was discovered as the result of a complaint or self-reported is not contained in the publicly available dataset.

In this paper, I limit the breach sample to the years for which I have data for the entire calendar year, 2010-2022. I also limit my focus to breaches that take place at healthcare providers, and in later sections limit the sample to only hospitals. Section 4.2 provides detailed descriptives for the HHS data. I limit all analysis to healthcare providers and exclude healthcare plans, business associates, and clearinghouses.

Figure 1 shows the number of breaches that were reported to HHS from 2010-2022. Consistent with expectations, breaches rose in the later part of the 2010s, saw a spike in 2020 likely related to the pandemic, and have remained high.

Further, later years show more individuals being affected by data breaches, as seen in Figure 2. In an alternative presentation, Figure 3 presents a histogram of individuals affected by each data breach from 2010-2022. The minimum is of course 500 – the minimum number that requires reporting to HHS – and follows linearly a log scale, meaning there are many outliers affecting over ten million individual records.

Figure 1: Total Breaches Over Time



4.2.1 Classification: Cyber vs. Physical

I use the description in conjunction with the reported Location of Breach to classify it as a "cyber" or "physical" breach. Cyber-breaches take place when the data breach did not require anyone to have physical access to the office or the records. For example, a ransomware attack that breaches a network server is remote. On the other hand, a physical attack relies in some way on having access to the specific provider's office. Note that physical breaches my still involve technology, such as someone breaking into an office and stealing a laptop on which ePHI is stored.

I manually label each breach "cyber" or "physical." When there is no web description available (about 25% of cases), I label breaches that are located in a"Network Server," "Elec-

Figure 2: Total Individuals Affected in Over Time

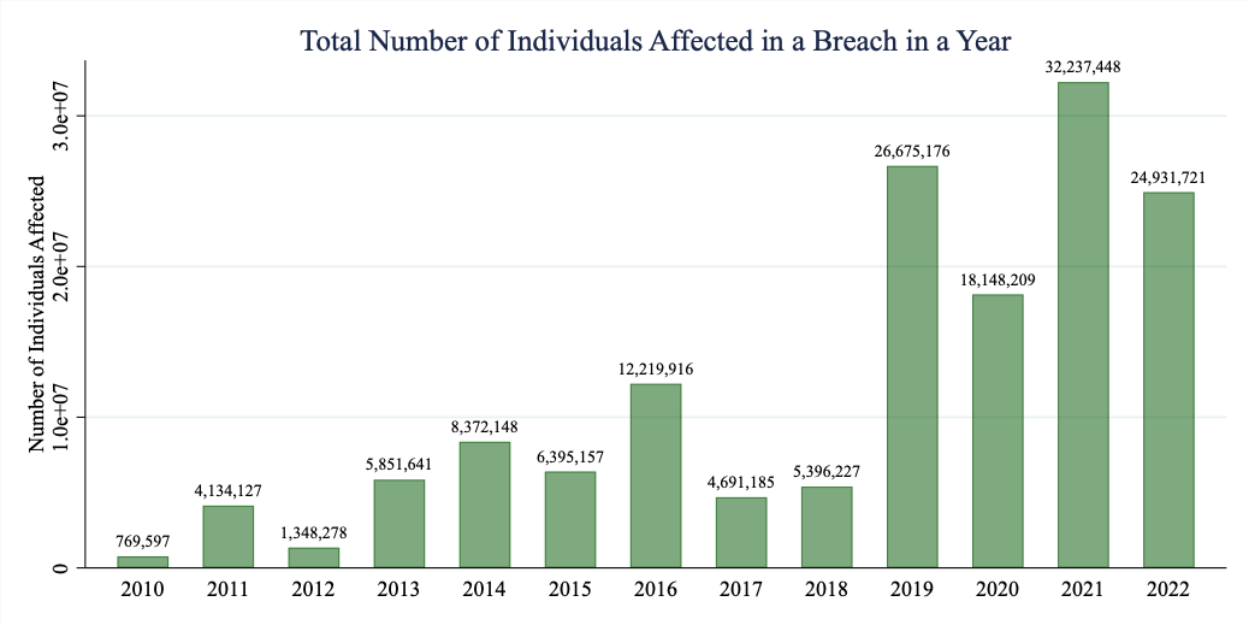
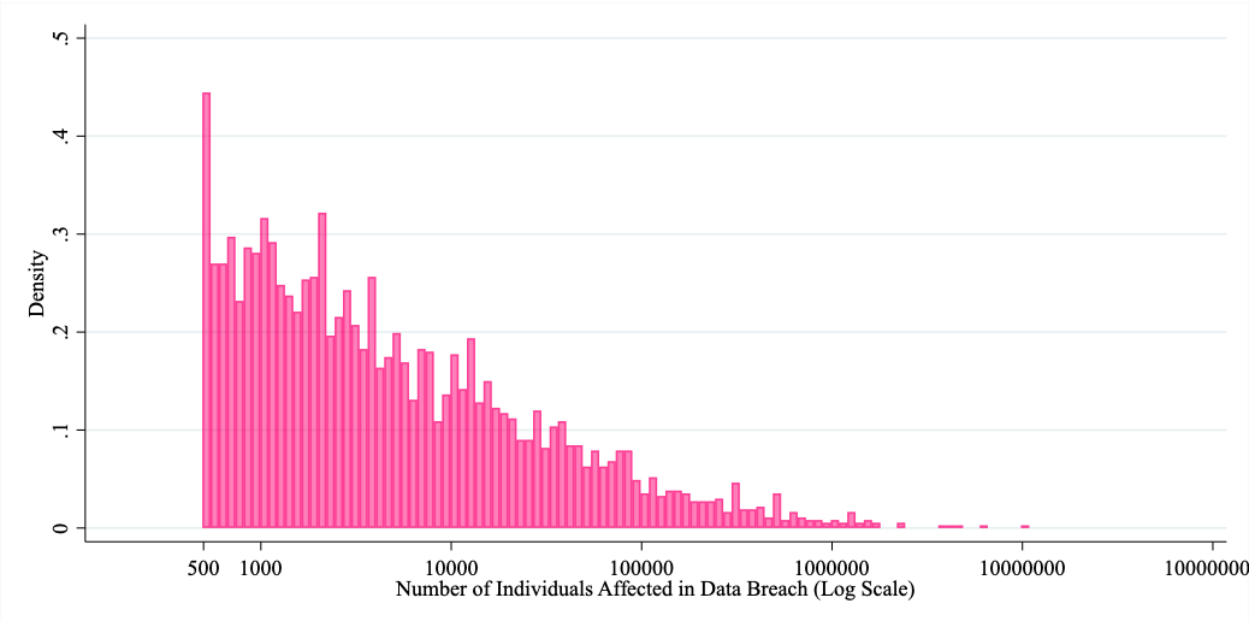


Figure 3: Individuals Affected in Data Breaches



tronic Health Record," or "Email" as well as those whose causes are Hacking/IT Incident (see Section refsec:crime) as cyber and the rest as physical. The procedure is manual, and in the most recent breaches, no description is available, so the process is subject to Type I or Type

II errors. However, to my knowledge, no other systematic classification of the breaches exist; other papers (e.g. Kim and Kwon (2019) rely on the same manual labeling process).

4.2.2 Classification: Crimes vs. Mistakes

I also use the description in conjunction with the listed Type of Breach to classify it as a "crime" or a "mistake." A breach is considered a crime if a malicious third party – possibly including an inside employee – deliberately breached records for the purpose of fraud. A crime may be local (stolen papers) or remote (a hack). Mistakes are often local (lost keys to a file cabinet, a fire) but can also be remote (a vulnerability in a technology that was discovered but not yet exploited).

I manually label each breach as a "crime" or "mistake." When there is no description available, I label any breach that includes hacking/IT incidents, thefts, or unauthorized access in its list of types as a crime and the rest as mistakes.

The goal is to distinguish breaches that may be actively perpetrated by a strategic cyber-criminal from those that may have been just as possible in the pre-digitization era. Cyber breaches explicitly rely on digital technology to occur remotely and are possibly scaled, while physical breaches rely on proximity. Similarly, mistakes are not the result of a strategic attacker, while crimes are. Hospitals may be more or less concerned about each category of breach, and strategies to mitigate them may differ as a result.

Table 1 shows the breakdown of total breaches by their location and type according to my classification.

Table 1: Types and Individuals Affected in Reported Data Breaches at Healthcare Providers

Year	**Count:**					**Individuals Affected:**			
	Total	Crimes	Mistakes	Cyber	Physical	Mean	Min.	Max.	SD
2010	122	89	33	42	80	6,308.2	500.0	83,945.0	12,139.4
2011	135	104	31	58	77	30,623.2	500.0	1,055,489.0	134,333.5
2012	150	123	27	62	88	8,988.5	500.0	315,000.0	29,799.8
2013	190	133	57	59	131	30,798.1	500.0	4,029,530.0	296,581.7
2014	197	131	66	65	132	42,498.2	500.0	6,121,158.0	437,074.2
2015	195	121	74	36	159	32,795.7	500.0	4,500,000.0	322,918.4
2016	256	163	93	109	147	47,734.0	500.0	3,620,000.0	280,767.1
2017	283	187	96	136	147	16,576.6	500.0	697,800.0	60,943.9
2018	272	173	99	145	127	19,839.1	500.0	566,236.0	65,673.6
2019	392	297	95	271	121	68,048.9	500.0	10,251,784.0	537,114.8
2020	512	419	94	382	129	35,445.7	500.0	1,045,270.0	85,096.7
2021	483	412	71	394	89	66,744.2	500.0	2,413,553.0	227,271.3
2022	479	456	23	450	29	52,049.5	500.0	1,608,549.0	165,819.3

Source: Full Health and Human Services Office of Civil Rights List of Breaches as of August 2023. The data used in this table cover all breaches across any healthcare provider. The data used in the analysis later removes some breaches, a process described in Section 4.6

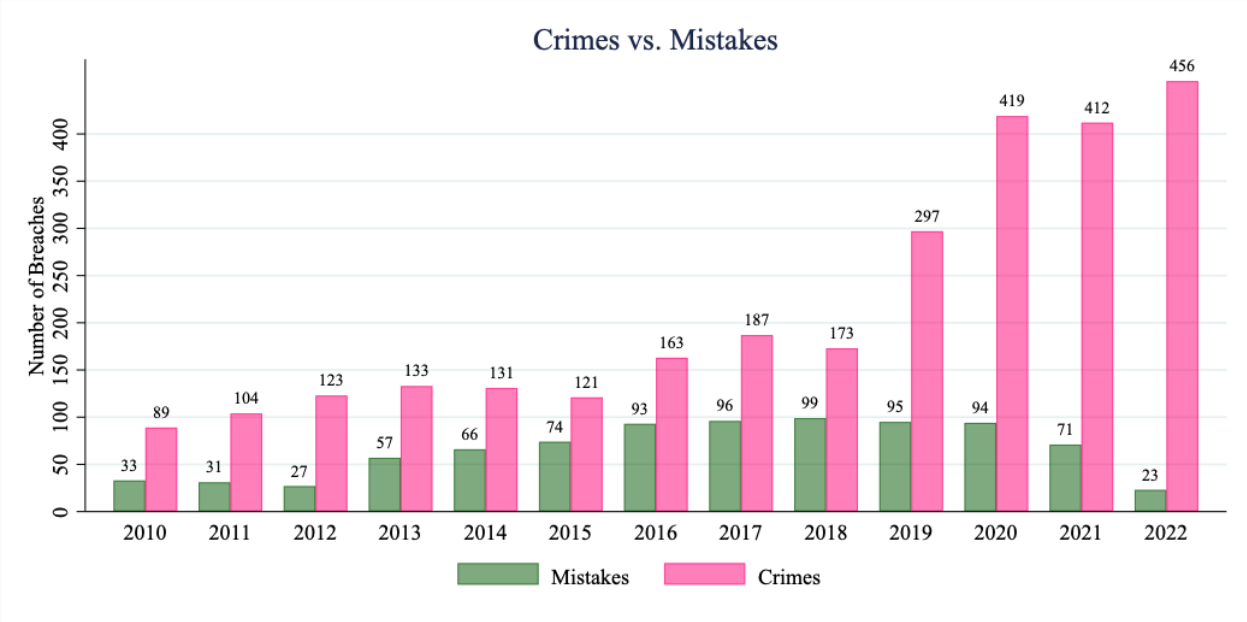
Figures 4, 5, and 6 show the growth in each type of breach over time.

In Figure 4, we see that while the number of mistakes per year has stayed relatively constant, crimes have been growing drastically over time – suggestive evidence that healthcare crimes have been – or should be – more of a concern for hospitals.

Similarly, Figure 5 shows that cyber breaches are growing while physical breaches have again stayed relatively constant. At the start of the sample period, physical breaches outnumbered cyber ones; by 2022 that is definitely no longer the case. The "death of distance" suggests again that there are greater opportunities for breaches available via digital technologies, while physical breaches may have some natural limit.

Finally, Figure 6 shows the breach type (crime or mistake) in conjunction with the breach location (remote or local). The category that has grown the most in the sample period is the set of remote-crimes: cyberattacks, in other words, that exploit digital technologies away from the site of the breach to use data for the purpose of fraud.

Figure 4: HHS: Number of Crimes vs. Mistakes



4.2.3 Examples of Breaches

Example of a Cyber-Crime *"On June 9, 2014, Revere Health, the covered entity (CE), discovered that cybercriminals had compromised one of its Internet-facing servers containing electronic protected health information (ePHI), affecting 31,677 patients."*

Example of a Cyber-Mistake *"Texas Health Harris Methodist Hospital Stephenville, the covered entity (CE), reported that a program coding error within its billing system allowed electronic protected health information (ePHI) to be mismatched with the incorrect account guarantor. This led to PHI being sent to the wrong recipients."*

Figure 5: HHS: Number of Cyber vs. Physical Breaches

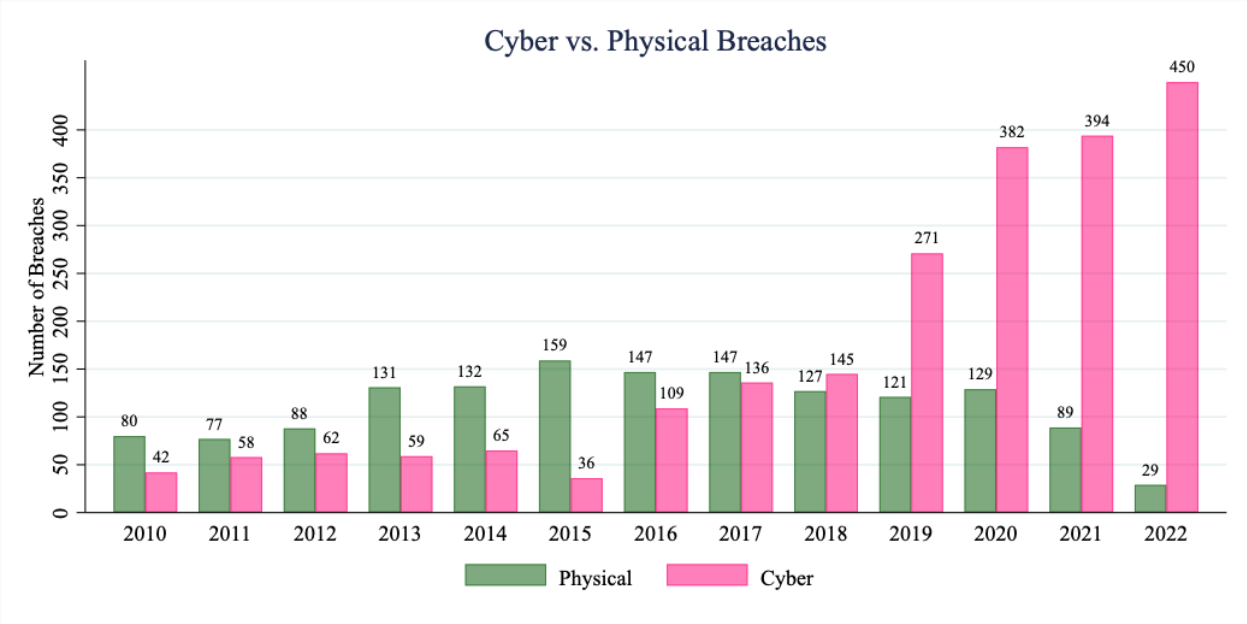
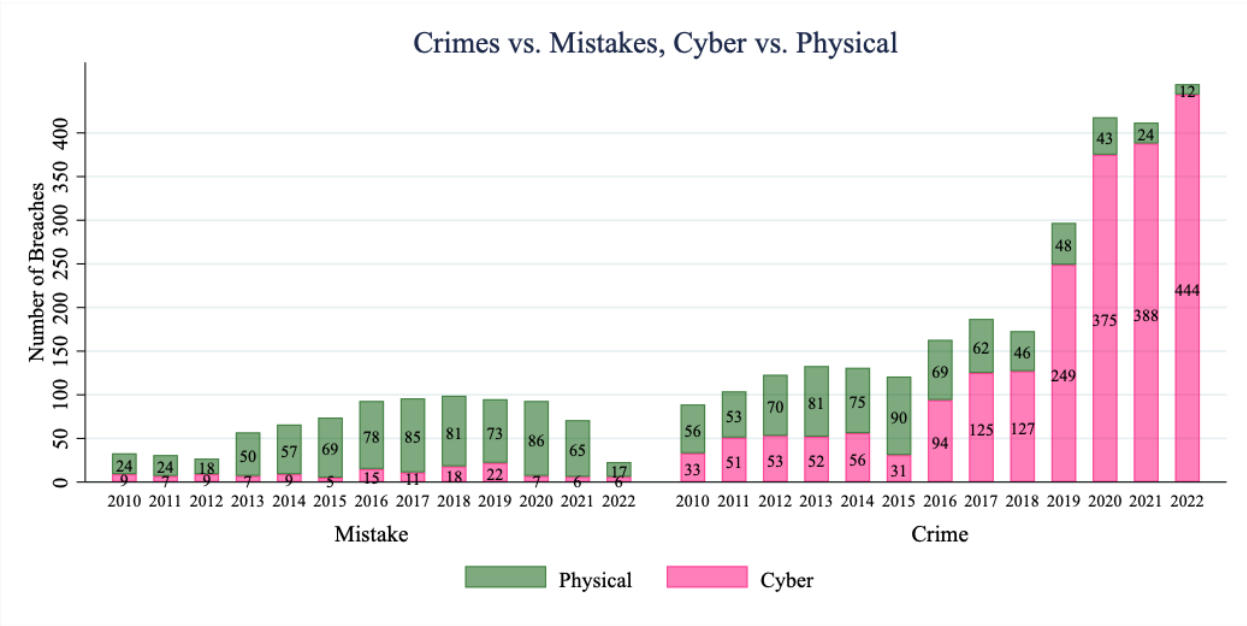


Figure 6: HHS: Number of Crimes vs. Mistakes *and* Cyber vs. Physical Breaches



Example of a Physical-Mistake *"On November 18, 2017, a physician employee removed patient files from the covered entity (CE), MidMichigan Medical Center-Alpena, and left them in a public parking lot in an unsecured container, which spilled out into the parking*

lot, and the wind subsequently scattered the records over several blocks."

Example of a Physical-Crime *"A clinical intern at the covered entity (CE), University of Florida Health Jacksonville (UFHJ) (formerly Shands Jacksonville Medical Center), took photographs of protected health information (PHI) and emailed the PHI to an unauthorized third person for the purpose of filing fraudulent tax returns. The PHI included the names, addresses, social security numbers, dates of birth, and treatment information of 1,025 individuals. Law enforcement agencies that learned of the breach informed the CE and requested delays of breach notification."*

4.3 Healthcare Information Management Systems Society (HIMSS)

I use data on hospital characteristics and hospital choices of technology vendors and products from the Healthcare Information Management Systems Society (HIMSS) Analytics Legacy Database. The HIMSS data span from 1989[3] to 2017, and are the result of a survey run by the organization surveying healthcare providers on their technology categories, applications, vendor choices, product choices, and more. The data are extremely detailed and paint a full picture of a healthcare provider's technology choices. Because the data are the result of a survey, standard cautions apply with respect to human error, mis- or under-reporting of technology use, and incomplete data. The data have been used extensively in other studies of healthcare digitization with success, suggesting the above issues do not preclude careful analysis.

The HIMSS data cover hospitals, systems, ambulatory and sub-acute facilities. In line with the HHS data, I focus on hospitals, which form their own market and are the most salient entities seeking to protect patient data and compete for insurance coverage. There are just under **5,000 hospitals** covered by the HIMSS data during the sample period of 2009 to 2017.

The next several figures illustrate the type and the level of detail of data available. In each, I show the number of hospitals that reported using a digital technology in the relevant category over time. We have, for each hospital, for each technology *category*, the specific application for which a particular *vendor* is contracted, and in many cases even the *product* used.

First, Figure 7 shows that more hospitals are reporting using more technologies in more categories over time – i.e., digitization is increasing, almost tautologically. Note that the increase is not uniform across categories; some categories, such as Telemedicine or Health Information Exchange, saw a greater increase than others.[4]

I focus the majority of this paper on three specific categories: Electronic Medical Record, Health Information Management, and Security. Figures 8, 9, and 10 show how within each category, the set of applications now digitized has also increased. Again, the increase has not been uniform: in 2010, for example, many entities already had a Clinical Data Repository technology provider, but not all had Computerized Practitioner Order Entry.

Next, the survey also asked healthcare providers to report the *specific vendor* that providers use for each technology application. In the below figures, I show the variation in concentration

[3]The questions asked in the survey have changed significantly over time. I primarily focus on the edition of the survey from 2005 onwards.

[4]I relabel the categories in some cases to combine old and new phrasing, or information across different sections of the HIMSS data. Other papers have usually focused on just the "Application" table of the HIMSS data; I make use of extra tables such as "Security" and "CDSS" to gain a more complete picture. I check the data to avoid double-counting and for internal inconsistencies at the hospital level.

Figure 7: Broad Categories of Technology Reported

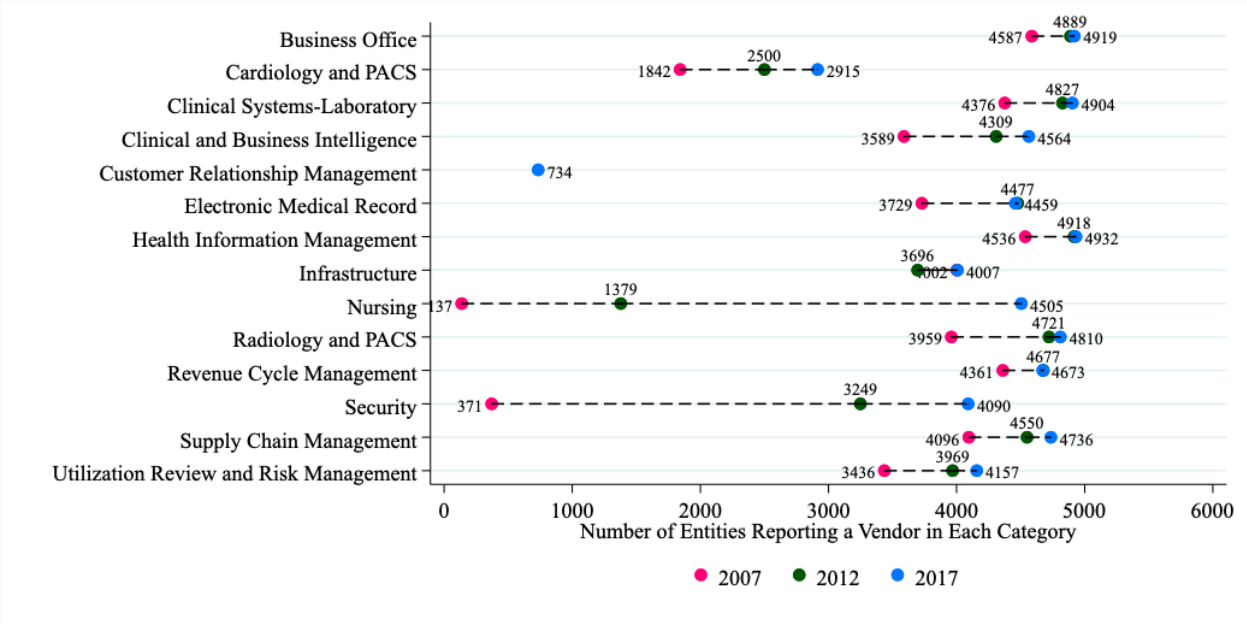
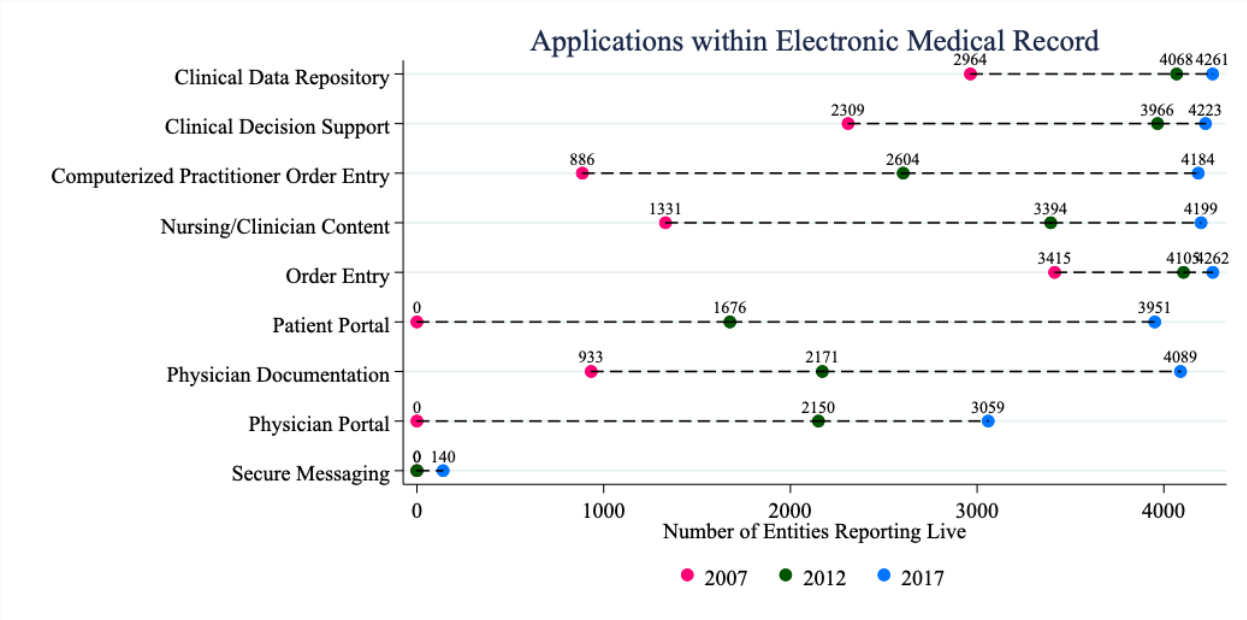


Figure 8: Within a Category, Specific Applications: Electronic Medical Record



across providers, within each application.[5] There is variation both within years and across

[5]Less than 1% of hospital-application combinations have more than one vendor reported, even after tables are

Figure 9: Within a Category, Specific Applications: Health Information Exchange

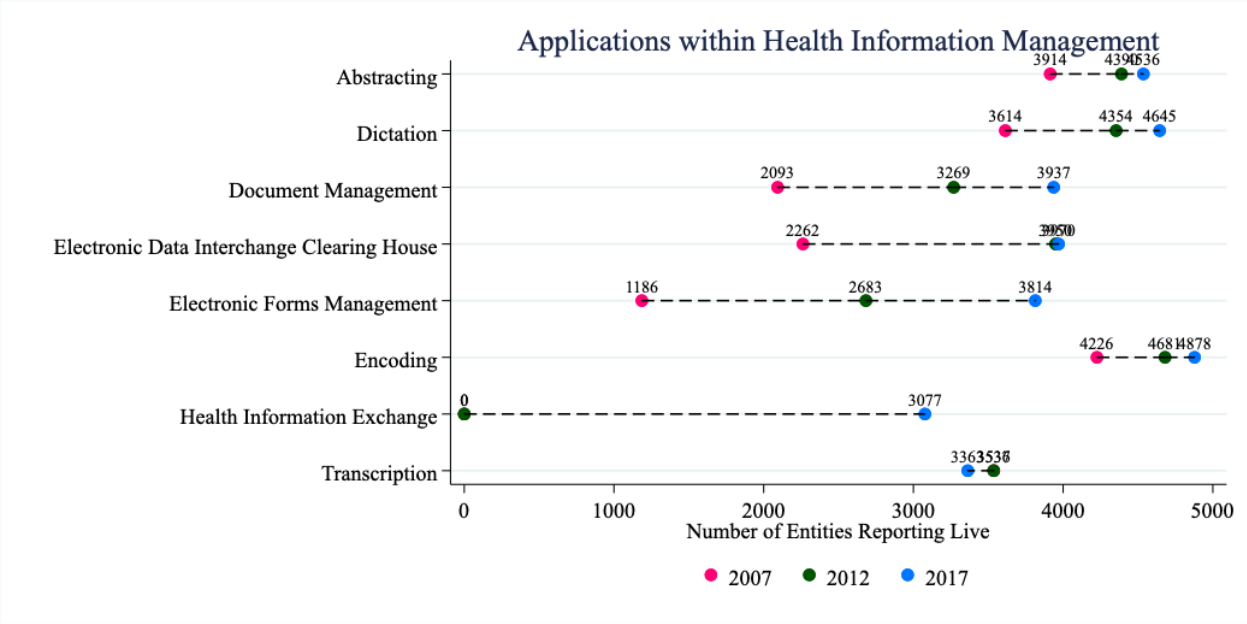
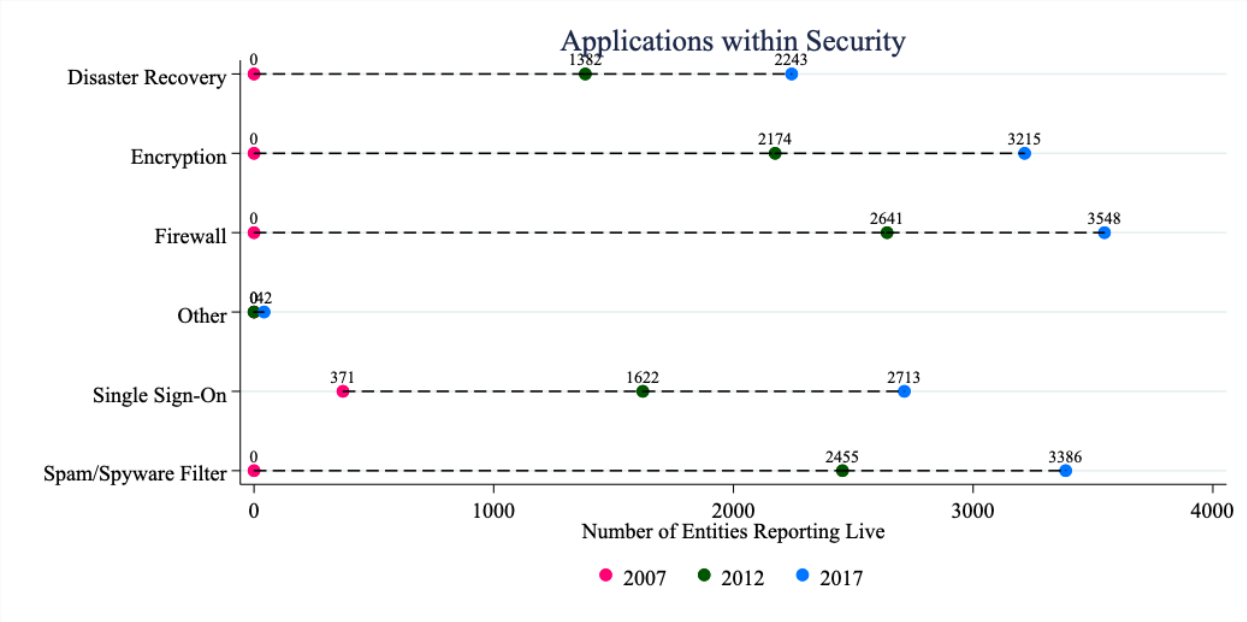


Figure 10: Within a Category, Specific Applications: Security

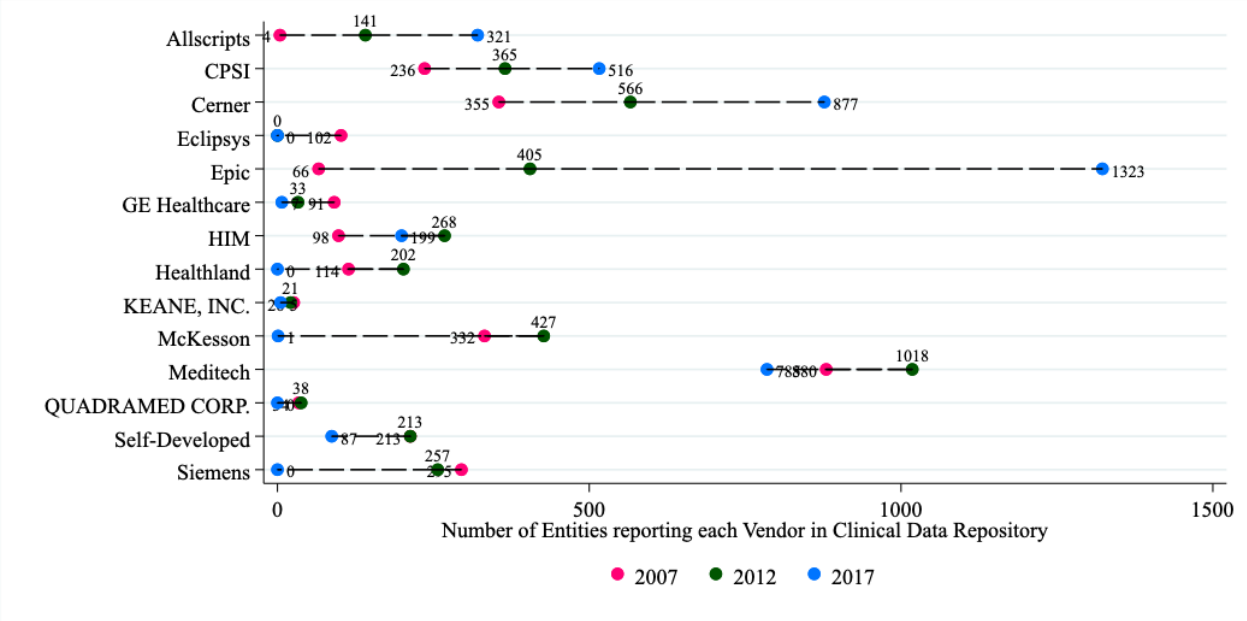


combined. If more than one vendor is reported, I lexicographically choose the vendor that either (1) is used again the next year or (2) has the higher market share.

years in the dominant vendor, the concentration of the market, and the overall market size within each application– each of which will be used in the analyses to come.

For example, in the Clinical Data Repository application (the main Electronic Medical Record application), Meditech dominated in the market in 2009, but by 2017 Epic Systems has the highest market share.

Figure 11: Vendors within Application: Clinical Data Repository



The data contain even more specific details. Within each vendor, we can see which *products* are chosen by specific healthcare providers. For the purposes of this paper, I do not use product-level information, which is often missing or incomplete.

4.4 Matching Process: HHS and HIMSS

Because the HHS data contain only the name of the breached entity and no other identifiers, I use a near-manual matching process to find the corresponding entity in the HIMSS data. I first match on the name and state as accurately as possible. Individual doctors' offices are matched to the healthcare entity to which the doctor belongs, if any; standalone providers are excluded. I exclude dentists, pharmacies, rehabilitation clinics, plastic surgery offices, and any other breached entities that would not be expected to have a corresponding official healthcare provider in the HIMSS data.

Out of about 3500 breaches from 2010 to 2017, I match 1,181 to hospitals in the HIMSS Data. However, I generally limit my analysis to breaches that occurred before 2017, which covers about 1500 breaches total and misses the (likely) pandemic-induced spike in cyber-crime. Within the matched breaches only, we still see the same pattern of rising cyber and rising crime breaches, and no dramatic shifts year-to-year.

Figure 12: Vendors within Application: Encryption

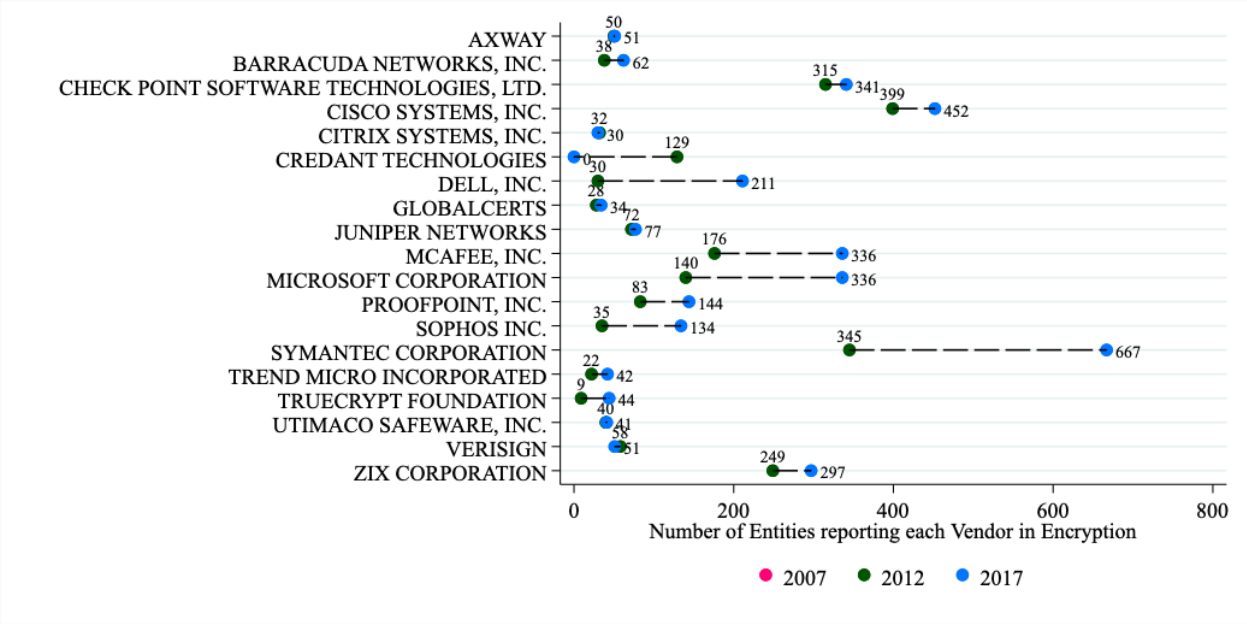
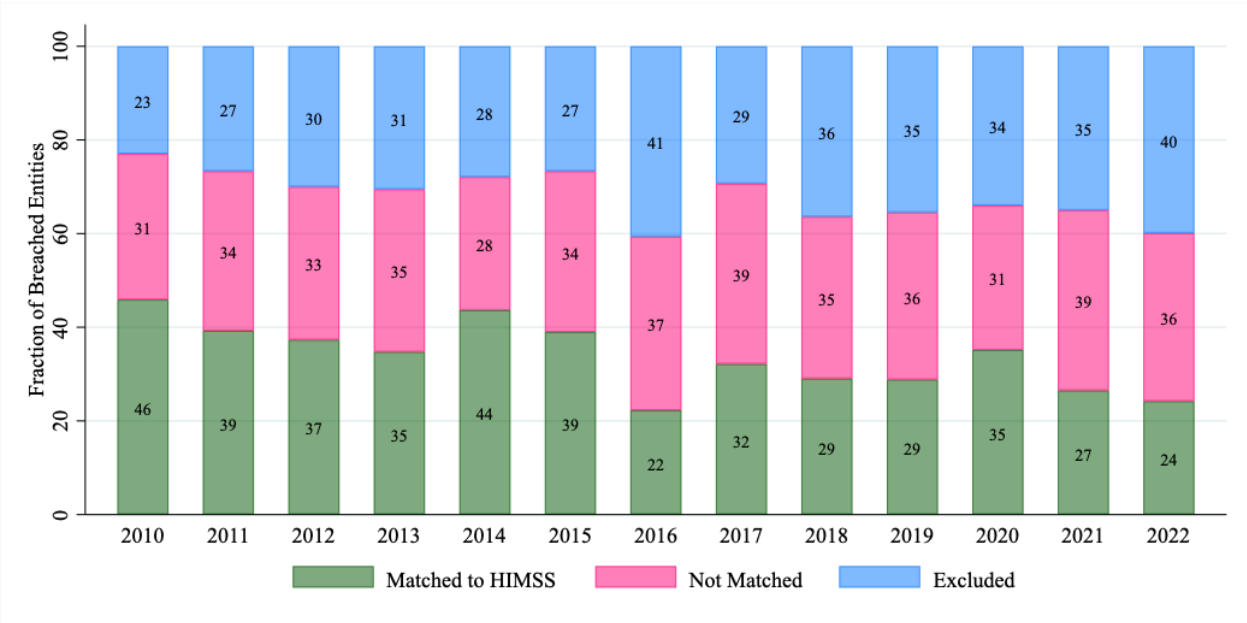


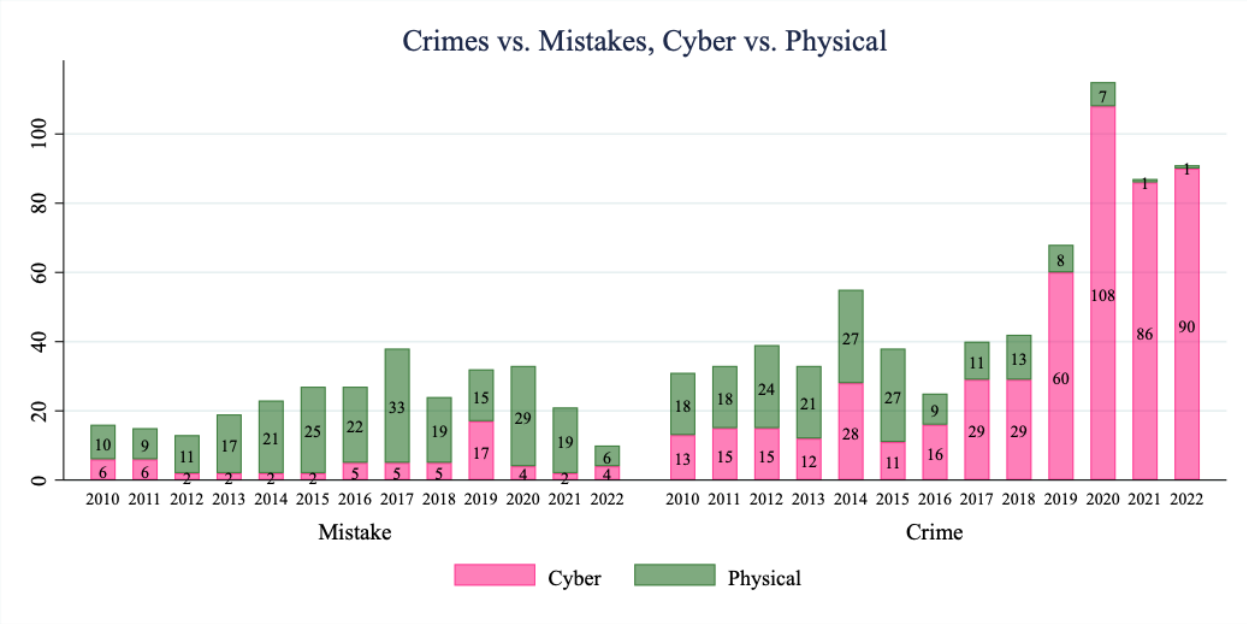
Figure 13: HHS to HIMSS Data: Fractions Matched and Excluded



4.5 American Hospital Association (AHA)

I also utilize data from the American Hospital Association (AHA) Annual Survey, which began in 1972 and runs to the present. I specifically use data from the Information Technology (IT)

Figure 14: HHS to HIMSS Data: Matched Crimes vs. Mistakes, Cyber vs. Physical



supplement, which began in 2008. I use the sample from 2008 to 2018. About 3,500 hospitals report their IT information to the AHA, which includes data on EMR use, meaningful use attestation, HIE participation, as well as subjective questions about what might be keeping the hospital from full adoption of digital technologies and their future plans.

The main AHA survey provides basic hospital information, such as size, the type of hospital (teaching, residency status). The IT supplement includes much more detailed information about hospitals' technology choices, up to and including specific functionalities of their EHR/EHR systems, the level to which they are involved with their regional Health Information Exchanges, and their main concerns and planned changes in their IT strategies. The survey changes slightly from year to year. The questions are designed both explicitly to address the Meaningful Use components of the HITECH Act, both explicitly after the Act was passed, and implicitly as they reflect the major healthcare technology policy concerns at the time of its passage.

4.5.1 Discussion: No Prices in the Data

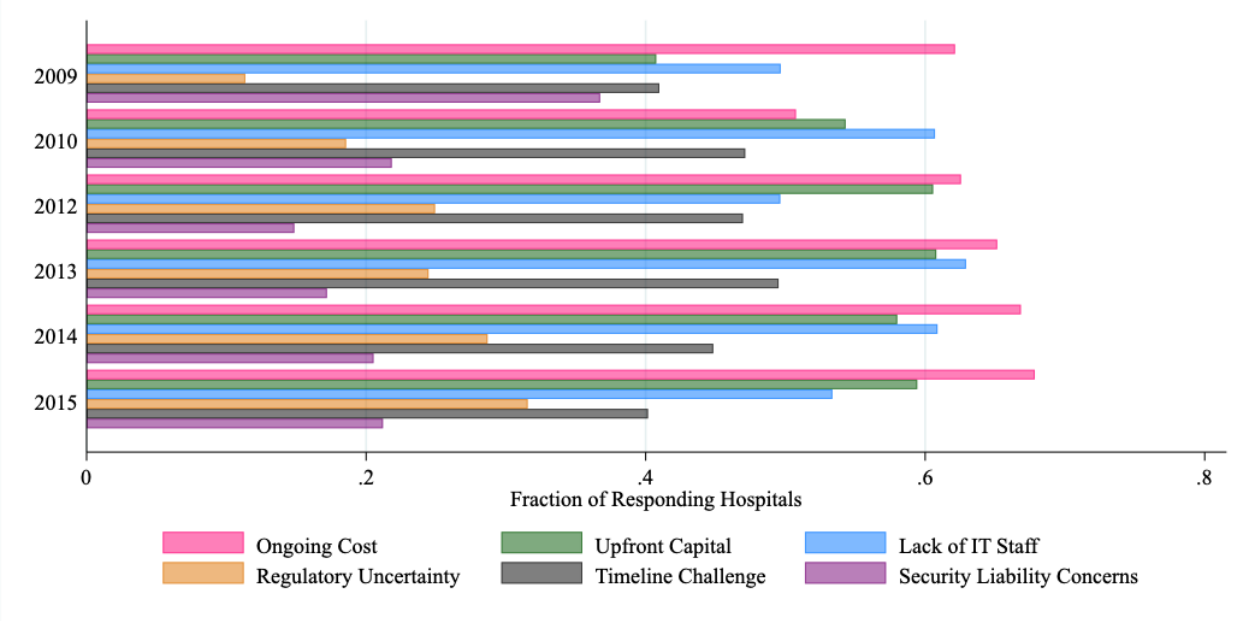
The data I described in Section 4.1, the HIMSS and AHA datasets, contain a wealth of information about hospitals and their technology choices over time. However, to an economist, they each lack one key piece of information: the price healthcare providers are paying for these technologies.

The price of an EMR system is often negotiated individually between hospitals and the EMR vendor, with prices ranging from thousands to millions. The market is therefore subject to price discrimination (first, second, and third degree), where comparable hospitals pay different prices for either the same or slightly different products, depending on characteristics like their location, revenue model, size, and even bargaining ability of their executives. The

lack of systematically reported prices is both a feature of the market – facilitating individual bargaining and therefore price discrimination – and a bug of this analysis, which must find alternative ways to address hospitals' price concerns.

Of course, we know that hospitals are price sensitive when installing an EMR system: they explicitly compare expected cost savings (realized over many years) with initial and ongoing costs. Furthermore, costs were listed as the main concern in Figure 15. The HITECH Act specifically sought to alleviate high prices through subsidies that decreased over time: addressing high upfront costs and then smoothing the ongoing costs for the hospital.

Figure 15: AHA Data: Why are you not pursuing Meaningful Use?



In the theoretical model, I consider the selection of software (EMR vendor) by the user (the hospital) to have taken place fully before security outcomes are realized. Using HIMSS data, I show that hospitals do not switch providers very much throughout the sample and the HITECH period, suggesting a lock-in effect. In particular, they do not switch after they experience a data breach, suggesting a lack of sensitivity to security features.

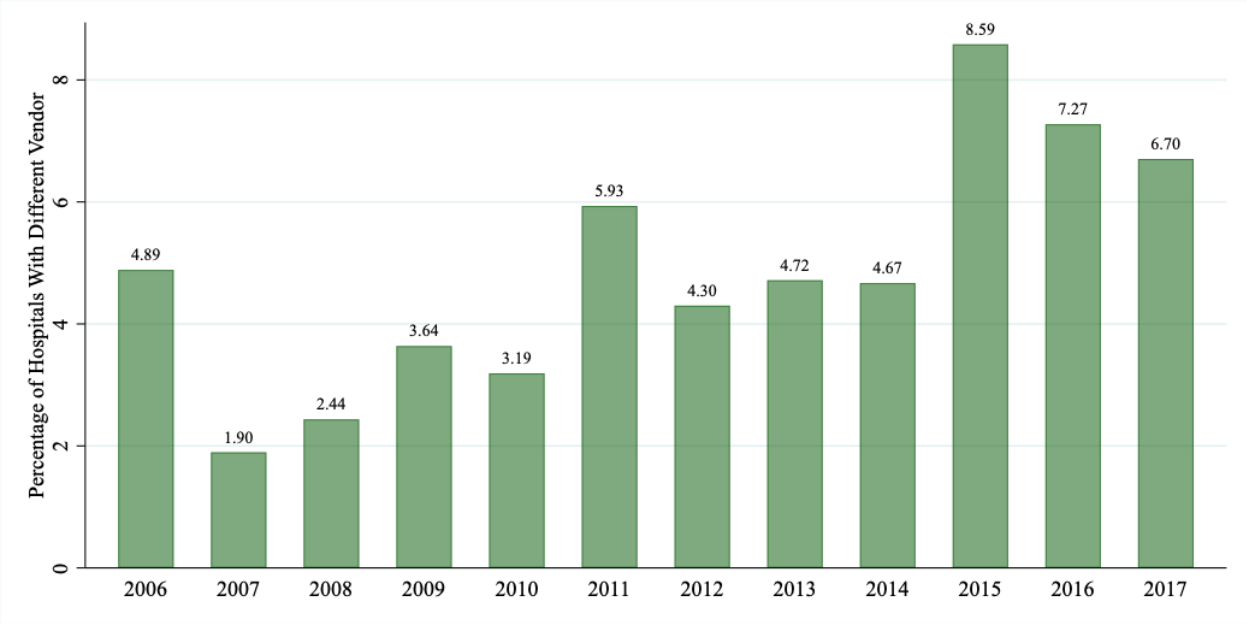
Figure 16 shows that no more than nine percent of hospitals switched EMR vendors in any given year in the HIMSS data, conditional on having adopted by the year in question (i.e., I do not count new adoptees as switchers).

4.5.2 How Hospitals Choose Their EMR Vendor

In separate analyses (see Marti (2024)), I investigate how hospitals select their EMR vendors. In that analysis, I find that nearly all of hospital choice of vendor can be explained by:

1. HITECH Act subsidies on the extensive margin (i.e. when a hospital chooses to contract with an EMR vendor)

Figure 16: HIMSS: Percentage of Hospitals Who Switch EMR Vendors Each Year



2. Basic hospital characteristics on the intensive margin (i.e. which among many vendors a hospital choosesn)
3. The identity of the current market leader on the intensive margin
4. Peer influence and the potential for positive network effects on the intensive margin.

I find that hospitals do *not* switch vendors often.

4.6 Matching Technologies and AHA Data

I use data on basic hospital characteristics from the American Hospital Association's annual survey. I match AHA hospitals to those in HIMSS primarily on the basis of Medicare Number, and in other cases after manual matching.[6]

Further, I also only focus on hospitals who report their technologies to the HIMSS Survey. The survey is voluntary, and hospitals may elect to only respond to part of it. I only include hospitals that report having at least one Vendor in both the EMR and the Security application categories. Although this does exclude hospitals that do not adopt either, note that separate estimates from the AHA data suggest 94% of hospitals had am EHR system by 2018, and the laggards are likely to be very small practices who do not accept many Medicare patients.

4.7 Who Experiences Data Breaches?

In this Section, I provide descriptive evidence showing that the types of hospitals who experience data breaches in the sample period differ in some fundamental ways from those that

[6]The authors of Gabriel et al. (2018) provided me with a crosswalk matching part of the AHA and HIMSS data, for which I thank them.

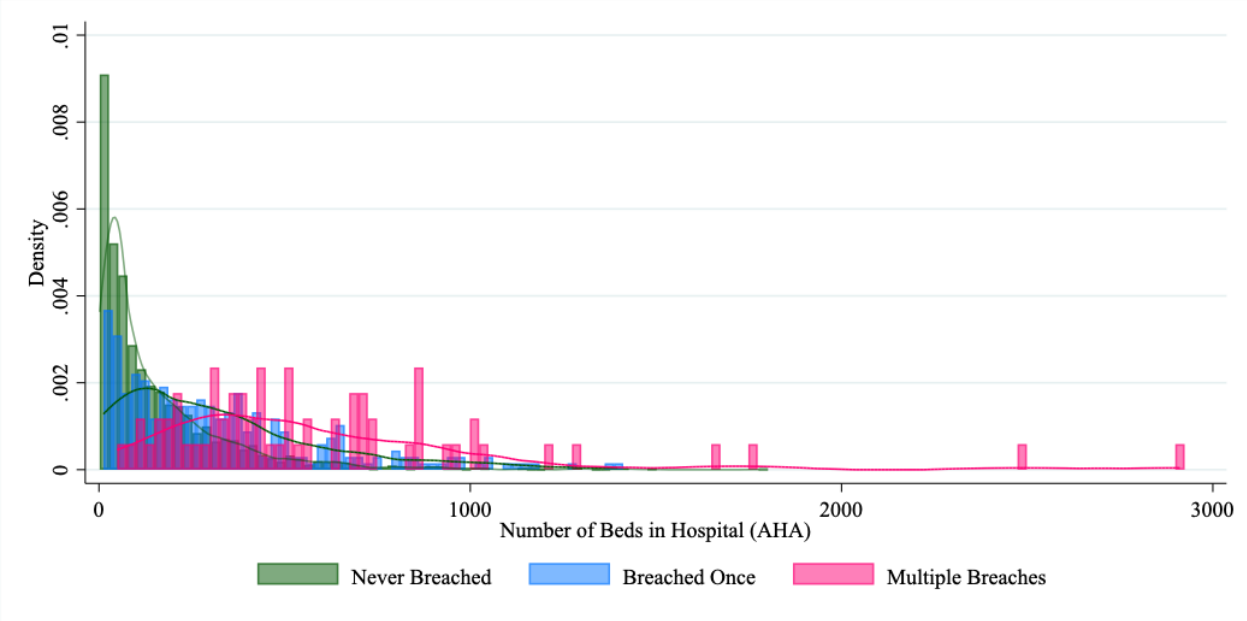
Table 2: Final Count of Hospitals in the HIMSS Technology and AHA Data

All HIMSS	All AHA	Both
5061	5136	4234

do not. Hospitals that experience data breaches are, in general, larger, more likely to be in metro areas, more likely to be Children's hospitals, and more digitized. I break down these differences into different types of breaches as well.

First, and perhaps most obviously, breached hospitals tend to be larger. One explanation is mechanical; breaches are only reported when they affect 500 or more records, which means larger hospitals are more likely to meet that threshold by virtue of having more records. Figure 17 shows the density of bed size split by hospitals that experience no breaches, just one breach, and multiple breaches.

Figure 17: HHS & AHA: Sizes of Breached Hospitals



The pattern holds as we move to what might be called more complex, or simply more scalable, attacks: cyberattacks and crime attacks each strike larger hospitals relative to physical attacks and mistake attacks respectively. Figure 18 shows the size of cyber vs. physically breached hospitals, and Figure 19 shows the same for crime vs. mistake breached hospitals, both with never-breached hospitals as a control group. Note some overlap, as a hospital may have experienced multiple types of breaches.

Across states and territories, the most breached (in terms of fractions of its hospitals breached) is the District of Columbia – perhaps attackers find records in the nation's capital particularly valuable (Figure 20).

Figure 18: HHS & AHA: Sizes of Breached Hospitals: Cyber vs. Physically Breached

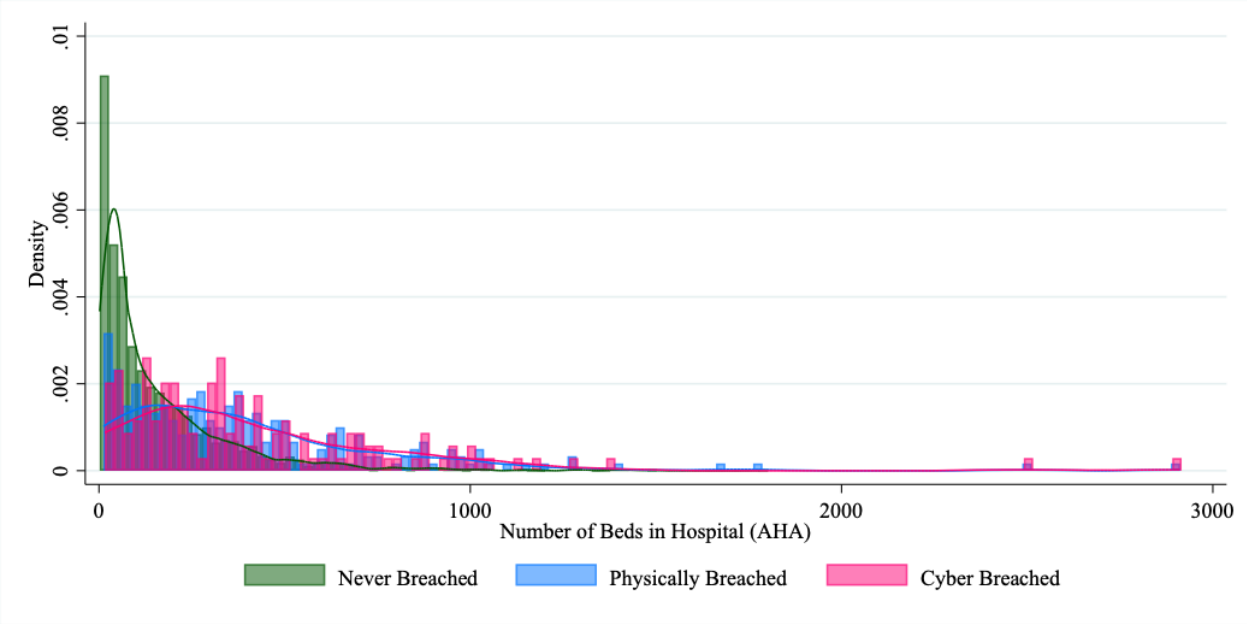
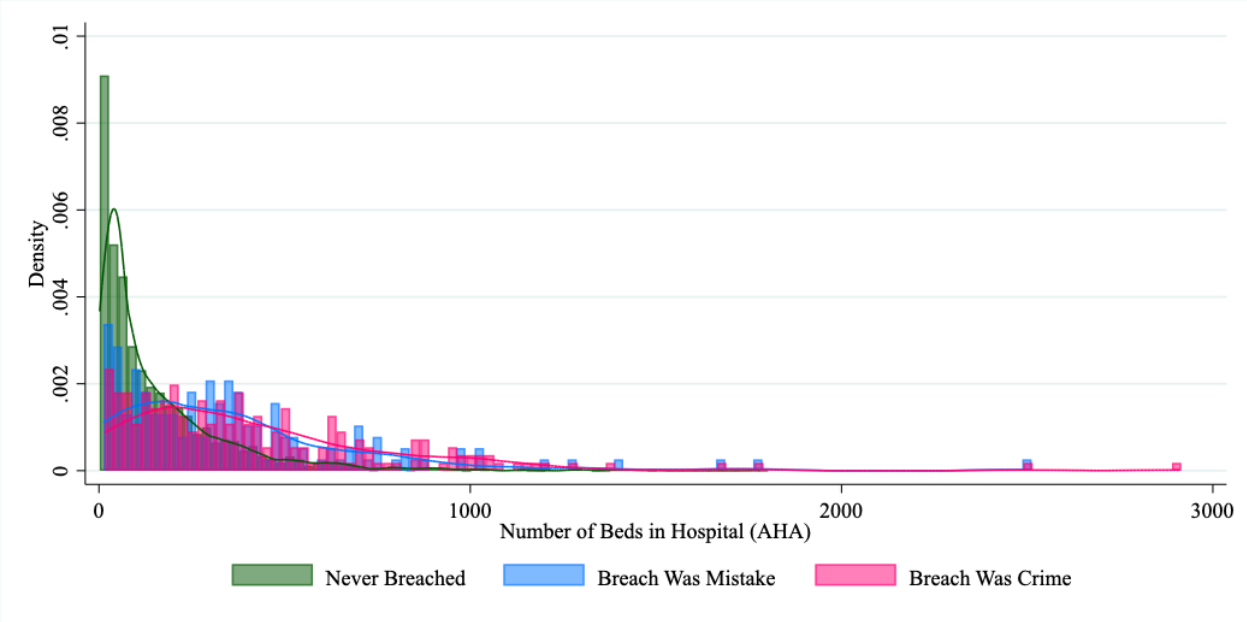


Figure 19: HHS & AHA: Sizes of Breached Hospitals: Crime vs. Mistake Breached



Children's hospitals are more likely to be breached than other types, perhaps due to the known greater value of children's data in secondary markets (Figure 21).

Government and not-for-profit hospitals are also more likely to experience breaches relative

Figure 20: HHS & AHA: States of Breached Hospitals

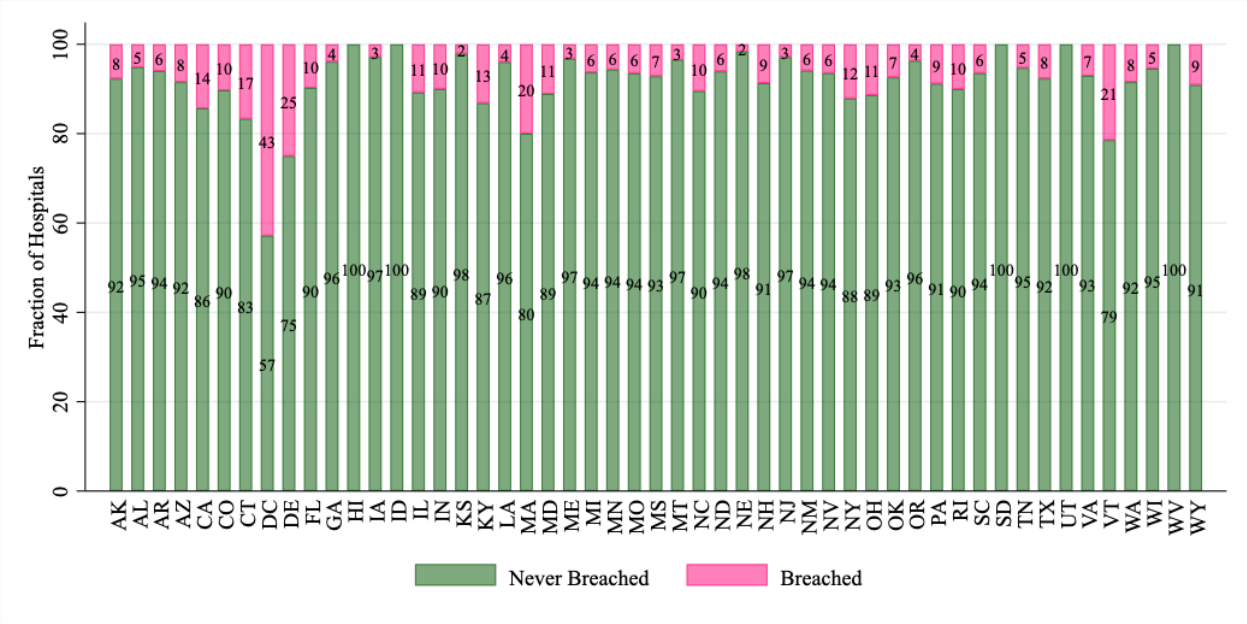
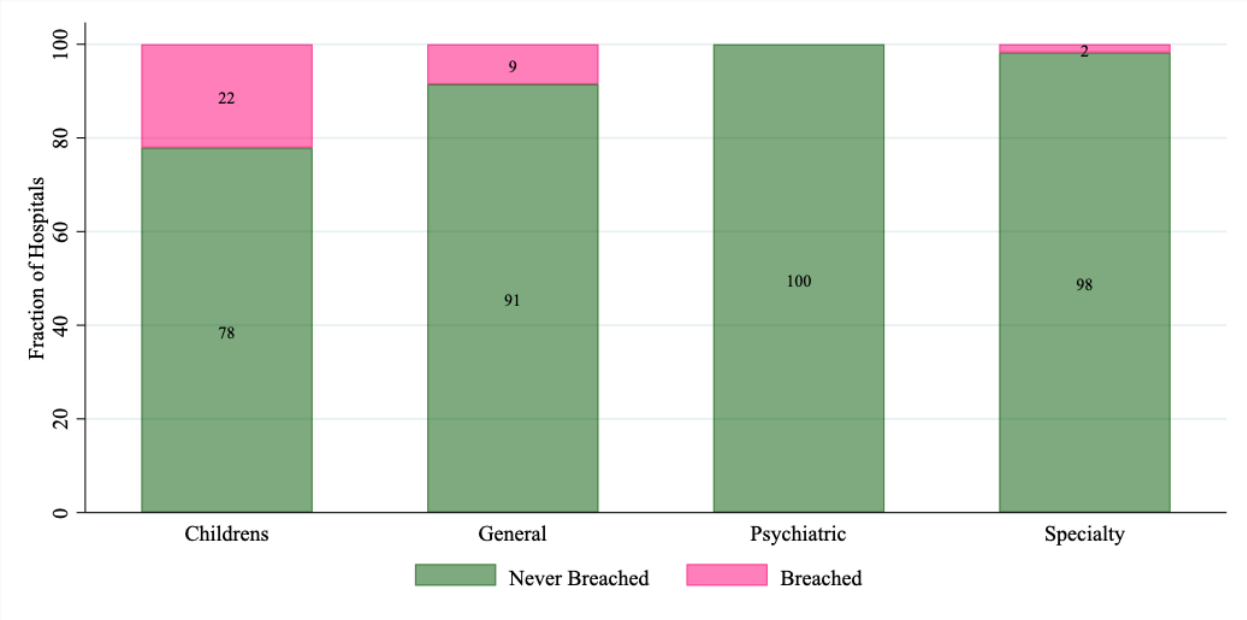


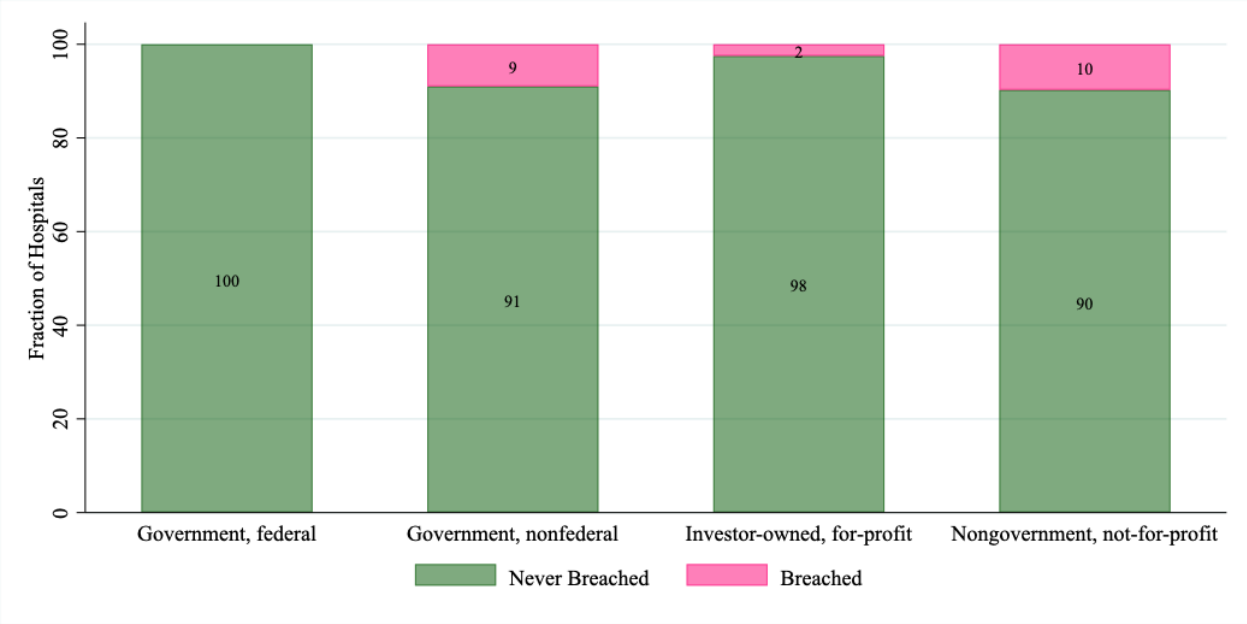
Figure 21: HHS & AHA: Service Type of Breached Hospitals



to their peers, suggesting for-profit hospitals' incentives around data protection may differ from those of other hospitals not necessarily motivated by shareholder returns (Figure 22).

Hospitals in Metro and Division CBSAs (i.e. urban) were more likely to be breached than

Figure 22: HHS & AHA: Control Type of Breached Hospitals



those in Micro and Rural areas (Figure 23).

Figure 23: HHS & AHA: Control Type of Breached Hospitals

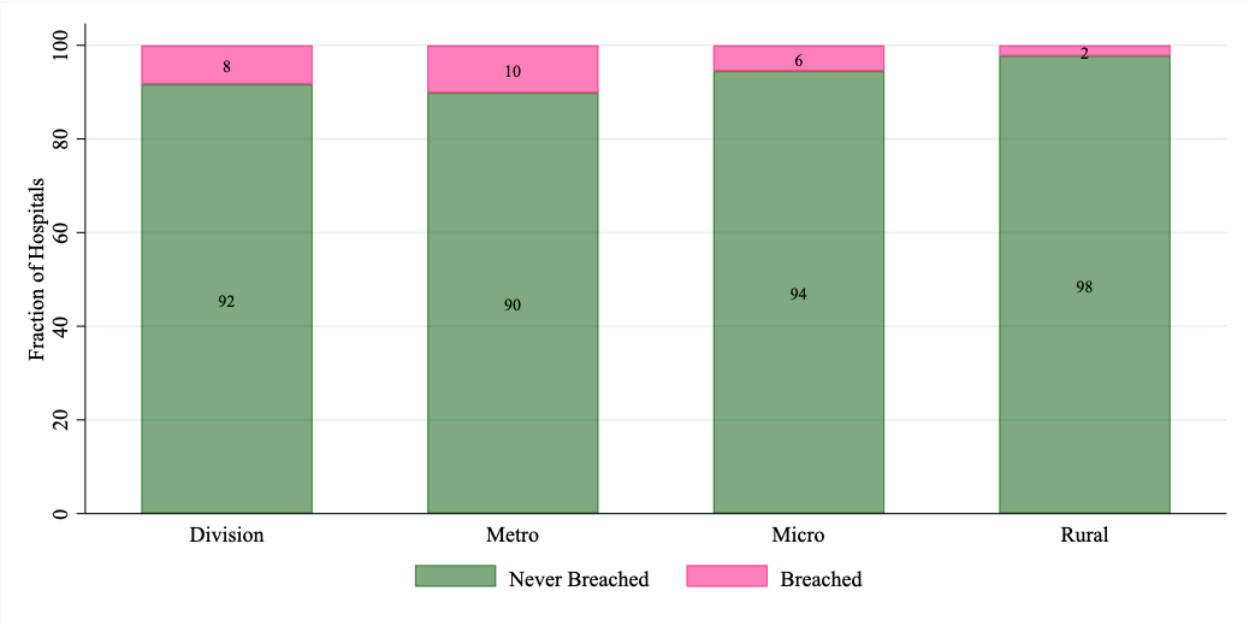
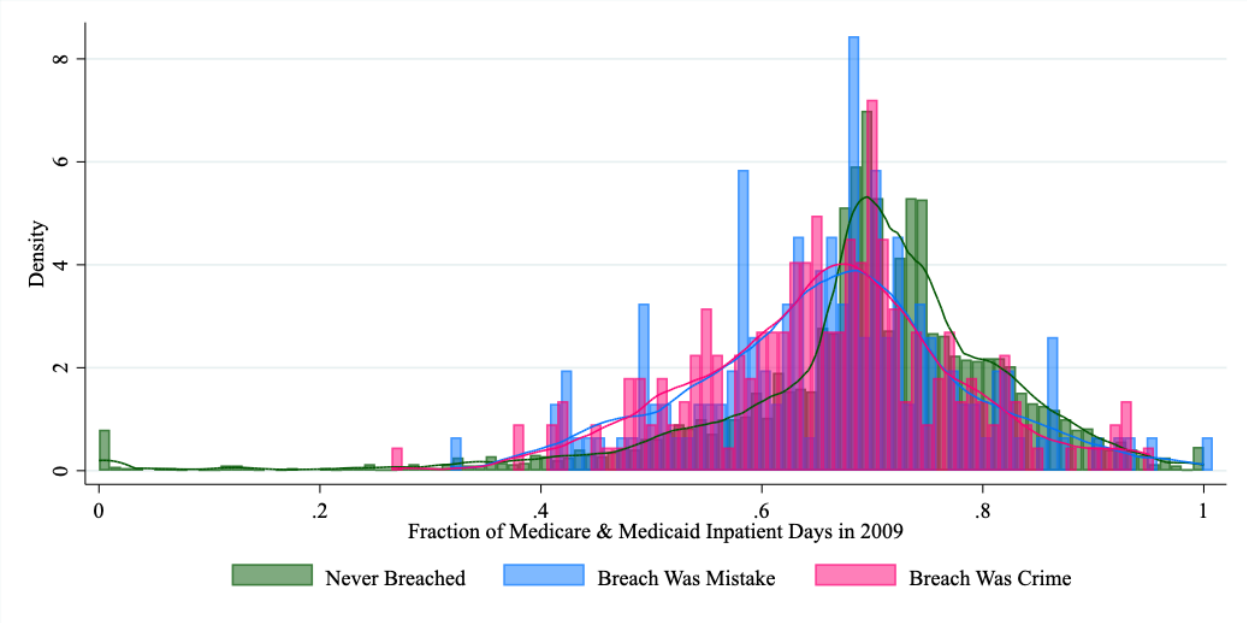


Figure 24 shows the density of the fraction of inpatient days for the hospital that were for

Medicare or Medicaid patients – i.e., the exact input the HITECH Act used when determining incentive payments. I plot the 2009 levels, before the HITECH Act was passed. I find that hospitals with higher fractions of Medicare & Medicaid inpatient days seem to be less likely to have had a breach – perhaps incentive payments allowed hospitals to upgrade systems or meaningful use criterion were implicitly effectively security polices. Alternatively, perhaps hospitals with low fractions of Medicare/Medicaid patients were more valuable to attackers thanks to their data on private insurers.

Figure 24: HHS & AHA: Fraction of Medicare & Medicaid



Finally, the only security information in the AHA data asks whether the hospital had MFA implemented. I use the 2009 level of MFA as a control for general security posture pre-HITECH Act. Figure 25 shows the breakdown of breached hospitals by whether they had MFA in 2009. As is a known issue in cybersecurity research, we appear to suffer from reverse causality here: hospitals that have implemented MFA are actually *more* likely to be breached later. The omitted variable is "concern about being breached" – hospitals that believe they might be breached are more likely to implement security tools like MFA – but also more likely to actually experience the breach. We will therefore need to be careful with causality of digitization and breach activity.

The result holds when we focus on cyber breaches only, though with the caveat that few hospitals are cyber breached in general.

4.7.1 Market Conditions and Data Breaches

This paper is concerned with how market structure shapes cybersecurity outcomes. In this Section, I provide descriptive evidence that hospitals in more concentrated markets – both in terms of the concentration of hospitals and of technology vendors – experience more breaches.

Figure 25: HHS & AHA: Had Multi-Factor Authentication in 2009 vs. Breached

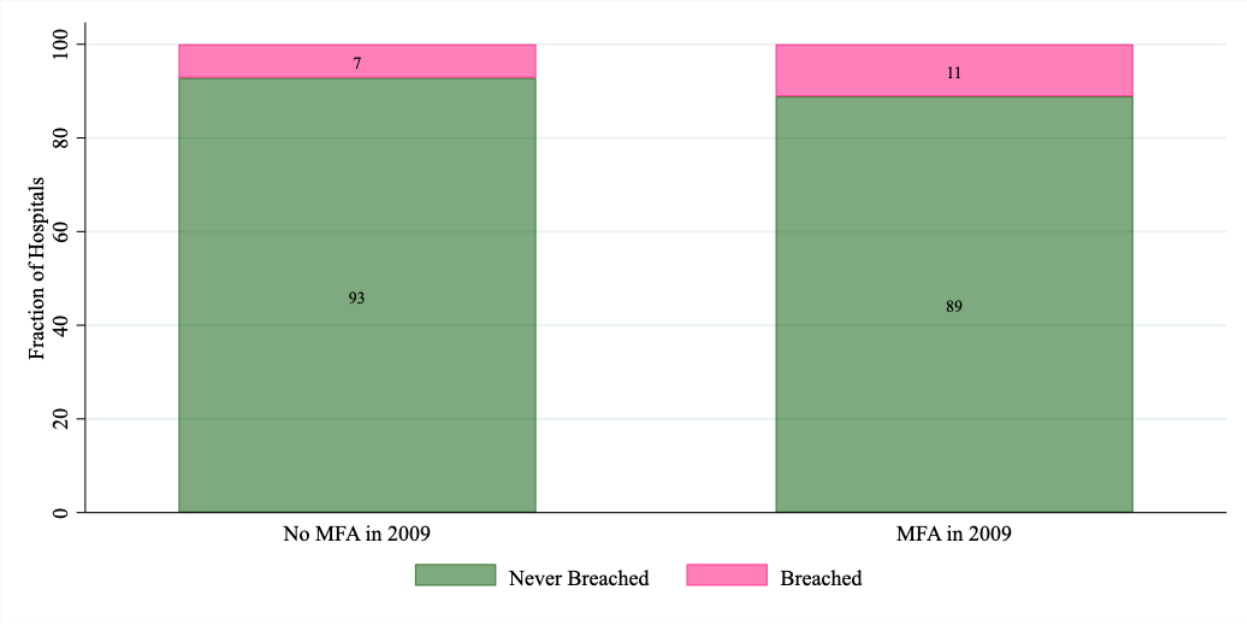
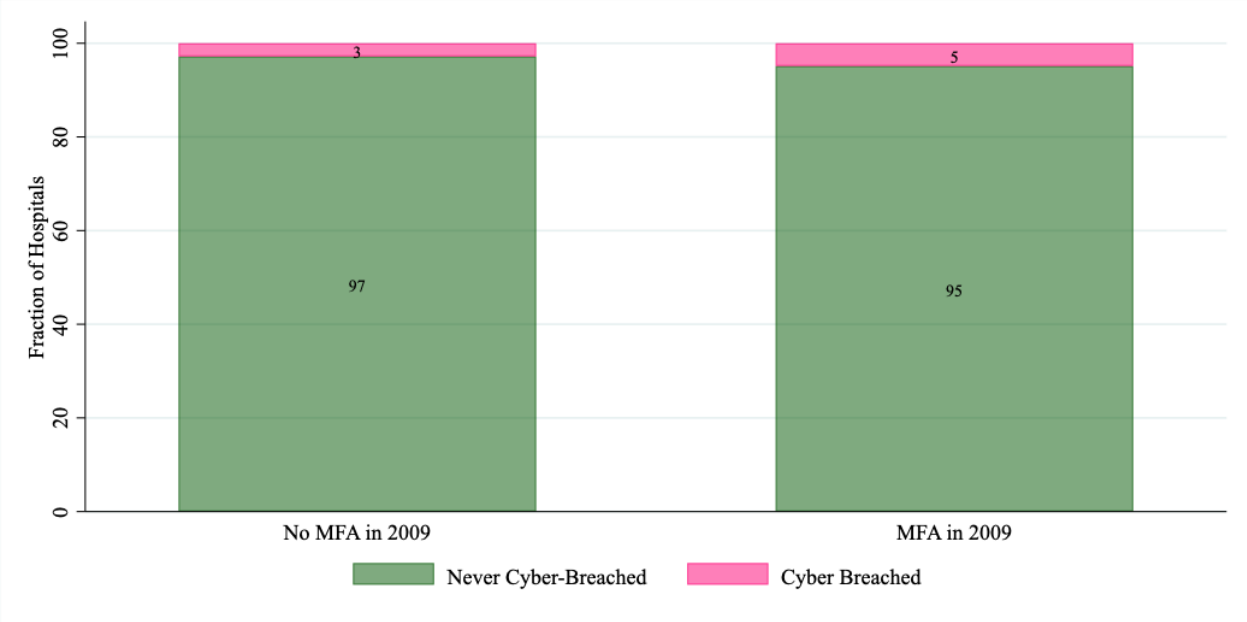


Figure 26: HHS & AHA: Had Multi-Factor Authentication in 2009 vs. Cyber Breached

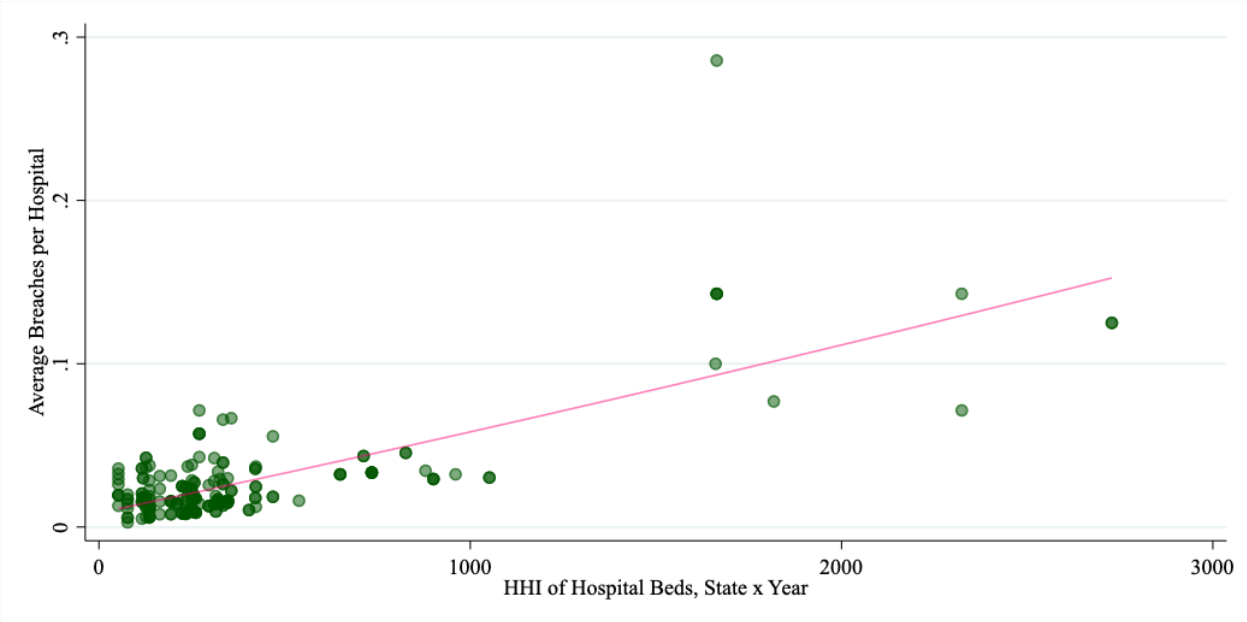


4.7.2 The Hospital Market: Beds

Under a Magnet Effect, if the object being targeted for vulnerabilities is the hospital, then a hospital that commands a large market share in its state should expect to see more breaches.

Figure 27 shows the correlation between the HHI of the hospital sector – where market share is the share of beds each hospital has in the state and year – and the average number of breaches experienced by hospitals in that state and year. The market here is the market for hospital beds, not yet technology. Here, we can clearly see that hospitals in more concentrated states – i.e. where patients have fewer options, and individual hospitals are larger – experience on average more breaches than in their less concentrated peers. The result mirrors that of Hydari et al. (2012) which found patient data was somewhat better protected in competitive hospital markets. However, we must control for hospital size – a driving factor of the attack in the first place to a strategic attacker.

Figure 27: HHS & AHA: HHI of Hospitals and Average Breaches



4.7.3 The Technology Market: EMR Vendors, State-Level

On the other hand, if the exploit comes not from the hospital but from the *technology* the hospital uses – that is, the relevant market is healthcare technology – we would expect to see a similar pattern for technology concentration as well.

Figure 29 shows the correlation between the HHI of EMR vendors within a state and year and the average number of breaches per hospital in that state and year. Here, we are concerned with concentration at the *technology* level – the vector through which an attacker can access a hospital's records, the upstream conduit, but not the final target. We see the same pattern: concentrated markets are associated with each individual hospital experiencing on average more breaches.

In Figure 30, the result holds even when we limit focus to just cyber breaches, again illustrative of the role a concentrated technology market can play in facilitating scaled attacks. A similar pattern holds in Figure 31, focusing on crime breaches only.

Figure 28: HHS & AHA: HHI of Hospitals and Average Cyber Breaches per Hospital

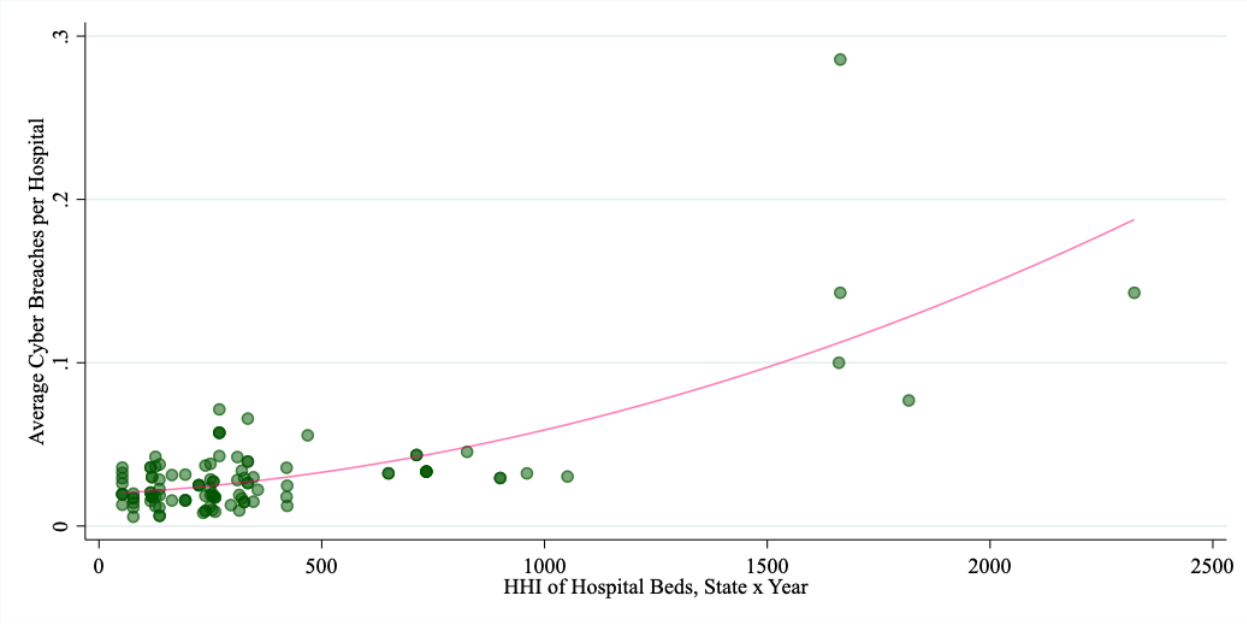
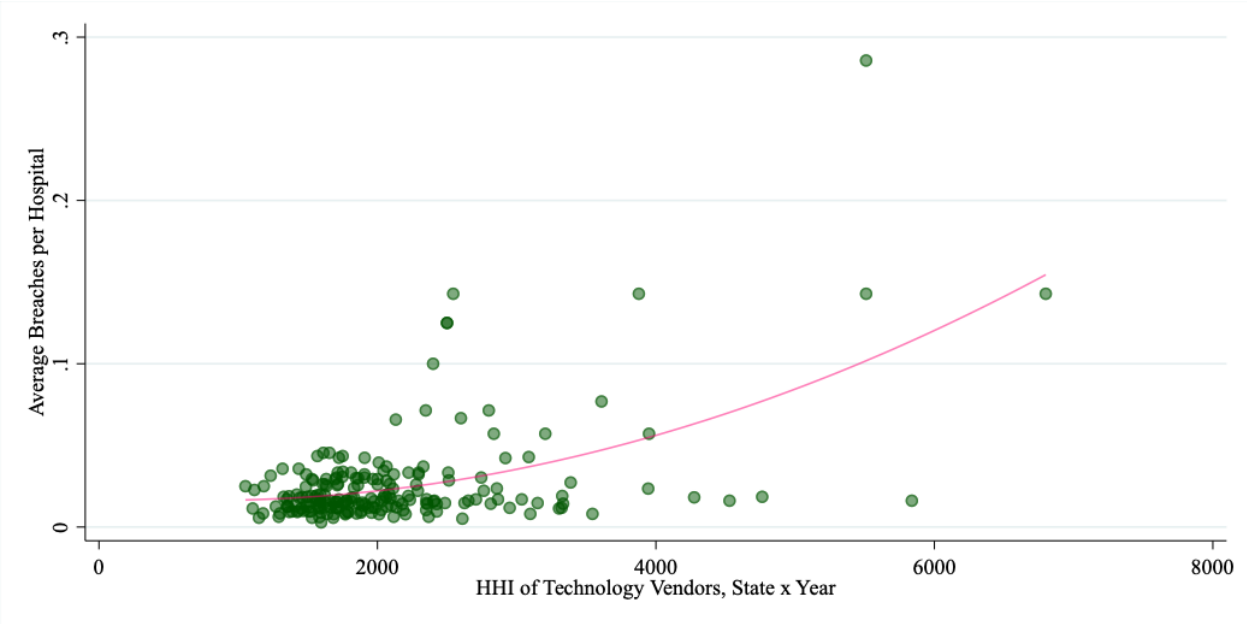


Figure 29: HHS, HIMSS, & AHA: HHI of EMR Vendors and Average Breaches per Hospital



4.7.4 The Collective Buyers Market: EMR Vendors, GPO-Level

I next look at vendor concentration within the GPO to which a hospital belongs. The GPO generally defines the *actual* set of choices a hospital has for its technology vendor contracts,

Figure 30: HHS, HIMSS, & AHA: HHI of EMR Vendors and Average Cyber Breaches per Hospital

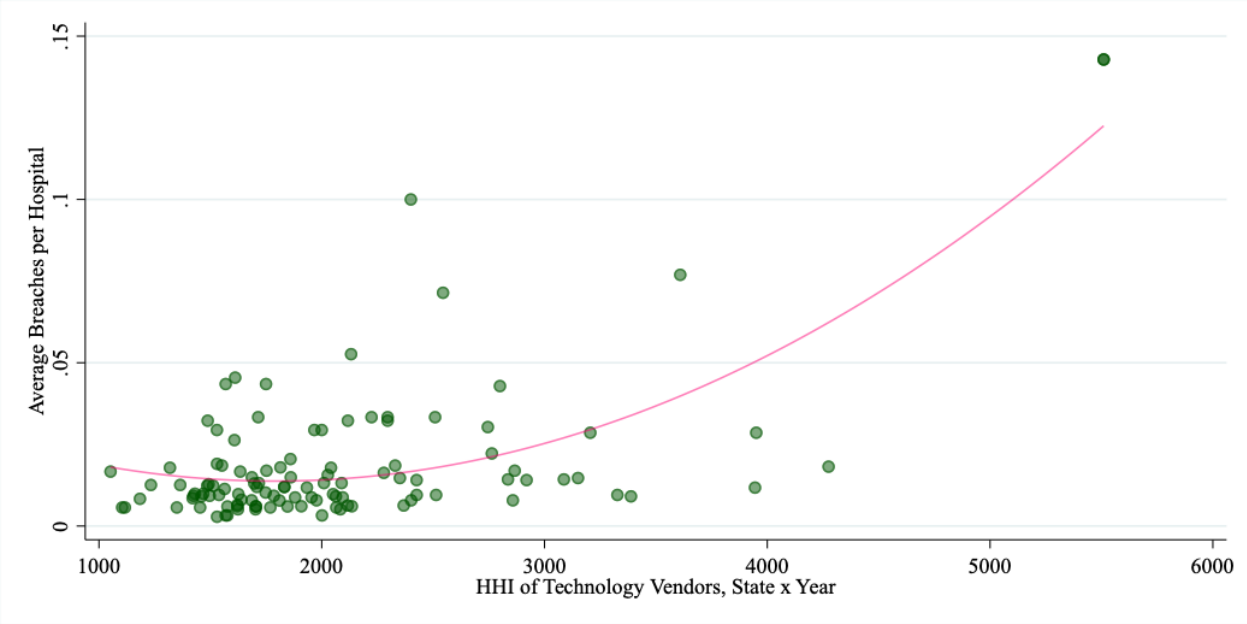


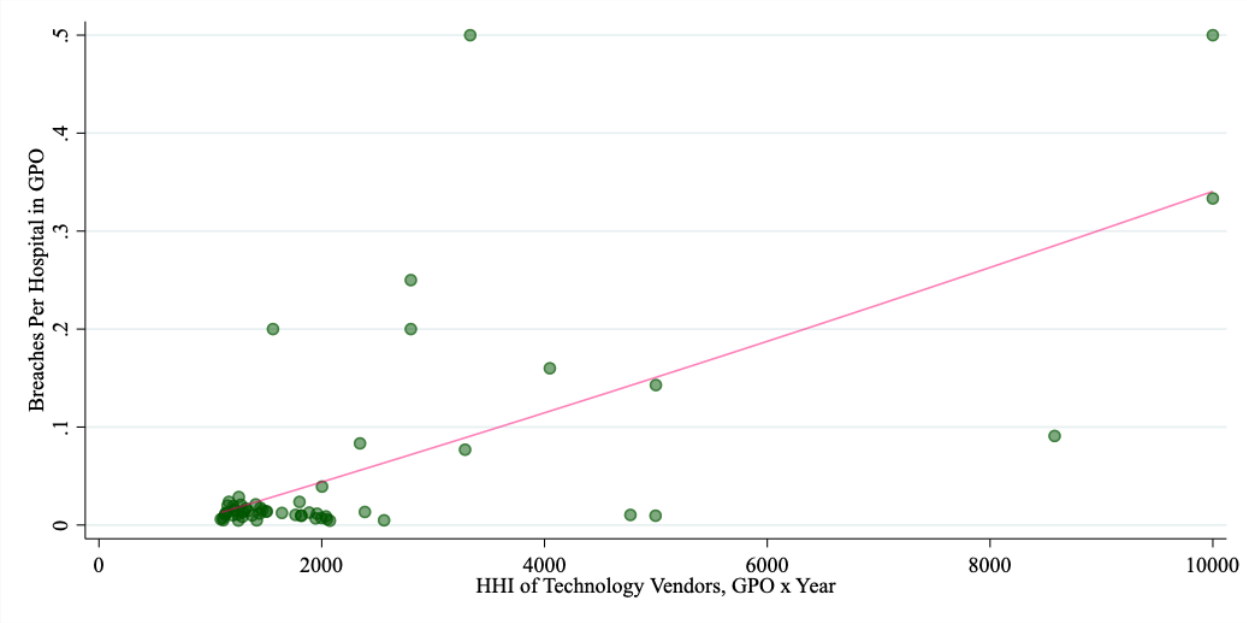
Figure 31: HHS, HIMSS, & AHA: HHI of EMR Vendors and Average Crime Breaches per Hospital

and therefore may provide a better sense of the market for technology vendors faced by an individual hospital. The GPO-level HHI here measures how concentrated, within a GPO, hospitals' choices of vendors are.

We once again see the same result in Figures 32 and 33: hospitals that are in GPOs with fewer vendor choices and more concentration are more likely to experience breaches, especially cyber breaches.

Taken together, the figures in Section 4.7.1 suggest that in environments where hospitals have fewer choices and are more concentrated in their technology vendors, or environments where patients have little hospital choice and large hospitals dominate, more breaches occur.

Figure 32: HHS, HIMSS, & AHA: HHI of EMR Vendors within a GPO and Average Breaches per Hospital



As a test, I check if physical breaches show the same pattern. I find that physical breaches do not seem to particularly correlate with the HHI of the hospital (Figure 34), suggesting something about the method of attack may be relevant to the market concentration correlation observed.

4.8 Methodology: The Duration Model, Time to Breach

In this Section, I empirically investigate how the digitization process has affected cybersecurity outcomes in the healthcare sector. I do so by breaking down the analysis into two margins: (i) Extensive Margin (ii) Intensive Margin. Within the Intensive Margin, I conduct tests that estimate the importance of each of the four externalities described in the model from Section 3:

1. **Floodgates Effect**: Attacks scale, meaning as long as the technology provider is the same, the attacker has a near-zero marginal cost of attacking each additional hospital.

Figure 33: HHS, HIMSS, & AHA: HHI of EMR Vendors within a GPO and Average Cyber Breaches per Hospital

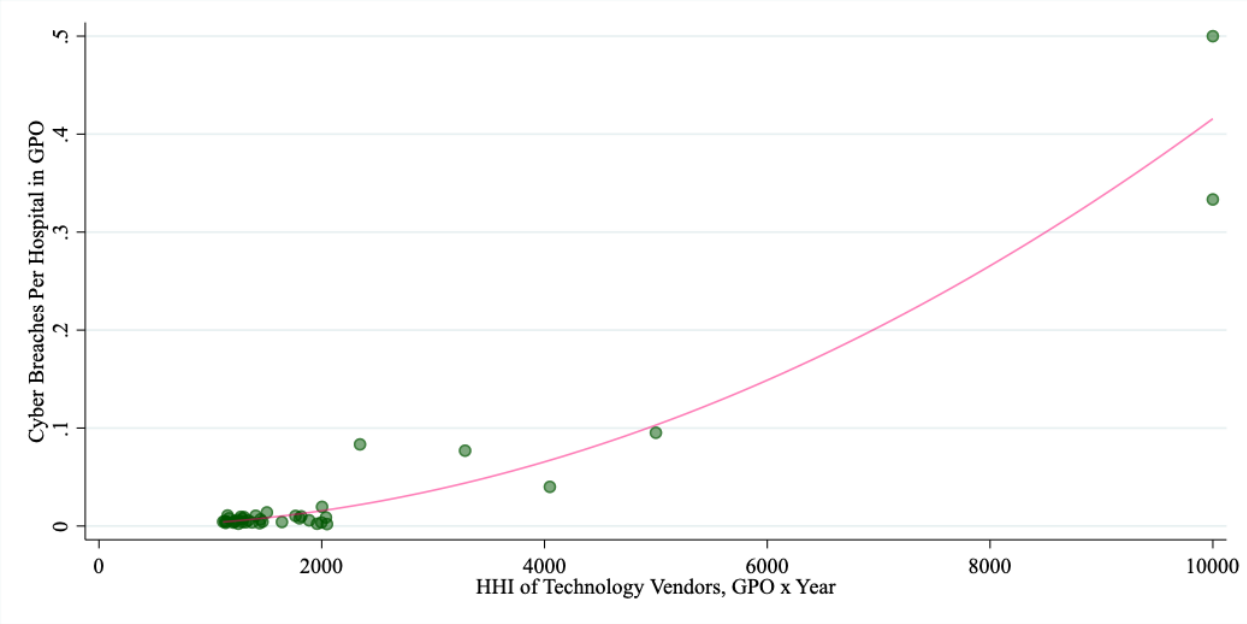
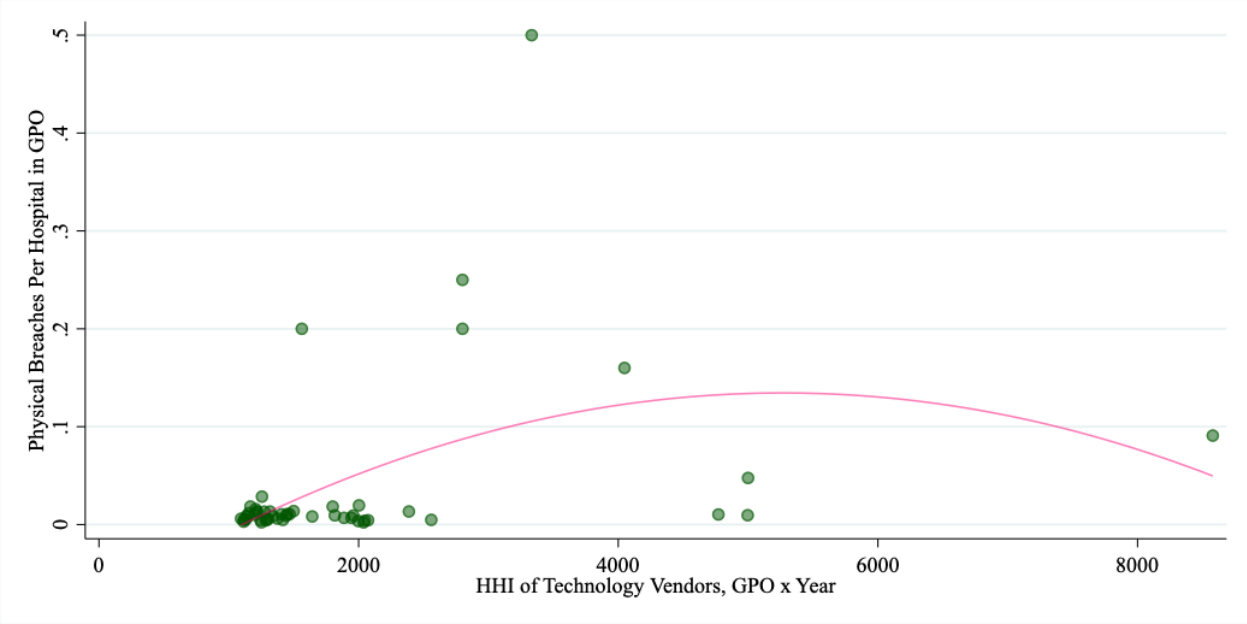


Figure 34: HHS, HIMSS, & AHA: HHI of EMR Vendors within a GPO and Average Physical Breaches per Hospital



Therefore, once a single hospital is breached and the exploit has been developed, all other hospitals who have the same vulnerability are at increased breach risk. We therefore would expect to see many hospitals with the technology vulnerability in common all breached at once.

- **Test**: once one hospital has experienced a breach, are its peers all more likely to experience breaches as well?

2. **Gatekeeper Effect**: As the counter to the floodgates effect, additional investment in security by the technology vendor would keep all of its users safe. Software developers may experience economies of scale in security investments, able to keep all their users safe with one investment, while in a competitive market each developer has to independently invest in costly security.

   - **Test**: are larger (i.e. more popular) vendors better able to protect their hospitals than others? Who invests in security?

3. **Magnet Effect**: Attackers are strategic, and therefore will choose to invest in developing an exploit for software that has a higher expected benefit than cost. A technology vendor that serves many hospitals therefore is an attractive target, assuming there is a low-to-zero marginal cost of deploying the same attack on multiple targets when the attacker only has to develop an exploit once. Any hospital that uses the same technology provider as others creates a *negative externality* on all other users by increasing the expected value of the attack.

   - **Test**: Are larger vendors more likely to be associated with breaches than smaller vendors?

4. **Sum-of-Efforts Effect**: On the other hand, I find that as long as individual hospitals are able to *add* their own security on top of what the technology provider has, each hospitals' investment can act as a *deterrant* to the attacker. That is, individual security reduces the overall expected value to the attacker of exploiting the technology in the first place. Therefore, each additional security choice by an individual hospital has a *positive externality* on all other users by decreasing the expected value of the attack.

   - **Test**: if hospitals implement security technologies, do they better protect not just themselves, but their peers as well?

My primary methodology for evaluating the role various technologies play in cyberattacks is the *duration model* (Van Den Berg (2001)). Here, we seek to answer the question, *"What influences how long a hospital can go without having a data breach?"*

Implicit in that question is the presumption that hospitals will *inevitably* experience a data breach; their inherent characteristics, their technology choices, and other covariates I explore in this paper all will influence the *when* of the data breach. Good choices and characteristics make data breaches more "rare" and therefore hospitals can go longer without them; bad choices and characteristics make them more common and possible earlier.

Let t_h be the amount of time from the start of the world that passes before hospital h experiences any data breach. I assume the "start of the world" is 2010, the first year the HITECH Act mandated data breach disclosure. Of course, hospitals may have experienced breaches before 2010, and indeed that may affect their probabilities of future breaches. I use both a multiple-breaches model to account for hospitals experiencing more than one breach during the sample period, and I also assume that the state of a hospital in 2010 captures any differences between hospitals that were and weren't breached before then.

The data are *right-censored*, meaning there exist hospitals in the sample that do not experience *any* data breaches, though they may in the future. The data exist as a "flow," meaning we observe hospitals over time, and can identify the years in which they experienced data breaches (rather than looking at just the breached hospitals in their year of breach, or a single cross-subsectional year).

We are interested in the distribution of the time-to-breach, t_h across hospitals. Let $F(t) = \Pr(T < t)$, the probability the hospital "survives" without a breached only until some t. Conversely, $S(t) = 1 - F(t)$ is the probability the hospital is breached within t. Then, $h(t) = \frac{f(t)}{S(t)}$ is the *hazard ratio*, the probability the hospital is breached exactly at time t, conditional on having survived until t. It must depend both on the external conditions at time t and on the characteristics of the hospital that let it survive this long without a breach (Van Den Berg (2001)). It is the latter fact – that hospitals that do not experience breaches are somehow different from those that do – that makes the hazard ratio useful here.

4.8.1 Cyber vs. Physical Breaches

One would expect a priori that after records are digitized, there would be more cyberattacks and fewer physical attacks. On the other hand, physical attacks may simply hold steady while cyberattacks increase, leading to an overall increase in breaches. I break down the extensive margin analysis into total, cyber, and physical breaches to determine if digitization itself is correlated with more cyberattacks, or if other factors simply mean more breaches in general.

4.8.2 Mistake Data Breaches

Let us assume that the first category of possible data breaches are *mistakes*. Hospitals can be prone to or mitigate the occurrence of mistakes (e.g. papers that scatter in a parking lot, Section 4.2.3). However, they do not necessarily depend on the occurrence of mistakes at *other* hospitals.

Suppose each hospital has its own individual rate of mistakes, $\theta_h(t)$. For exposition purposes but not in the final analysis, let's also assume that the rate of mistakes does not change over time for a hospital, $\theta_h(t) = \theta_h$. Then, the probability that a mistake has occurred by time t is $F(t) = 1 - \exp(-\theta)$, which results in a hazard rate of exactly θ_h. Therefore, the object we are interested in is itself the hospital-specific hazard rate.

In the analysis, $\theta_h(t)$ will depend parametrically on hospital characteristics and choices, like its size (constant over time) and its choice of technology systems (time-varying). In particular, I will assume those covariates enter the hazard function proportionally, leading us to a Cox Proportional Hazards model (Cox (1972)):

$$\theta_h(t) = \lambda(t) \exp(x_{h,t}\beta_{h,t})$$

meaning the baseline hazard rate $\lambda(t)$ is common across hospitals, but the actual in practice rate for a hospital will depend on factors $x_{h,t}$ that vary jointly across hospitals and time. It is the coefficients on these covariates, $\beta_{h,t}$, that I aim to estimate.

I will assume $\lambda(t) = 1$, meaning it is constant over time, neither generally increasing or decreasing; including a constant in the list of covariates $x_{h,t}$ will capture the shifts over time or across hospitals.

4.8.3 Crime Data Breaches

In this section, I provide a simple structure to ground the concepts driving the results of the duration model.

Let us suppose there exists an attacker who makes a hospital- and time-specific investment in attacking an institution at some particular time α_{ht}. The investment may depend on the technology choice of the hospital (e.g. if the attacker has been able to breach that technology before) and the overall stock of attackers (high during the pandemic, for example), and other factors that vary both by hospital and time. Similarly, a hospital makes an investment in securing itself from crimes, s_{ht}. Let's assume they interact in a Tullock-style contest, i.e. independently and simultaneously, so that the overall probability of attack becomes:

$$\Pr_{ht}(\text{Successful Attack}) = \frac{\alpha_{ht}}{\alpha_{ht} + s_{ht}}$$

The hazard rate, as described in Section 4.8.2, turns into exactly the probability of experiencing a successful attack.

However, the key difference is that while s_{ht} captures the [effective] security investment of the hospital – consisting of their technology choices as well as invariant characteristics such as size and type that will affect the efficacy of such choices – we also have the component αht, which is a separate object not under the hospital's control. Furthermore, the attacker's effective investment αht depends on the investments of *every other* hospital in the system. Therefore, as every other hospital in the system gets more secure, the attacker will (a) shift investment towards hospitals that have less security and (b) increase overall investment if they are not budget constrained (c) target technologies that have the highest promised returns, i.e. those used by the largest-weakest hospitals.

Overall, there is an implicit *dependence* in α_{ht} of the hospital's attack threat level on *every other hospital in the system* that was not present in the case of mistakes. We therefore *expect* that the differentiating factor between a hospital that experiences a *crime* data breach and a *mistake* data breach should be *precisely* the possibility of network effects brought upon by the *strategic and scaling* attacker that do not appear in the mistake breaches.

In the case of EHRs, for example, we expect that the probability of experiencing a crime and cyber-based data breach may depend on not just one's own digitization but also the *digitization of one's peers*: as more hospitals begin to use EHR systems, the overall gain to an attacker of *learning how to breach an EHR system* increases, leading to increases in α_{ht} as described above.

4.8.4 Identification

The identification assumption here is that there is no reverse causality, directly or indirectly, in the breach outcome on digitization behavior of the hospitals. That is, unobserved factors are not influencing both the digitization behavior and data breaches. The assumption is strong. Other papers instrument digitization behavior with variables that are supposed to be unrelated to data breaches, such as average digitization in the state (Kim and Kwon (2019)). However, my exact argument in Section 4.8.3 is that the occurrence of breaches at other hospitals may in fact be *directly* related to breaches at one's own hospital, as a strategic attacker can take advantage of common technologies to run a scaled attack.

I instead rely on the evidence of [redcated] which showed that hospitals' choice to adopt EHR systems is primarily driven by the perceived costs and benefits rather than security concerns. These costs and benefits change over time thanks to the HITECH subsidies (Dranove

et al. (2014b), Adler-Milstein and Jha (2017)), might vary by location (Dranove et al. (2014a)), or basic hospital characteristics (Adler-Milstein et al. (2015)). Further, as I show in separate analyses hospitals do not seem to respond in measurable ways to the experience of a breach (either their own or in their state), suggesting some role of insurance and/or inevitability (Marti (2024)).

I also control for other factors that may influence the tech-savvy of a hospital (and therefore their data breach outcomes) while being orthogonal to the EHR and security decisions. Following Kim and Kwon (2019) I include the use of technologies unrelated to the storage of records as control variables. I use the following:

1. **Health Information Exchange (HIE)** Per Choi et al. (2023), joining an HIE is possibly related to data breach risk. By 2017 about 68% of sample hospitals report being in a HIE.
2. **Electronic Data Interchange Clearing House (EDI)** In 2024, the massive Change Healthcare ransomware attack brought down the major clearinghouse that serves, as of this writing, about a third of the healthcare market and therefore about *five percent of U.S. GDP*. EDIs were common targets for ransomware attacks even before the Change attack. I include whether or not the hospital uses an EDI. In 2017 83% of sample hospitals contract with an EDI.
3. **Radiology Information Systems (RIS)** similar to the Cardiology Information system also used in bluecitekimkwon, a type of information system not directly governed by HITECH initiatives and therefore another indicator of how an implementing hospital may recognize a higher benefit to healthcare IT implementation. I use the Radiology Information System instead of the Cardiology because of a quirk in the data: there are two places where hospitals are asked about their RIS (in Applications and under Service Delivery), so we are likely to get responses from more hospitals.[7] By 2017, 89% of hospitals in the final sample have a radiology information system.

4.9 Results: Extensive Margin

In this Section, I present results on the extensive margin for how adopting a Basic EMR, and then an Advanced EMR, along with other security technologies, impacts a hospital's breach risk. The coefficient on each acts multiplicatively on the overall hazard rate of experiencing a breach in the Cox Proportional Hazards model. Therefore, coefficients greater than one increase risk, while coefficients lower than one will decrease it.

As would be expected if digitization enables attackers to access records, adopting Basic EMR in point estimate is associated with a small increase in the breach hazard rate, while advanced is even lower (Table 3). Spam filters help somewhat, while single sign-on software might actually facilitate attacks. Coefficients on Firewalls and Encryption are not statistically significant, especially when we add in EMR technologies and basic hospital characteristics (all the ones discussed in Section 4.2.2). I present the size coefficients for exposition and to pin down orders of magnitude: the largest hospitals have three times the hazard rate of the smallest hospitals in the sample.

[7] Hospitals may report their Vendors in both surveys, or they may – due to error – report in only one.

Table 3: Hazard Ratios, Breach: Extensive Margin

	Breach		Breach	
Basic EMR			1.235	(0.577)
Advanced EMR			1.016	(0.918)
Firewall	1.027	(0.893)	1.023	(0.909)
Spam/Spyware Filter	0.708*	(0.092)	0.708*	(0.091)
Encryption	1.243	(0.202)	1.239	(0.210)
Single Sign-On	1.218	(0.107)	1.215	(0.111)
HIE	0.894	(0.361)	0.891	(0.350)
EDI	1.082	(0.689)	1.076	(0.710)
RIS	1.012	(0.977)	0.926	(0.864)
Medium	1.177	(0.485)	1.171	(0.498)
Large	3.015***	(0.000)	2.995***	(0.000)
Has MFA in 2009	1.057	(0.635)	1.054	(0.652)
Pseudo-R-Squared	0.06		0.06	
Hospital Count	4232		4232	
Breach Type Count	381		381	

Results gain meaning when we consider not just all breaches as a monolith but break them down into the *type* of breach (Table 4). Here, most notably, implementing a Basic EMR is associated with a *six times higher* cyber breach hazard rate, while the physical breach rate is actually much lower. The results confirm the hypothesis that digitization did not just change hospital procedures: it gave attackers new avenues through which they could access records and conduct breaches. A lack of paper records might then lead to, in some compensation, fewer physical breaches. The pattern does not necessarily hold for crimes and mistakes (Table 5), as no coefficients are statistically significant. The failure to rule out differences between crimes and mistakes suggests the extensive margin does not exactly impact third-party breaches differently from internal ones.

Table 4: Hazard Ratios, Each Breach Type: Extensive Margin

	Breach		Cyber		Physical	
Basic EMR	1.235	(0.577)	6.770*	(0.087)	0.803	(0.575)
Advanced EMR	1.016	(0.918)	1.123	(0.622)	1.007	(0.971)
Firewall	1.023	(0.909)	1.558	(0.140)	0.859	(0.506)
Spam/Spyware Filter	0.708*	(0.091)	0.541**	(0.031)	0.766	(0.260)
Encryption	1.239	(0.210)	1.229	(0.426)	1.295	(0.204)
Single Sign-On	1.215	(0.111)	1.179	(0.427)	1.230	(0.158)
HIE	0.891	(0.350)	1.044	(0.846)	0.904	(0.494)
EDI	1.076	(0.710)	0.993	(0.981)	1.186	(0.481)
RIS	0.926	(0.864)	0.364*	(0.060)	1.870	(0.339)
Medium	1.171	(0.498)	1.351	(0.471)	1.127	(0.667)
Large	2.995***	(0.000)	3.639***	(0.001)	2.903***	(0.000)
Has MFA in 2009	1.054	(0.652)	1.061	(0.737)	1.046	(0.758)
Pseudo-R-Squared	0.06		0.10		0.06	
Hospital Count	4232		4232		4232	
Breach Type Count	381		138		255	

Table 5: Hazard Ratios, Each Breach Type: Extensive Margin

	Breach		Crime		Mistake	
Basic EMR	1.235	(0.577)	1.713	(0.249)	0.707	(0.480)
Advanced EMR	1.016	(0.918)	1.088	(0.645)	1.004	(0.987)
Firewall	1.023	(0.909)	1.012	(0.962)	1.076	(0.798)
Spam/Spyware Filter	0.708*	(0.091)	0.602*	(0.063)	0.870	(0.588)
Encryption	1.239	(0.210)	1.560**	(0.042)	0.840	(0.474)
Single Sign-On	1.215	(0.111)	1.154	(0.355)	1.345	(0.111)
HIE	0.891	(0.350)	0.809	(0.172)	1.211	(0.364)
EDI	1.076	(0.710)	1.200	(0.464)	0.943	(0.839)
RIS	0.926	(0.864)	0.620	(0.301)	1.590	(0.559)
Medium	1.171	(0.498)	1.226	(0.503)	1.167	(0.658)
Large	2.995***	(0.000)	3.669***	(0.000)	2.547***	(0.002)
Has MFA in 2009	1.054	(0.652)	0.992	(0.955)	1.140	(0.476)
Pseudo-R-Squared	0.06		0.08		0.06	
Hospital Count	4232		4232		4232	
Breach Type Count	381		239		155	

4.10 Results: The Market Effects

Next, I look at how the choice of technology vendors and the resulting market structure affect cybersecurity outcomes, testing the hypotheses of the theoretical model.

I seek to understand if certain technologies and market structures are actually hastening breaches for U.S. hospitals.

Of course, the question of measuring the security of a software is extremely difficult, since we do not know which attacks were attempted and thwarted by *good* software; we only observe attacks in the HHS data that were *actually* successful. Furthermore, the HHS data do not specifically *assign fault* to a particular technology in their public reports – as described in Section 4.2.2. In the terminology of Section 3, we are not able to observe β_i directly, the security posture of the technology service provider. If we could, we would directly be able to compare numbers and externalities.

Instead, I combine various datasets that each offer us one piece of the puzzle and use econometric analysis to draw revelations about the security posture of technology vendors. The HHS data show us when breaches occur in hospitals. The AHA data provide basic hospital characteristics. The HIMSS data provide the technology and security posture of the hospital over time. We do not, as is almost always the case, have specific data on the vulnerabilities in technology nor whether the hospital was exposed to a vulnerability (see Murciano-Goroff et al. (2024) for a rare case with specific vulnerability data, but at the cost of not having any associated breach data).

In this Section, I take a backwards-induction-style approach, looking at breach *outcomes* to infer something about breach *sources*. That is, we can use information about how an attack spreads to infer whether an exploitable technology might have been involved. If the attack is a mistake, or takes place physically, it occurs outside the technology choices of the hospital. For cyber-crimes, however, looking at the technologies used in the same locations where an attack occurred gives us information about the security of those technologies. If a lot of bikes using the same bike lock are stolen, one might consider investigating the safety of that lock.

On the other hand, one might expect hospitals themselves may be making choices about which EMR vendor to contract with as a function of their own concern about security, along with follow-on choices about which additional security technologies to implement. However, as I showed in Section 3, it appears hospitals are choosing their EMR vendor based on factors near-exogenous to their *individual* security concerns, such as who offers a certified EHR system, the HITECH subsidy levels, the dominant contracts in their Group Purchasing Organizations, their state leaders, the possibility of *positive* network effects at the state level, and simple stickiness to their past vendor. In addition, Marti (2024) shows that hospitals do not necessarily increase their switching behavior for their EMR vendor – or any other vendor – after experiencing a breach. I therefore in this Section treat the specific choice of EMR vendor as near-exogenous to the product's advertised security.

I run a Cox Proportional Hazards Model, as in Section 4.8, including as explanatory variables the specific choice of EMR Vendor used by the hospital over time. Note that hospitals may change their EMR vendors over time, though not often, so the variable is time-varying. The choice of EMR vendor therefore acts multiplicatively on the hazard rate of a breach in any given year. As a "binary" choice – whether to use each EMR vendor is a binary choice, with the restriction that they can only use one – the effect is allowed to vary with time.

I bottom-code vendor choice to be "Other" for any vendor that is not, in any year of the sample, in the Top 10 of Vendors chosen that year. The "Other" then is really any "small" vendor. All hazard ratios are presented with the baseline of "Other," i.e. how much more

or less hazardous it is to use e.g. Allscripts relative to any of the very small vendors. I first separate the vendors into three types: Any Large Vendor, Small Vendors, and Self-Developed. Then, I separate out each of the Top 10 vendors to analyze how secure each is.

4.10.1 Aside: The Reflection Problem

Before I present results, it is prudent to discuss the common "reflection problem" of social effect analysis, well described in an early analysis by Manski (1993). The paper breaks down social effects into three types, each of which is econometrically indistinguishable from the others without additional data or theoretical structure.

The issue is this: when a hospital is part of a network (e.g. a group of hospitals that all use the same EMR vendor), and the network shares outcomes, what we observe to be the result – a hospital experiencing a breach at the same time as all its peers – could be due to endogenous, exogenous, or correlated effects. Manski (1993) uses slightly different, more general definitions; I adapt them here for the case of hospital breaches:

1. Endogenous Effect: other hospitals in the network experience a breach, and *therefore* the hospital in question experiences one too
2. Exogenous Effect: some outside factor results in *all hospitals* in the network experiencing a breach
3. Correlated Effect: some factor that every hospital in the network has in common led to the breach

Manski (1993) shows that without additional structure, a classical econometric scheme will not distinguish the three effects. How can we distinguish the three effects? And more fundamentally, should we?

When discussing the role of strategy and scale in cyberattacks, I specifically defined crimes as third-party-perpetrated attacks and mistakes as internal; cyber as scalable and remote, while physical attacks are non-scalable and in-person.

First, a scaled attack can appear and indeed encapsulates all three types of effects. To see a scaled attack, one needs only to observe that there is *any social effect at all*. If attacks are not scaled, and networks do not matter outside of the characteristics that determine them, then we should not observe any social effect. For example, if attacks do not scale through technology choice, but rather because the vendor itself is flawed (a correlated effect), then simply controlling for the choice of vendor as we did in Section 4.10 would provide more information than the additional peer effects I test later in this Section. On the other hand, if there are social effects of any kind, we will be able to learn something from adding in specific group outcome information. Further, we can attribute these choices in common to cyber and physical attacks separately: is the correlation through some kind of technology, as is required for a cyberattack? Or is it something about hospital characteristics driving the technology choice, which might be expected to show up in physical breaches too?

Second, a strategic attack is one conducted by a third party against many hospitals at once – a crime – facilitated by the network they share in common. The crime therefore encapsulates the exogenous effect if we think crimes are contemporary (i.e. occur all at once) or endogenous if we think crimes occur dynamically (the criminal tests the attack on one hospital first). Due to data limitations, I am not able to test the dynamics of cyberattacks any faster than on an annual basis. However, I think of the distinction between endogenous and exogenous as somewhat moot in the context of a scaled attack: the key is whether the attack was perpetrated by a third-party or not, and whether the third-party spurs any kind of social effect at all. Therefore, I

instead add structure to the analysis by comparing crimes and mistakes, specifically to detect if the presence of a third party in the breach induces the exogenous social effect that we would not expect to see in mistakes.

Throughout all analyses, I control for hospital characteristics and ultimate vendor choice, which I show in Section 3 can explain most hospitals' eventual technology choices. Further, as hospitals tend to be *reactive* rather than *proactive* in their security posture, and their vendor choices are made without first-order concern for security, I further treat network structure as orthogonal to security outside the explicit effect of the technology and the set of peers.

4.10.2 Going After the Largest Vendors: The Magnet Effect

I begin the analysis with the "Magnet Effect": the finding that technology products that serve either *more* users or *larger* users could induce attackers to develop exploits and conduct scaled attacks. The Magnet Effect is a user-to-user negative externality that cannot be corrected, since there is no reasonable way for a large user to reimburse a small user for the increase in risk.

I therefore begin the analysis by looking at whether a hospital served by a popular provider – independent of the identity of that provider – faces in and of itself an increase in breach risk. Here, I add into the specification from Section 4.9 one additional explanatory variable:[8]

$$\text{Group Size}_{g,t} = 100 \times \frac{\text{Number of Hospitals in Group } g \text{ at } t}{\text{Total Number of Hospitals at } t}$$

I then run the specification with a number of "groups" as described: states, Health Information Exchange (HIE), and EMR networks. In each case, I treat the network choice as exogenous to the breach risk. I include the covariates from Section 4.10, including a fixed effect for each group, a control for Basic vs. Advanced EMR, the presence of security technologies, and basic hospital characteristics.

To test which peer groups matter when it comes to the network effect of a cyberattack, I test each specification below with a different possibly relevant network:

1. States: are attacks spreading at the local level? For example, an attacker in Maine may seek to exfiltrate many records from Maine patients.
2. HIE: does a Health Information Exchange, where hospitals share records, get compromised all at once?
3. EMR: does the implicit technology network of having the same software provider – which may also be the same as the de facto HIE due to interoperability – facilitate the spreading of attacks? The EMR network is the one I hypothesize results in the software monoculture and market concentration effects on cybersecurity outcomes posited at the start of this paper.
4. Three security technologies: Firewall, Encryption, and Spam/Spyware Filters.[9]

[8] I also remove the control for "Basic EMR" as near all hospitals that have an EMR at all have a Basic EMR, distorting coefficients through a small sample size.

[9] While I did not show results for specific vendors here, I examine if there might be a network effect of using the same, e.g. spam filter as everyone else. The story is natural: a phishing email sent *en masse* would be treated equally by a spam filter regardless of the hospital, so if it slips through one it could slip through all. However, here we do suffer from small sample sizes and a time endogeneity of hospitals who experience past breaches implementing security technologies only *reactively* rather than proactively.

I also test whether or not the networks hold for each type of attack: cyber vs. physical, and crime vs. mistake. The general idea, as discussed in Section 4.8.3 is that crime breaches, perpetrated by a third party who seeks to scale attacks, may show network effects. Physical breaches, on the other hand, would only show network effects if the network relies somehow on physical distance. Mistake breaches

I finally repeat the analysis on *contemporary* networks (i.e. outcomes your peer hospitals in the same year, excluding yourself) and on *lagged* networks (i.e. outcomes your peer hospitals faced last year, excluding yourself). I find no qualitative differences between the results.

I show the results for each group in Tables 6, 7, and 8. Then, I look at security technologies specifically: Firewall (Table 10), Encryption (Table 11), and Spam/Spyware Filters (Table 12). I will guide the reader through each set of results carefully so we can be sure exactly what is and is not being tested in each case.

Table 6: Hazard Ratios: State Magnet Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	1.068*	1.109	1.106**	1.030	1.003
	(0.066)	(0.105)	(0.016)	(0.485)	(0.963)
Advanced EMR	1.110	1.389	1.231	1.054	1.046
	(0.502)	(0.177)	(0.277)	(0.783)	(0.859)
Firewall	1.014	1.640	0.979	0.818	1.094
	(0.944)	(0.109)	(0.933)	(0.376)	(0.757)
Spam/Spyware Filter	0.709*	0.541**	0.610*	0.765	0.851
	(0.095)	(0.032)	(0.072)	(0.254)	(0.527)
Encryption	1.240	1.208	1.576**	1.318	0.844
	(0.215)	(0.490)	(0.043)	(0.184)	(0.499)
Single Sign-On	1.234*	1.208	1.156	1.250	1.383*
	(0.078)	(0.342)	(0.339)	(0.121)	(0.071)
HIE	0.950	1.091	0.859	0.979	1.285
	(0.672)	(0.693)	(0.323)	(0.887)	(0.221)
EDI	1.041	1.007	1.166	1.135	0.915
	(0.837)	(0.983)	(0.537)	(0.605)	(0.761)
RIS	1.030	0.591	0.793	1.745	1.352
	(0.941)	(0.317)	(0.587)	(0.362)	(0.685)
Medium	1.339	1.660	1.451	1.247	1.282
	(0.204)	(0.244)	(0.225)	(0.414)	(0.464)
Large	3.636***	4.896***	4.715***	3.363***	2.894***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Pseudo-R-Squared	0.05	0.07	0.06	0.04	0.04
Hospital Count	4232	4232	4232	4232	4232
Breach Type Count	381	138	239	255	155

Let us use Table 6 as a guide for the rest of this Section. Here, I am testing how the size of the peer group to which the hospital belongs – as a share of all hospitals, so size is really

relative here – affects the hospital's breach risk. Does being in a larger state, in this case (where again, "larger" means "more hospitals," not population per-se), worsen breach risk?

Here, I find yes – for the cases of cyber and crime breaches in particular, breach risk is increasing. A one-percentage point increase in the size of the group (the share of hospitals located in that state) increases cyber- and crime-breach risk by about 10%. Both physical and mistake breach risk increase on point estimate, but the estimates are small and not statistically significant.

The magnet effect distinguishes a strategic attack from a general shock – only a strategic attacker would consciously seek to attack a large group in hopes of capturing as much value as possible at once. Here, it seems hospitals in larger states experience higher breach risk. One explanation is that being in a larger state may be associated with more attacks on hospitals in general, as is the case in California, New York, and other states with major metropolitan areas and many hospitals where an attack may carry prestige for the attacker. That effect only translates minimally to physical breaches, and mistakes even less so. The attacker may therefore be induced to attack the group not necessarily due to attack scalability but rather because they expect some positive effect of capturing many state records at once.

Table 7: Hazard Ratios: HIE Magnet Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	0.996	1.000	0.994	0.998	1.001
	(0.464)	(0.993)	(0.397)	(0.691)	(0.914)
Advanced EMR	0.962	0.605	0.928	1.096	1.221
	(0.911)	(0.308)	(0.870)	(0.837)	(0.725)
Firewall	0.958	1.118	1.308	0.945	0.646
	(0.890)	(0.812)	(0.545)	(0.867)	(0.175)
Spam/Spyware Filter	0.801	1.225	0.587	0.703	1.159
	(0.581)	(0.695)	(0.354)	(0.372)	(0.687)
Encryption	1.372	1.032	1.585	1.460	1.069
	(0.329)	(0.945)	(0.239)	(0.302)	(0.889)
Single Sign-On	1.165	1.015	1.072	1.228	1.309
	(0.403)	(0.961)	(0.782)	(0.357)	(0.306)
HIE	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
EDI	1.965	2.742	2.256	1.866	1.762
	(0.183)	(0.315)	(0.272)	(0.299)	(0.428)
RIS	9.56e+14	2.55e+14	7.00e+14	5.16e+14	1.24e+15
	(.)	(.)	(.)	(.)	(.)
Medium	1.143	2.134	1.050	1.032	1.228
	(0.704)	(0.333)	(0.930)	(0.939)	(0.652)
Large	3.343***	8.479***	4.244***	2.509***	2.823***
	(0.000)	(0.001)	(0.001)	(0.007)	(0.008)
Pseudo-R-Squared	0.05	0.09	0.08	0.04	0.04
Hospital Count	2948	2948	2948	2948	2948
Breach Type Count	176	67	97	115	84

I see no statistically significant impacts of the size of a hospital's health information exchange on a hospital's breach risk. Of note is that hospitals are generally not members of HIEs until later in the same, with higher participation from 2012 onwards as the HITECH Act created exchanges. The sample is also much smaller, with less than 3000 hospitals in HIEs (compared to 4200 in the full sample). Earlier results found joining an HIE had no effect on breach risk (Table 3). Choi et al. (2023) found the act of joining an HIE increases short-term breach risk, but do not control for the actual technologies adopted to join the HIE (i.e. the EMR), while I do. The increase in risk from a joining HIE may come from the accompanying technology adoption rather than the HIE itself.

Table 8: Hazard Ratios: EMR Vendor Magnet Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	0.972**	0.983	0.972	0.975	0.991
	(0.034)	(0.434)	(0.114)	(0.125)	(0.644)
Advanced EMR	0.973	1.101	1.070	0.946	0.978
	(0.869)	(0.705)	(0.734)	(0.787)	(0.935)
Firewall	1.021	1.643	0.977	0.799	1.049
	(0.923)	(0.129)	(0.929)	(0.369)	(0.879)
Spam/Spyware Filter	0.840	0.634	0.704	0.914	1.023
	(0.467)	(0.134)	(0.250)	(0.750)	(0.938)
Encryption	1.033	1.020	1.374	1.110	0.695
	(0.852)	(0.938)	(0.142)	(0.648)	(0.148)
Single Sign-On	1.157	1.168	1.125	1.154	1.214
	(0.238)	(0.449)	(0.449)	(0.342)	(0.306)
HIE	0.869	0.938	0.819	0.870	1.049
	(0.315)	(0.793)	(0.261)	(0.411)	(0.833)
EDI	0.955	0.865	1.005	1.085	0.892
	(0.819)	(0.631)	(0.983)	(0.754)	(0.720)
RIS	0.728	0.454	0.608	1.178	0.920
	(0.389)	(0.111)	(0.203)	(0.772)	(0.902)
Medium	1.391	1.996	1.484	1.218	1.325
	(0.193)	(0.147)	(0.229)	(0.502)	(0.463)
Large	3.501***	6.330***	4.702***	2.809***	2.606***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)
Pseudo-R-Squared	0.05	0.08	0.07	0.05	0.06
Hospital Count	4227	4227	4227	4227	4227
Breach Type Count	345	133	220	224	137

For the EMR vendor, I find that joining a larger EMR vendor – again, iindependent for the identity of that EMR vendor – is actually associated with a small *decline* in a hospital's breach risk. When more hospitals group together under a single provider, then, may be associated with an increase in safety simply because one can now access services that smaller hospitals may find difficult to provide. Alternatively, vendors may switch security strategies as they grow, which I cannot control for as I do not have information on product characteristics, only identity.

I repeat the analysis looking at the size of the network in terms of the number of beds served, which helps us distinguish between vendors that serve a few large hospitals from those that just serve many smaller hospitals. Again, I do not see any magnet effect (i.e., no coefficient on Group Size greater than one).

Table 9: Hazard Ratios: EMR Vendor Magnet Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Beds Share)	0.974**	0.987	0.977	0.973*	0.985
	(0.021)	(0.477)	(0.125)	(0.069)	(0.373)
Advanced EMR	0.969	1.106	1.072	0.937	0.961
	(0.846)	(0.694)	(0.728)	(0.752)	(0.882)
Firewall	1.023	1.646	0.978	0.801	1.050
	(0.916)	(0.128)	(0.934)	(0.373)	(0.877)
Spam/Spyware Filter	0.839	0.632	0.703	0.913	1.026
	(0.464)	(0.132)	(0.247)	(0.748)	(0.931)
Encryption	1.034	1.021	1.375	1.111	0.694
	(0.849)	(0.935)	(0.141)	(0.646)	(0.147)
Single Sign-On	1.157	1.167	1.124	1.156	1.215
	(0.237)	(0.452)	(0.451)	(0.336)	(0.303)
HIE	0.874	0.933	0.818	0.881	1.064
	(0.333)	(0.777)	(0.256)	(0.453)	(0.782)
EDI	0.954	0.864	1.003	1.084	0.892
	(0.812)	(0.629)	(0.989)	(0.758)	(0.719)
RIS	0.735	0.453	0.610	1.193	0.930
	(0.402)	(0.110)	(0.206)	(0.754)	(0.914)
Medium	1.391	1.997	1.484	1.217	1.323
	(0.194)	(0.147)	(0.229)	(0.505)	(0.466)
Large	3.496***	6.329***	4.698***	2.803***	2.600***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)
Pseudo-R-Squared	0.05	0.08	0.07	0.05	0.06
Hospital Count	4227	4227	4227	4227	4227
Breach Type Count	345	133	220	224	137

Finally, Tables 10, 11, and 12 each show the role of using a more popular security technology. The same sizes are small, as not all hospitals have implemented security technologies by the end of the sample. I nonetheless find a small role for using the larger Spam Filters on cyber breaches specifically.

Table 10: Hazard Ratios: Firewall Magnet Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	1.002	1.002	1.004	1.002	0.998
	(0.401)	(0.596)	(0.211)	(0.470)	(0.692)
Advanced EMR	0.954	1.414	0.956	0.800	0.975
	(0.808)	(0.324)	(0.851)	(0.310)	(0.934)
Firewall	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
Spam/Spyware Filter	0.825	0.669	0.645	0.894	1.176
	(0.540)	(0.331)	(0.272)	(0.762)	(0.698)
Encryption	1.077	1.076	1.604	1.142	0.672
	(0.734)	(0.828)	(0.159)	(0.611)	(0.148)
Single Sign-On	1.232	1.109	1.181	1.283	1.303
	(0.150)	(0.669)	(0.382)	(0.149)	(0.205)
HIE	0.713***	0.686*	0.667**	0.774*	0.816
	(0.008)	(0.069)	(0.015)	(0.098)	(0.291)
EDI	1.118	1.628	1.354	0.975	0.814
	(0.714)	(0.342)	(0.460)	(0.941)	(0.578)
RIS	0.654	0.365*	0.566	1.148	0.565
	(0.370)	(0.094)	(0.277)	(0.855)	(0.451)
Medium	1.655*	2.180	1.520	1.540	1.897
	(0.070)	(0.181)	(0.246)	(0.165)	(0.126)
Large	4.160***	7.894***	4.574***	3.285***	3.994***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Pseudo-R-Squared	0.05	0.08	0.07	0.04	0.04
Hospital Count	3709	3709	3709	3709	3709
Breach Type Count	278	107	169	179	117

Table 11: Hazard Ratios: Encryption Magnet Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	1.003	0.989	1.010	1.008	0.987
	(0.795)	(0.579)	(0.495)	(0.540)	(0.429)
Advanced EMR	0.840	1.240	0.868	0.731	0.922
	(0.391)	(0.568)	(0.554)	(0.181)	(0.813)
Firewall	1.052	2.191	1.264	0.829	0.880
	(0.861)	(0.207)	(0.541)	(0.555)	(0.758)
Spam/Spyware Filter	0.595*	0.481*	0.481**	0.621	0.859
	(0.066)	(0.071)	(0.040)	(0.122)	(0.692)
Encryption	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
Single Sign-On	1.127	0.982	1.008	1.163	1.369
	(0.404)	(0.942)	(0.963)	(0.379)	(0.158)
HIE	0.724**	0.692*	0.663**	0.776	0.870
	(0.013)	(0.087)	(0.013)	(0.105)	(0.508)
EDI	1.354	1.811	1.586	1.314	0.892
	(0.299)	(0.315)	(0.222)	(0.433)	(0.799)
RIS	0.549	0.442	0.518	0.655	0.391
	(0.219)	(0.287)	(0.222)	(0.503)	(0.225)
Medium	1.650	1.628	1.492	1.702	1.979
	(0.113)	(0.455)	(0.308)	(0.138)	(0.175)
Large	4.344***	7.599***	4.743***	3.600***	4.178***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)
Pseudo-R-Squared	0.05	0.09	0.07	0.05	0.05
Hospital Count	3348	3348	3348	3348	3348
Breach Type Count	261	96	163	174	105

Table 12: Hazard Ratios: Spam Filter Magnet Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	1.003	1.051**	1.012	0.986	0.999
	(0.802)	(0.019)	(0.416)	(0.337)	(0.954)
Advanced EMR	1.068	1.238	0.984	0.991	1.236
	(0.767)	(0.565)	(0.950)	(0.970)	(0.550)
Firewall	1.456	4.382	1.237	1.088	2.381
	(0.286)	(0.146)	(0.605)	(0.824)	(0.217)
Spam/Spyware Filter	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
Encryption	1.033	0.912	1.284	1.128	0.735
	(0.884)	(0.792)	(0.418)	(0.653)	(0.323)
Single Sign-On	1.178	0.994	1.062	1.245	1.356
	(0.264)	(0.980)	(0.750)	(0.206)	(0.170)
HIE	0.679***	0.714	0.637***	0.704**	0.790
	(0.003)	(0.119)	(0.007)	(0.025)	(0.233)
EDI	1.050	1.502	1.330	0.937	0.689
	(0.871)	(0.425)	(0.480)	(0.849)	(0.319)
RIS	0.435*	0.292*	0.372**	0.632	0.423
	(0.063)	(0.058)	(0.040)	(0.467)	(0.257)
Medium	1.529	1.899	1.427	1.468	1.731
	(0.125)	(0.296)	(0.316)	(0.211)	(0.195)
Large	3.901***	7.638***	4.260***	3.052***	3.741***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)
Pseudo-R-Squared	0.05	0.10	0.06	0.03	0.04
Hospital Count	3545	3545	3545	3545	3545
Breach Type Count	266	100	162	173	111

The results of this Section therefore suggest a limited role for the Magnet Externality, the user-to-user externality that shows up in the size of the network, when the network is measured as the set of hospitals that use some particular EMR vendor or other security technology. Instead, I see evidence that attackers scale their attacks at the state level, whereby either state-level policies or differences in the lucrativeness of various hospitals' records drive differences in group outcomes.

4.10.3 The Gatekeeper Effect

The "Gatekeeper" Effect emerges when a large EMR vendor is able to better secure its users when there are many than when there are few thanks to economies of scale. That is, the Gatekeeper Effect allows larger vendors to better secure their hospitals than the smaller ones.

Table 13 begins by analyzing how looking at the identity of the specific vendor changes a hospital's breach risk.

Table 13: Hazard Ratios, Breach: Intensive Margin

	Breach		Breach		Breach	
Basic EMR	1.235	(0.577)	1.033	(0.933)	1.066	(0.868)
Advanced EMR	1.016	(0.918)	1.080	(0.630)	0.987	(0.935)
Firewall	1.023	(0.909)	0.968	(0.873)	0.999	(0.995)
Spam/Spyware Filter	0.708*	(0.091)	0.800	(0.315)	0.811	(0.378)
Encryption	1.239	(0.210)	1.152	(0.401)	1.104	(0.564)
Single Sign-On	1.215	(0.111)	1.152	(0.267)	1.154	(0.263)
HIE	0.891	(0.350)	0.930	(0.576)	0.822	(0.152)
EDI	1.076	(0.710)	1.005	(0.980)	1.028	(0.890)
RIS	0.926	(0.864)	0.612	(0.234)	0.646	(0.281)
Medium	1.171	(0.498)	1.213	(0.430)	1.196	(0.475)
Large	2.995***	(0.000)	2.959***	(0.000)	2.719***	(0.000)
Large Vendor			1.554	(0.225)		
Self-Developed			4.136***	(0.008)		
Small Vendor			1.000	(.)		
Allscripts					2.076**	(0.045)
CPSI					1.100	(0.836)
Cerner					1.466	(0.259)
Epic					1.720	(0.112)
HIM					0.690	(0.541)
Healthland					0.445	(0.438)
McKesson					1.621	(0.192)
Meditech					0.937	(0.850)
Other					1.000	(.)
Self-Developed					3.318**	(0.024)
Siemens					1.326	(0.500)
Pseudo-R-Squared	0.06		0.06		0.06	
Hospital Count	4232		4227		4227	
Breach Type Count	381		345		345	

I find that simply using any large vendor is worse than using any small vendor for a hospital's breach risk, suggesting that large vendors do not benefit from any kind of Gatekeeper Effect. Similar to the magnet effect, using *any* large vendor is worse than *any* small vendor, though the difference here is in using the identity of the vendor rather than just its size.

Notably, breach risk lowers if a hospital moves from *self-developed* software to a large vendor: going it alone is certainly worse than using an established vendor's product. However, I see the opposite of a gatekeeper effect, whereby large vendors seem to be providing worse security to their customers.

The results persist in Table 14, which examine cyber vs. physical breaches, where cyber breaches are exactly the kind that good software should be able to prevent. Here, we see an even worse effect on breaches from using Self-Developed software, but also from using any Large Vendor over any Small Vendor on cyber breaches but no statistical significance for

physical breaches. That is, using a large vendor appears to hasten cyber breaches for hospitals

Table 14: Hazard Ratios, Total vs. Cyber vs. Physical Breaches

	Breach		Cyber		Physical	
Basic EMR	1.033	(0.933)	3.853	(0.240)	0.737	(0.447)
Advanced EMR	1.080	(0.630)	1.062	(0.808)	1.128	(0.550)
Firewall	0.968	(0.873)	1.449	(0.234)	0.791	(0.322)
Spam/Spyware Filter	0.800	(0.315)	0.658	(0.162)	0.840	(0.511)
Encryption	1.152	(0.401)	1.116	(0.659)	1.245	(0.309)
Single Sign-On	1.152	(0.267)	1.112	(0.623)	1.173	(0.302)
HIE	0.930	(0.576)	1.034	(0.885)	0.979	(0.891)
EDI	1.005	(0.980)	0.889	(0.701)	1.139	(0.604)
RIS	0.612	(0.234)	0.301**	(0.021)	1.118	(0.860)
Medium	1.213	(0.430)	1.519	(0.338)	1.133	(0.669)
Large	2.959***	(0.000)	4.397***	(0.000)	2.624***	(0.000)
Large Vendor	1.554	(0.225)	1.31e+09***	(0.000)	1.108	(0.780)
Self-Developed	4.136***	(0.008)	7.62e+09	(.)	2.173	(0.237)
Small Vendor	1.000	(.)	1.000	(.)	1.000	(.)
Pseudo-R-Squared	0.06		0.10		0.06	
Hospital Count	4227		4227		4227	
Breach Type Count	345		133		224	

I report similar results for crime vs. mistake breaches, which again suggest that the gatekeeper effect – whereby a large vendor should be better at preventing cyberattacks thanks to economies of scale in the cost of security across its users – is not present in the case of hospitals.

Table 15: Hazard Ratios, Total vs. Crime vs. Mistake Breaches

	Breach		Crime		Mistake	
Basic EMR	1.033	(0.933)	1.375	(0.553)	0.633	(0.367)
Advanced EMR	1.080	(0.630)	1.161	(0.441)	1.110	(0.699)
Firewall	0.968	(0.873)	0.938	(0.799)	0.977	(0.938)
Spam/Spyware Filter	0.800	(0.315)	0.680	(0.178)	0.985	(0.957)
Encryption	1.152	(0.401)	1.519**	(0.046)	0.776	(0.289)
Single Sign-On	1.152	(0.267)	1.118	(0.486)	1.204	(0.340)
HIE	0.930	(0.576)	0.864	(0.359)	1.253	(0.315)
EDI	1.005	(0.980)	1.105	(0.677)	0.872	(0.656)
RIS	0.612	(0.234)	0.439*	(0.053)	0.951	(0.946)
Medium	1.213	(0.430)	1.212	(0.542)	1.261	(0.531)
Large	2.959***	(0.000)	3.589***	(0.000)	2.510***	(0.005)
Large Vendor	1.554	(0.225)	7.245**	(0.044)	0.594	(0.149)
Self-Developed	4.136***	(0.008)	25.850***	(0.003)	1.066	(0.943)
Small Vendor	1.000	(.)	1.000	(.)	1.000	(.)
Pseudo-R-Squared	0.06		0.09		0.06	
Hospital Count	4227		4227		4227	
Breach Type Count	345		220		137	

Table 16 shows the results for the proportional hazards model when we include the identity of the vendor as a covariate. Recall that coefficients greater than one *increase* the hazard rate via multiplication.

Table 16: Hazard Ratios, Total vs. Cyber vs. Physical Breaches

	Breach		Cyber		Physical	
Basic EMR	1.066	(0.868)	3.857	(0.235)	0.751	(0.479)
Advanced EMR	0.987	(0.935)	1.057	(0.833)	0.967	(0.869)
Firewall	0.999	(0.995)	1.525	(0.193)	0.817	(0.418)
Spam/Spyware Filter	0.811	(0.378)	0.655	(0.172)	0.858	(0.594)
Encryption	1.104	(0.564)	1.085	(0.740)	1.180	(0.450)
Single Sign-On	1.154	(0.263)	1.140	(0.553)	1.164	(0.325)
HIE	0.822	(0.152)	0.937	(0.791)	0.842	(0.301)
EDI	1.028	(0.890)	0.929	(0.810)	1.165	(0.552)
RIS	0.646	(0.281)	0.355**	(0.047)	1.131	(0.845)
Medium	1.196	(0.475)	1.559	(0.345)	1.110	(0.720)
Large	2.719***	(0.000)	4.397***	(0.001)	2.339***	(0.002)
Allscripts	2.076**	(0.045)	7.014*	(0.066)	1.801	(0.146)
CPSI	1.100	(0.836)	5.763	(0.113)	0.711	(0.537)
Cerner	1.466	(0.259)	4.060	(0.173)	1.228	(0.576)
Epic	1.720	(0.112)	5.403	(0.103)	1.472	(0.294)
HIM	0.690	(0.541)	7.092*	(0.095)	0.171*	(0.096)
Healthland	0.445	(0.438)	0.000	(.)	0.398	(0.382)
McKesson	1.621	(0.192)	7.279*	(0.057)	1.002	(0.997)
Meditech	0.937	(0.850)	2.778	(0.321)	0.780	(0.503)
Other	1.000	(.)	1.000	(.)	1.000	(.)
Self-Developed	3.318**	(0.024)	25.455***	(0.003)	1.812	(0.383)
Siemens	1.326	(0.500)	4.037	(0.207)	1.009	(0.985)
Pseudo-R-Squared	0.06		0.11		0.06	
Hospital Count	4227		4227		4227	
Breach Type Count	345		133		224	

Finally, Tables 17 and 18 show the results for crime and cyber breaches only – the most relevant to this paper.

Table 17: Hazard Ratios, Crime: Intensive Margin

	Crime		Crime		Crime	
Basic EMR	1.713	(0.249)	1.375	(0.553)	1.480	(0.468)
Advanced EMR	1.088	(0.645)	1.161	(0.441)	1.077	(0.710)
Firewall	1.012	(0.962)	0.938	(0.799)	0.978	(0.932)
Spam/Spyware Filter	0.602*	(0.063)	0.680	(0.178)	0.695	(0.225)
Encryption	1.560**	(0.042)	1.519**	(0.046)	1.455*	(0.075)
Single Sign-On	1.154	(0.355)	1.118	(0.486)	1.134	(0.437)
HIE	0.809	(0.172)	0.864	(0.359)	0.781	(0.142)
EDI	1.200	(0.464)	1.105	(0.677)	1.146	(0.572)
RIS	0.620	(0.301)	0.439*	(0.053)	0.486*	(0.081)
Medium	1.226	(0.503)	1.212	(0.542)	1.231	(0.522)
Large	3.669***	(0.000)	3.589***	(0.000)	3.487***	(0.000)
Large Vendor			7.245**	(0.044)		
Self-Developed			25.850***	(0.003)		
Small Vendor			1.000	(.)		
Allscripts					3.043**	(0.050)
CPSI					1.985	(0.281)
Cerner					1.849	(0.259)
Epic					2.418	(0.111)
HIM					1.970	(0.352)
Healthland					0.000	(.)
McKesson					2.110	(0.197)
Meditech					1.423	(0.518)
Other					1.000	(.)
Self-Developed					6.779***	(0.004)
Siemens					1.950	(0.276)
Pseudo-R-Squared	0.08		0.09		0.09	
Hospital Count	4232		4227		4227	
Breach Type Count	239		220		220	

Table 18: Hazard Ratios, Cyber: Intensive Margin

	Cyber		Cyber		Cyber	
Basic EMR	6.770*	(0.087)	3.853	(0.240)	3.857	(0.235)
Advanced EMR	1.123	(0.622)	1.062	(0.808)	1.057	(0.833)
Firewall	1.558	(0.140)	1.449	(0.234)	1.525	(0.193)
Spam/Spyware Filter	0.541**	(0.031)	0.658	(0.162)	0.655	(0.172)
Encryption	1.229	(0.426)	1.116	(0.659)	1.085	(0.740)
Single Sign-On	1.179	(0.427)	1.112	(0.623)	1.140	(0.553)
HIE	1.044	(0.846)	1.034	(0.885)	0.937	(0.791)
EDI	0.993	(0.981)	0.889	(0.701)	0.929	(0.810)
RIS	0.364*	(0.060)	0.301**	(0.021)	0.355**	(0.047)
Medium	1.351	(0.471)	1.519	(0.338)	1.559	(0.345)
Large	3.639***	(0.001)	4.397***	(0.000)	4.397***	(0.001)
Large Vendor			1.31e+09***	(0.000)		
Self-Developed			7.62e+09	(.)		
Small Vendor			1.000	(.)		
Allscripts					7.014*	(0.066)
CPSI					5.763	(0.113)
Cerner					4.060	(0.173)
Epic					5.403	(0.103)
HIM					7.092*	(0.095)
Healthland					0.000	(.)
McKesson					7.279*	(0.057)
Meditech					2.778	(0.321)
Other					1.000	(.)
Self-Developed					25.455***	(0.003)
Siemens					4.037	(0.207)
Pseudo-R-Squared	0.10		0.10		0.11	
Hospital Count	4232		4227		4227	
Breach Type Count	138		133		133	

The results here confirm that indeed, some vendors are associated more strongly with data breach risk than others. The standout is the coefficient on Self-Developed software, which triples the rate of any breach. When we investigate cyber vs. physical breaches, the mechanism emerges: self-developed software enables cyber breaches, multiplying the hazard rate by *twenty-five* relative to the "other" category (which comprises any vendor that serves only a few hospitals).

Similar results hold for other vendors: those that appear to significantly increase general breach risk are, when broken down into cyber vs. physical breaches, really only increasing cyber breach risk. The results therefore pass a sanity check – we would not necessarily expect software to worsen physical breach risk, but maybe even mitigate it. Each of the large vendors is worse for the hazard rate than the "Other" category, consisting of small vendors. Only Healthland, a relatively small vendor, has no associated cyber breaches.

These results are even unexpectedly expected: Allscripts was known to have had an exploitable vulnerability in its products during the sample period that was only discovered years later (Davis (2018)). We can therefore in a sense *uncover* possible common vulnerabilities even when official descriptions exclude them.

The results are threatened if a hospital's choice of vendor is correlated with other unobserved factors that lead them to choose the vendor and also face differential breach risk. However, in [redacted] I showed that nearly all hospital choices could be explained by observables: basic hospital characteristics, their location, and their networks. That is, in practice, hospitals are not necessarily comparing product characteristics to determine which vendor best suits their needs and their security posture; rather, they are choosing what their peers, their neighbors or their GPO are choosing. I include those observable characteristics as controls (not all coefficients are displayed) in every specification.

In Table 19 I look at how breaches made by a third-party (crimes) differ from those that were internal (mistakes). Here, no particular vendor stands out other than Allscripts and Self-Developed software. Once again, self-developed software is likely to be worse than commercially available software given the lack of hospital expertise in software development; and Allscripts' known vulnerability was likely exploited by hospitals who were not aware of or did not patch the vulnerability in time (Davis (2018)).

Table 19: Hazard Ratios, Total vs. Crime vs. Mistake Breaches

	Breach		Crime		Mistake	
Basic EMR	1.066	(0.868)	1.480	(0.468)	0.619	(0.344)
Advanced EMR	0.987	(0.935)	1.077	(0.710)	0.952	(0.856)
Firewall	0.999	(0.995)	0.978	(0.932)	1.006	(0.984)
Spam/Spyware Filter	0.811	(0.378)	0.695	(0.225)	0.969	(0.918)
Encryption	1.104	(0.564)	1.455*	(0.075)	0.747	(0.233)
Single Sign-On	1.154	(0.263)	1.134	(0.437)	1.180	(0.393)
HIE	0.822	(0.152)	0.781	(0.142)	1.073	(0.758)
EDI	1.028	(0.890)	1.146	(0.572)	0.884	(0.693)
RIS	0.646	(0.281)	0.486*	(0.081)	0.914	(0.905)
Medium	1.196	(0.475)	1.231	(0.522)	1.231	(0.584)
Large	2.719***	(0.000)	3.487***	(0.000)	2.232**	(0.020)
Allscripts	2.076**	(0.045)	3.043**	(0.050)	1.189	(0.720)
CPSI	1.100	(0.836)	1.985	(0.281)	0.563	(0.381)
Cerner	1.466	(0.259)	1.849	(0.259)	1.100	(0.811)
Epic	1.720	(0.112)	2.418	(0.111)	1.083	(0.837)
HIM	0.690	(0.541)	1.970	(0.352)	0.000***	(0.000)
Healthland	0.445	(0.438)	0.000	(.)	0.596	(0.638)
McKesson	1.621	(0.192)	2.110	(0.197)	1.058	(0.905)
Meditech	0.937	(0.850)	1.423	(0.518)	0.554	(0.146)
Other	1.000	(.)	1.000	(.)	1.000	(.)
Self-Developed	3.318**	(0.024)	6.779***	(0.004)	1.239	(0.818)
Siemens	1.326	(0.500)	1.950	(0.276)	0.630	(0.494)
Pseudo-R-Squared	0.06		0.09		0.06	
Hospital Count	4227		4227		4227	
Breach Type Count	345		220		137	

Once again, the vendor itself is not directly attributed with the breach in any of the HHS reported breaches – in all cases, hospitals only report that the breach happened but rarely if ever have uncovered the cause of the breach by the time of reporting. Instead, I infer which vendors might have had vulnerabilities by looking at which in the end were used by hospitals at the time of their breaches. The results are reasonable, and line up with news reports of vulnerabilities and common sense. The backwards induction process I describe here is extremely promising for settings in which we do not have explicitly vulnerability information but do have information on outcomes and correlates: we can discover past possibly common vulnerabilities by the existence of common outcomes. Essentially: we know which vendors to watch out for.

The Gatekeeper effect is empirically related to the Magnet Effect in that here, we are concerned with the size of the vendor (and the number of targets). The addition is the focus on vendor identity, so we can see if particular vendors are better or worse at providing security in general. The Magnet Effect specifically asks how the market share of the vendor influences breach risk – as more hospitals gather under one umbrella, is each hospital more at risk?

There, I find no additional Magnet Effect beyond what was captured in the Gatekeeper Effect: attackers target insecure software, but not necessarily because it is popular – rather, *popular software tends to be insecure*. The results together suggest the market incentive does not encourage good security to accompany a popular product.

In summary, I find little to no evidence of the Gatekeeper Externality, and instead find the opposite. Per Section 3, given that the force driving higher welfare under monopoly was the gatekeeper effect, the fact that it is empirically missing suggests antitrust policy faces almost no efficiency tradeoff.

4.10.4 Peers' Breaches: The Floodgates Effect

Next, I investigate the "floodgates effect": once a hospital has chosen to use a particular technology, and once a peer hospital who uses that technology is breached, is the hospital then more likely to experience a breach? Here, the concern is the *scalability* of the attack: once an attacker learns how to breach, say, Epic's EMR systems or Cisco's Firewalls, their marginal cost of repeating the attack is lower – making all other hospitals that use the same technology cheaper to attack. Is there a floodgates effect, and, if so, which technologies release it?

The Floodgates effect is essentially a conditional Magnet Effect: the magnet effect concerns the attacker's *extensive* margin choice of which vendor to attack, while the floodgates effect asks if, once the attacker has chosen to attack a vendor, if all users are at increased risk.

I test the Floodgates hypothesis by augmenting the specification from Section 4.10.2 with the following additional term:

$$\text{Perc. Group Breaches}_{g,t} = 100 \times \frac{\text{Number of Hospitals Breached in Group } g \text{ at } t}{\text{Number of Hospitals in Group } g \text{ at } t}$$

That is: what percentage of the peer group was breached that year? Each coefficient corresponds to the change in the hazard rate following a one percentage point increase in the share of peer hospitals breached. As with the Magnet Effect, I test different peer groups to identify the relevant group through which attacks might be scalable.

In the interest of concision, I exclude coefficients from the control variables from display.

Table 20: Hazard Ratios: State Floodgates Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	1.071*	1.138**	1.121***	1.046	1.023
	(0.064)	(0.043)	(0.008)	(0.311)	(0.738)
Perc. Group Breaches	0.981	0.763***	0.924	0.846**	0.564***
	(0.669)	(0.009)	(0.242)	(0.025)	(0.000)
Advanced EMR	1.112	1.398	1.235	1.063	1.062
	(0.495)	(0.170)	(0.269)	(0.750)	(0.807)
Firewall	1.014	1.670*	0.976	0.813	1.107
	(0.945)	(0.098)	(0.923)	(0.360)	(0.728)
Spam/Spyware Filter	0.709*	0.539**	0.613*	0.769	0.855
	(0.096)	(0.031)	(0.073)	(0.260)	(0.543)
Encryption	1.240	1.206	1.576**	1.319	0.844
	(0.214)	(0.497)	(0.042)	(0.182)	(0.504)
Single Sign-On	1.234*	1.208	1.152	1.251	1.404*
	(0.077)	(0.338)	(0.350)	(0.120)	(0.058)
HIE	0.951	1.085	0.857	0.987	1.311
	(0.678)	(0.709)	(0.315)	(0.927)	(0.194)
EDI	1.040	1.008	1.165	1.132	0.907
	(0.841)	(0.979)	(0.538)	(0.610)	(0.742)
RIS	1.030	0.589	0.795	1.742	1.340
	(0.940)	(0.318)	(0.592)	(0.363)	(0.694)
Medium	1.341	1.671	1.458	1.252	1.288
	(0.202)	(0.238)	(0.219)	(0.406)	(0.455)
Large	3.642***	4.928***	4.743***	3.392***	2.913***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Pseudo-R-Squared	0.05	0.07	0.06	0.05	0.05
Hospital Count	4232	4232	4232	4232	4232
Breach Type Count	381	138	239	255	155

I find that, if anything, a hospital breached in the same state *lowers* your own breach risk, perhaps as information on vigilance is spread at the state level. The vigilance-effect is particularly strong for mistake breaches, suggesting information may be spread on how to not repeat unnecessary breaches internally.

Table 21: Hazard Ratios: HIE Floodgates Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	0.997	1.001	0.995	0.998	1.001
	(0.559)	(0.906)	(0.463)	(0.763)	(0.834)
Perc. Group Breaches	1.033***	1.078***	1.044***	1.027**	1.093**
	(0.002)	(0.005)	(0.000)	(0.023)	(0.032)
Advanced EMR	0.991	0.694	1.094	1.111	1.175
	(0.980)	(0.525)	(0.869)	(0.821)	(0.776)
Firewall	0.952	1.063	1.308	0.949	0.650
	(0.879)	(0.900)	(0.565)	(0.879)	(0.186)
Spam/Spyware Filter	0.832	1.299	0.610	0.725	1.233
	(0.656)	(0.626)	(0.410)	(0.423)	(0.561)
Encryption	1.345	1.001	1.606	1.426	1.007
	(0.365)	(0.998)	(0.240)	(0.336)	(0.988)
Single Sign-On	1.152	0.975	1.084	1.213	1.286
	(0.440)	(0.933)	(0.753)	(0.388)	(0.339)
HIE	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
EDI	1.936	2.666	2.171	1.856	1.722
	(0.193)	(0.331)	(0.297)	(0.303)	(0.447)
RIS	6.34e+14	6.09e+14	6.86e+13***	2.17e+14	1.60e+14
	(.)	(.)	(0.000)	(.)	(.)
Medium	1.166	1.876	1.011	1.060	1.305
	(0.672)	(0.428)	(0.984)	(0.889)	(0.578)
Large	3.340***	8.005***	4.067***	2.537***	2.909***
	(0.000)	(0.001)	(0.002)	(0.009)	(0.009)
Pseudo-R-Squared	0.05	0.10	0.09	0.04	0.05
Hospital Count	2935	2935	2935	2935	2935
Breach Type Count	174	66	96	114	83

In health information exchanges, however, I see across-the-board increases in one's own breach risk when a peer has experienced a breach. There, it may be that the technology itself is flawed, or that in the later period – when HIEs were commonplace – the entire HIE is compromised (as Choi et al. (2023) claim is possible). Again, the small sample size means results must be taken with caution.

Table 22: Hazard Ratios: EMR Vendor Floodgates Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	0.971**	0.984	0.972	0.971*	0.992
	(0.028)	(0.469)	(0.113)	(0.097)	(0.712)
Perc. Group Breaches	0.818**	0.911	0.985	0.625***	0.579**
	(0.032)	(0.652)	(0.870)	(0.002)	(0.046)
Advanced EMR	0.966	1.103	1.071	0.926	0.926
	(0.831)	(0.700)	(0.732)	(0.708)	(0.777)
Firewall	1.026	1.641	0.977	0.812	1.062
	(0.904)	(0.130)	(0.929)	(0.400)	(0.848)
Spam/Spyware Filter	0.833	0.635	0.704	0.895	1.011
	(0.441)	(0.135)	(0.250)	(0.694)	(0.970)
Encryption	1.028	1.019	1.373	1.095	0.686
	(0.874)	(0.942)	(0.143)	(0.690)	(0.133)
Single Sign-On	1.170	1.170	1.124	1.170	1.249
	(0.204)	(0.445)	(0.449)	(0.298)	(0.243)
HIE	0.876	0.934	0.818	0.883	1.090
	(0.344)	(0.783)	(0.262)	(0.465)	(0.700)
EDI	0.957	0.863	1.005	1.097	0.902
	(0.827)	(0.627)	(0.982)	(0.726)	(0.752)
RIS	0.813	0.451	0.607	1.651	1.680
	(0.599)	(0.107)	(0.201)	(0.467)	(0.586)
Medium	1.384	1.994	1.484	1.217	1.318
	(0.200)	(0.148)	(0.230)	(0.505)	(0.473)
Large	3.487***	6.332***	4.700***	2.802***	2.579***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)
Pseudo-R-Squared	0.05	0.08	0.07	0.05	0.06
Hospital Count	4225	4225	4225	4225	4225
Breach Type Count	343	133	220	222	135

For EMRs, we see exactly the floodgates effect expected if third-party attackers are scaling attacks through technology: the crime breaches in particular show the strongest (and only statistically significant) floodgates effect, with a 1% increase in the share of peers who experience a breach increasing one's own breach risk by 15%. We see no effect for physical or mistake breaches, exactly because those breaches do not scale well and do involve a third-party attacker, respectively. On point estimate, cyber breaches show the same floodgates effect, but the crime result shows that it really is the presence of a third party that ensures the attack is, in the end, repeated.

For our security technologies, Firewalls show a small floodgates effect, while the other technologies do not.

Table 23: Hazard Ratios: Firewall Floodgates Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	1.002	1.002	1.004	1.003	0.999
	(0.428)	(0.606)	(0.209)	(0.397)	(0.888)
Perc. Group Breaches	0.922	0.936	1.021*	0.736	0.581***
	(0.127)	(0.571)	(0.069)	(0.162)	(0.006)
Advanced EMR	0.964	1.419	0.954	0.816	0.984
	(0.853)	(0.320)	(0.844)	(0.356)	(0.958)
Firewall	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
Spam/Spyware Filter	0.822	0.666	0.647	0.903	1.198
	(0.531)	(0.328)	(0.276)	(0.783)	(0.666)
Encryption	1.087	1.079	1.597	1.158	0.681
	(0.704)	(0.821)	(0.163)	(0.574)	(0.162)
Single Sign-On	1.237	1.109	1.177	1.299	1.326
	(0.143)	(0.668)	(0.391)	(0.134)	(0.177)
HIE	0.724**	0.687*	0.665**	0.797	0.850
	(0.011)	(0.071)	(0.014)	(0.147)	(0.407)
EDI	1.116	1.621	1.348	0.978	0.839
	(0.719)	(0.346)	(0.467)	(0.948)	(0.635)
RIS	0.664	0.366*	0.566	1.179	0.574
	(0.386)	(0.095)	(0.277)	(0.827)	(0.465)
Medium	1.640*	2.169	1.509	1.519	1.858
	(0.075)	(0.183)	(0.255)	(0.178)	(0.139)
Large	4.146***	7.872***	4.529***	3.268***	3.952***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Pseudo-R-Squared	0.05	0.08	0.07	0.04	0.05
Hospital Count	3700	3700	3700	3700	3700
Breach Type Count	278	107	169	179	117

Table 24: Hazard Ratios: Encryption Floodgates Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	1.003	0.988	1.010	1.010	0.987
	(0.784)	(0.527)	(0.472)	(0.456)	(0.458)
Perc. Group Breaches	0.982	0.988	0.934	0.950	1.018
	(0.552)	(0.885)	(0.273)	(0.360)	(0.749)
Advanced EMR	0.830	1.241	0.855	0.718	0.911
	(0.363)	(0.565)	(0.517)	(0.157)	(0.785)
Firewall	1.034	2.155	1.227	0.814	0.879
	(0.908)	(0.215)	(0.592)	(0.518)	(0.754)
Spam/Spyware Filter	0.598*	0.484*	0.484**	0.622	0.859
	(0.069)	(0.073)	(0.042)	(0.126)	(0.692)
Encryption	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
Single Sign-On	1.131	0.981	1.029	1.172	1.334
	(0.391)	(0.939)	(0.877)	(0.355)	(0.195)
HIE	0.715**	0.693*	0.653**	0.760*	0.854
	(0.010)	(0.090)	(0.010)	(0.079)	(0.450)
EDI	1.340	1.793	1.574	1.309	0.882
	(0.314)	(0.321)	(0.227)	(0.441)	(0.780)
RIS	0.549	0.443	0.517	0.659	0.387
	(0.218)	(0.285)	(0.222)	(0.510)	(0.219)
Medium	1.687	1.609	1.488	1.756	2.110
	(0.106)	(0.467)	(0.313)	(0.129)	(0.163)
Large	4.377***	7.493***	4.673***	3.638***	4.381***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)
Pseudo-R-Squared	0.06	0.09	0.07	0.05	0.05
Hospital Count	3333	3333	3333	3333	3333
Breach Type Count	259	96	162	172	104

Table 25: Hazard Ratios: Spam Filter Floodgates Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	1.003	1.052**	1.017	0.986	0.999
	(0.783)	(0.018)	(0.280)	(0.347)	(0.971)
Perc. Group Breaches	1.002	1.030	0.852*	1.017	0.959
	(0.947)	(0.702)	(0.065)	(0.251)	(0.756)
Advanced EMR	1.101	1.235	0.984	1.033	1.380
	(0.669)	(0.570)	(0.951)	(0.900)	(0.381)
Firewall	1.457	4.344	1.243	1.095	2.376
	(0.285)	(0.148)	(0.598)	(0.811)	(0.218)
Spam/Spyware Filter	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
Encryption	1.009	0.895	1.286	1.097	0.702
	(0.966)	(0.751)	(0.415)	(0.727)	(0.252)
Single Sign-On	1.197	1.030	1.083	1.249	1.379
	(0.222)	(0.908)	(0.674)	(0.203)	(0.153)
HIE	0.674***	0.696*	0.625***	0.704**	0.799
	(0.002)	(0.093)	(0.005)	(0.026)	(0.260)
EDI	1.037	1.492	1.328	0.922	0.674
	(0.904)	(0.431)	(0.483)	(0.812)	(0.294)
RIS	0.429*	0.296*	0.374**	0.615	0.395
	(0.059)	(0.061)	(0.042)	(0.443)	(0.221)
Medium	1.554	1.903	1.432	1.497	1.798
	(0.117)	(0.295)	(0.312)	(0.198)	(0.182)
Large	3.909***	7.530***	4.218***	3.076***	3.824***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)
Pseudo-R-Squared	0.05	0.09	0.06	0.04	0.05
Hospital Count	3534	3534	3534	3534	3534
Breach Type Count	264	99	161	172	110

4.10.5 The Sum-of-Efforts Effect

Finally, I test the Sum-of-Efforts hypothesis, put forth in Florencio and Herley (2011) and formalized in Section 3. The attacker's *extensive margin* choice – whether to attack the group of hospitals – depends exactly on the *total effort* put forth by the hospitals, weighted by their value – the $\sum_i (1 - s_i)v_i$ in Section 3. Therefore, an increase in s_i by even just one hospital decreases the attacker's incentive for attacking the *entire group*. Each individual hospital's investment, then, has a *positive externality*, keeping not just itself safe but also all others.

I formally test whether there exists a positive externality of peer investment by explicitly adding in peer investment into the specification from Section 4.10.4 as follows: the fraction of peers in the group who have implemented each of the security technologies:

$$\text{Peer Security}_{g,t} = 100 \times \frac{\text{Number of Hospitals With Technology } s \text{ at } t}{\text{Number of Hospitals in Group } g \text{ at } t}$$

where technology s is one of the three main security technologies: a firewall, encryption, and a spam/spyware filter. Note that I am not controlling for the type of the software or its vendor – which itself defines one of the relevant groups – but rather just the *presence* of the technology. I continue to control for a hospital's own choice of technology.

Note that in the case of hospitals' EMR systems, these technologies are not necessarily built-in but must be requested and implemented by the hospital itself – e.g., not everyone who uses Epic will necessarily possess a Firewall. The mechanism also requires that attackers *know* and are making decisions about hospital breaches on the basis of the security technologies they may possess. If attackers *expect*, say, 60% of Epic-using hospitals to have a Firewall, they may make an informed choice about their threat investments, either by trying to bypass the Firewall or choosing a different group to target. However, if a hospital either keeps its security choices secret[10] then we may fail to detect any total efforts effect. Identification will require that hospitals do not make coordinated security technology choices, so that implementation is exogenous to the network formation process. In Marti (2024), I found hospitals do not respond to peers' breaches but do respond to their own by implementing security technologies. Until that breach occurs, however, the pattern of implementation simply follows time and basic hospital characteristics.

I begin with the within-state total efforts effect: as more of a state's hospitals implement security technologies, does each hospital benefit from the total efforts externality, reducing their own breach risk? Note that there, I take security technology activity as exogenous. Underlying the implementation of security technology may be a broader state-level policy encouraging security, or increased concerns about breaches in large states (see Table 6). I cannot differentiate, but can only comment on what happens to basic breach risk endogenously.

Further, group use of security technology may be correlated with group breaches if we expect security technology to actually be affected, leading to potential multicollinearity issues. However, including both helps us separate out the indirect effect of group firewall use on group breaches from the total efforts mechanism, group firewall use on scaled group attack attempts. Since the panel dataset is large, and previous estimates suggest reactive rather than proactive security technology use, I consider it reasonable to assume some independence.

[10]Not likely, as I use semi-public datasets throughout this analysis,

Table 26: Hazard Ratios: State Sum-of-Efforts Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	1.055	1.143**	1.105**	1.038	1.028
	(0.167)	(0.047)	(0.034)	(0.445)	(0.691)
Perc. Group Breaches	0.986	0.734***	0.926	0.850**	0.554***
	(0.750)	(0.005)	(0.252)	(0.030)	(0.000)
Perc. Group Firewall	1.000	1.027	0.999	0.986	1.007
	(1.000)	(0.414)	(0.961)	(0.443)	(0.772)
Perc. Group Encryption	0.999	1.008	0.997	1.002	1.008
	(0.864)	(0.549)	(0.772)	(0.874)	(0.471)
Perc. Group Spam Filter	0.993	0.977	0.995	1.004	0.993
	(0.665)	(0.433)	(0.805)	(0.847)	(0.765)
Advanced EMR	1.126	1.370	1.250	1.075	1.050
	(0.449)	(0.199)	(0.245)	(0.706)	(0.843)
Firewall	1.025	1.576	0.987	0.847	1.087
	(0.901)	(0.147)	(0.959)	(0.473)	(0.772)
Spam/Spyware Filter	0.718	0.569**	0.615*	0.764	0.869
	(0.113)	(0.048)	(0.081)	(0.256)	(0.583)
Encryption	1.253	1.162	1.604**	1.320	0.818
	(0.199)	(0.582)	(0.037)	(0.181)	(0.431)
Single Sign-On	1.241*	1.195	1.162	1.259	1.398*
	(0.071)	(0.366)	(0.326)	(0.110)	(0.060)
HIE	0.976	1.070	0.881	0.999	1.298
	(0.850)	(0.757)	(0.430)	(0.996)	(0.222)
EDI	1.045	0.984	1.173	1.142	0.895
	(0.825)	(0.958)	(0.520)	(0.586)	(0.706)
RIS	1.033	0.593	0.793	1.733	1.344
	(0.936)	(0.331)	(0.586)	(0.368)	(0.690)
Medium	1.338	1.686	1.451	1.251	1.297
	(0.207)	(0.230)	(0.227)	(0.409)	(0.442)
Large	3.635***	4.980***	4.723***	3.391***	2.940***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pseudo-R-Squared	0.05	0.07	0.06	0.05	0.05
Hospital Count	4232	4232	4232	4232	4232
Breach Type Count	381	138	239	255	155

Focus here on the last three covariates: the percentage of the group that had a Firewall, Encryption, and Spam Filter respectively. In Table 26, we see that no coefficient is significant, though the magnet effect and floodgates effect remain statistically significant for the cases of Cyber and Crime breaches. That is, the additional exploration of the total efforts effect – at least when measured as the fraction of hospitals that use a Firewall, Encryption, or Spam Filter – does not appear to add much to our understanding of how breaches spread at the state level.

Table 27: Hazard Ratios: HIE Sum-of-Efforts Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	0.995	1.000	0.994	0.996	0.998
	(0.318)	(0.980)	(0.375)	(0.480)	(0.811)
Perc. Group Breaches	0.994	1.063***	1.011	0.999	1.047
	(0.674)	(0.001)	(0.306)	(0.949)	(0.373)
Perc. Group Firewall	0.984**	0.982	0.984*	0.987*	0.982*
	(0.027)	(0.118)	(0.074)	(0.095)	(0.090)
Perc. Group Encryption	1.007	1.002	0.997	1.010	1.018
	(0.472)	(0.911)	(0.772)	(0.356)	(0.234)
Perc. Group Spam Filter	0.964***	0.973	0.975*	0.966***	0.960**
	(0.003)	(0.236)	(0.090)	(0.005)	(0.027)
Advanced EMR	1.207	0.578	1.256	1.325	1.388
	(0.641)	(0.286)	(0.681)	(0.545)	(0.573)
Firewall	1.197	1.500	1.737	1.154	0.775
	(0.625)	(0.521)	(0.320)	(0.709)	(0.461)
Spam/Spyware Filter	0.979	1.622	0.693	0.851	1.482
	(0.963)	(0.464)	(0.575)	(0.707)	(0.277)
Encryption	1.296	1.045	1.676	1.323	0.896
	(0.456)	(0.932)	(0.238)	(0.466)	(0.814)
Single Sign-On	1.203	0.975	1.108	1.254	1.336
	(0.331)	(0.936)	(0.693)	(0.320)	(0.282)
HIE	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
EDI	2.015	2.656	2.148	1.936	1.880
	(0.160)	(0.314)	(0.283)	(0.273)	(0.377)
RIS	9.36e+14	7.16e+14	7.81e+15	9.31e+13	3.79e+14
	(.)	(.)	(.)	(.)	(.)
Medium	1.147	1.936	0.987	1.041	1.286
	(0.707)	(0.399)	(0.981)	(0.924)	(0.600)
Large	3.287***	8.044***	3.968***	2.473**	2.867**
	(0.000)	(0.001)	(0.002)	(0.011)	(0.011)
Pseudo-R-Squared	0.06	0.12	0.10	0.05	0.06
Hospital Count	2935	2935	2935	2935	2935
Breach Type Count	174	66	96	114	83

Next, Table 27 looks at how attacks might spread through Health Information Exchanges. Under the magnet effect (Table 7) and floodgates effect (Table 21), I found a strong floodgates effect across all breach types, though no magnet effect. Here, that floodgates effect coefficient is only statistically significant for cyber breaches – following the general hypothesis of this Section – while in other cases, the total efforts effect suggests that as more of the group adopts a spam filter, the rest of the group becomes safer (recall that a coefficient less than one indicates postponed breaches, or prolonged security).

Table 28: Hazard Ratios: EMR Vendor Sum-of-Efforts Effect

	Breach	Cyber	Crime	Physical	Mistake
Group Size (Market Share)	0.981	0.994	0.984	0.980	0.996
	(0.189)	(0.793)	(0.402)	(0.256)	(0.843)
Perc. Group Breaches	0.795**	0.898	0.958	0.537***	0.555**
	(0.016)	(0.604)	(0.662)	(0.000)	(0.032)
Perc. Group Firewall	1.023	0.965	0.978	1.052	1.100
	(0.493)	(0.508)	(0.548)	(0.254)	(0.128)
Perc. Group Encryption	0.945**	1.019	0.964	0.918**	0.936
	(0.032)	(0.697)	(0.226)	(0.018)	(0.106)
Perc. Group Spam Filter	0.978	0.994	1.003	0.960	0.942
	(0.419)	(0.856)	(0.895)	(0.260)	(0.210)
Advanced EMR	1.038	1.117	1.134	0.999	0.983
	(0.825)	(0.671)	(0.545)	(0.996)	(0.952)
Firewall	0.998	1.647	0.979	0.774	1.000
	(0.994)	(0.124)	(0.935)	(0.308)	(1.000)
Spam/Spyware Filter	0.864	0.642	0.703	0.939	1.082
	(0.547)	(0.148)	(0.254)	(0.831)	(0.793)
Encryption	1.054	1.022	1.415	1.134	0.694
	(0.771)	(0.932)	(0.118)	(0.591)	(0.150)
Single Sign-On	1.215	1.179	1.160	1.220	1.271
	(0.126)	(0.430)	(0.351)	(0.201)	(0.216)
HIE	0.952	0.964	0.880	0.959	1.129
	(0.743)	(0.887)	(0.505)	(0.821)	(0.602)
EDI	0.968	0.866	1.014	1.118	0.931
	(0.874)	(0.636)	(0.955)	(0.681)	(0.833)
RIS	0.918	0.473	0.667	2.071	2.076
	(0.829)	(0.128)	(0.299)	(0.312)	(0.471)
Medium	1.363	1.986	1.473	1.194	1.302
	(0.226)	(0.153)	(0.242)	(0.550)	(0.496)
Large	3.459***	6.310***	4.685***	2.758***	2.557***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)
Pseudo-R-Squared	0.06	0.08	0.07	0.06	0.06
Hospital Count	4225	4225	4225	4225	4225
Breach Type Count	343	133	220	222	135

Finally, Table 28 affirms again that the most effective spillovers – positive externalities – come from the EMR-groups' implementation of spam filters. Those hospitals who use EMRs whose *other* users have implemented a spam filter see a 18-22% in the reduction of their breach hazard rate, varying by the type of breach. The reduction is greatest for cyber and crime breaches – i.e. those that are perpetrated by a third-party and involve digital technologies. Given that many breaches originate as phishing emails that then take over a computer or trick the user into sharing credentials – such as those to their EMR system – it makes sense that

group adoption of spam filters would generally deter attackers from using phishing tactics, instead shifting to other tactics or other targets.

5 Conclusion: Which Theoretical Effects Dominate Empirically?

In this Section, I empirically evaluate the state of cybersecurity in the U.S. healthcare sector by combining three datasets: cybersecurity outcomes from the HHS, technology status from HIMSS, and hospital characteristics from AHA. I present the first comprehensive analysis of how hospital digitization has affected cybersecurity outcomes that takes into consideration the structure of healthcare technology market competition – in particular, the lack of proper interoperability and the high start-up costs – showing how hospitals' own lack of true choices leaves them unable to respond quickly or sufficiently to data breaches.

Further, the act of digitization itself (the "extensive margin") creates new opportunities for cybercriminals to target hospitals. When hospitals adopt basic technologies without advanced directives or careful consideration, cyberattacks increase at partial expense of physical attacks. Larger hospitals, and those without security technologies, are more likely to experience data breaches at any given moment.

I use a "revealed security" approach to show how some technology vendors are likely to be used by hospitals that experience data breaches, even when the cause of the breach is unknown. My simple and flexible methodology allows anyone with breach outcome data to identify "hotspots," or vendors who likely had unknown-to-them vulnerabilities being exploited by third-party attackers.

Finally, I investigate the "network externalities" of cybersecurity. I find some evidence for the "magnet effect," where hospitals in large markets or using high-market-share EMR vendors are more likely to experience data breaches. I also find support for the "floodgates effect" as attacks spread through *technology networks*: hospitals that share an EMR vendor are more likely to experience breaches in conjunction, compared with hospitals that are simply in the same state, GPO, or health information exchange. That is, software monoculture allows attacks to capture more value all at once. I find little evidence of a gatekeeper effect, and instead find that large firms are providing their users with worse security, suggesting security costs do not scale well. Finally, I find some evidence that group-level security can act to deter attacks, through the total efforts effect, but only for extremely specific technologies like spam filters.

From Section 3, we know that when the technology vendor does not experience economies of scale in security costs, and when its own incentive to secure is lower than the value hospitals stand to lose, that a concentrated market is not the optimal market structure. When the positive externalities cannot outweigh the negative, the overall market may be better off forgoing any positive network effects and instead focusing on breaking the attacker's economies of scale. By doing so, the market can eliminate the Magnet and the Floodgates Externalities at little cost, keeping all hospitals, instead of commonly at risk of catastrophe, individually strong.

6 References

References

Acemoglu, D., Malekian, A., and Ozdaglar, A. (2016). Network security and contagion. Journal of Economic Theory, 166(C):536–585.

Adler-Milstein, J., DesRoches, C. M., Kralovec, P., Foster, G., Worzala, C., Charles, D., Searcy, T., and Jha, A. K. (2015). Electronic Health Record Adoption In US Hospitals: Progress Continues, But Challenges Persist. Health Affairs, 34(12):2174–2180. Publisher: Health Affairs.

Adler-Milstein, J. and Jha, A. K. (2017). HITECH Act Drove Large Gains In Hospital Electronic Health Record Adoption. Health Affairs, 36(8):1416–1422. Publisher: Health Affairs.

American Hospital Association (2024). AHA Survey: Change Healthcare Cyberattack Significantly Disrupts Patient Care, Hospitals' Finances | AHA.

Anderson, R. (2001). Why information security is hard: an economic perspective. Link.

Arce, D. G. (2020). Cybersecurity and platform competition in the cloud. Computers & Security, 93:101774.

August, T., Dao, D., and Niculscu, M. F. (2021). Economics of ransomware: Risk interdependence and large-scale attacks. Management Science, Forthcoming.

August, T., Niculescu, M. F., and Shin, H. (2014). Cloud Implications on Software Network Structure and Security Risks. Information Systems Research, 25(3):489–510. Publisher: INFORMS.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms*. The Quarterly Journal of Economics, 135(2):645–709.

Baqaee, D. and Farhi, E. (2019). The macroeconomic impact of microeconomic shocks: Beyond hulten's theorem. Econometrica, 87(4):1155–1203.

Bartz, D. (2022). U.S. sues to block UnitedHealth's $8 bln deal for Change Healthcare. Reuters.

Benkard, C. L., Yurukoglu, A., and Zhang, A. L. (2021). Concentration in product markets. Working Paper.

Bloomberg (2021). That Cream Cheese Shortage You Heard About? Cyberattacks Played a Part. Elizabeth Elkin and Deena Shanker, December 9th, 2021. Link.

Bloomberg (2022). Apple (AAPL) Defends App Store Security From New US Antitrust Bill, Critics. Leah Nylen, June 22, 2022.

Bouveret, A. (2018). Cyber Risk for the Financial Sector: A Framework for Quantitative Assessment. IMF Working Paper.

Chesney, R. (2021). SolarWinds and the Holiday Bear Campaign: A Case Study for the Classroom. Link.

Choi, J. P., Fershtman, C., and Gandal, N. (2007). Network Security: Vulnerabilities and Disclosure Policy. SSRN Scholarly Paper 1133779, Social Science Research Network, Rochester, NY.

Choi, S. J., Chen, M., and Tan, X. (2023). Assessing the impact of health information exchange on hospital data breach risk. International Journal of Medical Informatics, 177:105149.

Choi, S. J., Johnson, M. E., and Lehmann, C. U. (2019). Data breach remediation efforts and their implications for hospital quality. Health Services Research, 54(5):971–980. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1475-6773.13203.

Clement, N. (2023). M&A Effect on Data Breaches in Hospitals: 2010-2022. Working Paper.

Congiu, R., Sabatino, L., and Sapi, G. (2022). The Impact of Privacy Regulation on Web Traffic: Evidence From the GDPR. Information Economics and Policy, 61:101003.

Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2):187–202. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1972.tb00899.x.

Crosignani, M., Macchiavelli, M., , and Silva, A. F. (2021). Pirates without borders: The propagation of cyberattacks through firms' supply chains. New York Times Staff Reports, 937.

Davis, J. (2018). Allscripts sued over ransomware attack, accused of 'wanton' disregard. Technical report, Healthcare IT News.

Dranove, D., Forman, C., Goldfarb, A., and Greenstein, S. (2014a). The Trillion Dollar Conundrum: Complementarities and Health Information Technology. American Economic Journal: Economic Policy, 6(4):239–270.

Dranove, D., Garthwaite, C., Li, B., and Ody, C. (2014b). Investment Subsidies and the Adoption of Electronic Medical Records in Hospitals.

Duffie, D. and Younger, J. (2019). Cyber runs: How a cyber attack could affect U.S. financial institutions | Brookings. Brookings Working Paper.

Eisenbach, T. M., Kovner, A., and Lee, M. J. (2021). Cyber risk and the u.s. financial system: A pre-mortem analysis. New York Times Staff Reports, 909.

Farboodi, M. and Veldkamp, L. (2021). A growth model of the data economy. Working Paper.

Florencio, D. and Herley, C. (2011). Where Do All The Attacks Go? Economics of Information and Security III, Bruce Schneier.

Ford, A., Al-Nemrat, A., Ghorashi, S. A., and Davidson, J. (2021). The Impact of Data Breach Announcements on Company Value in European Markets. Workshop on the Economics of Information Security, page 8.

Gabriel, M. H., Noblin, A., Rutherford, A., Walden, A., and Cortelyou-Ward, K. (2018). Data breach locations, types, and associated characteristics among US hospitals. The American Journal of Managed Care, 24(2):78–84.

Galeotti, A., Golub, B., and Goyal, S. (2020). Targeting interventions in networks. Econometrica, 88(6):2445–2471.

Ganapati, S. (2021). Growing oligopolies, prices, output, and productivity. American Economic Journal: Microeconomics, 13(3):309–27.

Geer, D., Jardine, E., and Leverett, E. (2020). On market concentration and cybersecurity risk. Journal of Cyber Policy, 5(1):9–29.

Goldsmith, J. C. (2024). Will The Change Healthcare Incident Change Health Care? Health Affairs Forefront.

Goyal, S. and Vigier, A. (2014). Attack, defence, and contagion in networks. The Review of Economic Studies, 81(4):1518–1542.

Greenberg, A. (2024). Hackers Behind the Change Healthcare Ransomware Attack Just Received a $22 Million Payment. Wired. Section: tags.

Herley, C. and Florêncio, D. (2008). A profitless endeavor: phishing as tragedy of the commons. In Proceedings of the 2008 New Security Paradigms Workshop, NSPW '08, pages 59–70, New York, NY, USA. Association for Computing Machinery.

Hulten, C. R. (1978). Growth Accounting with Intermediate Inputs. The Review of Economic Studies, 45(3):511–518.

Hydari, M. Z., Gaynor, M., and Telang, R. (2012). Is Patient Data Better Protected in Competitive Healthcare Markets? In Workshop on the Economics of Information Security.

Hyun, J., Kim, D., and Shin, S.-R. (2020). The role of global connectedness and market power in crises: Firm-level evidence from the covid-19 pandemic. Working Paper.

Jamilov, R., Rey, H., and Tahoun, A. (2021). The anatomy of cyber risk. Working Paper.

Katz, M. L. and Shapiro, C. (1985). Network externalities, competition, and compatibility. The American Economic Review, 75(3):424–440.

Kim, S. H. and Kwon, J. (2019). How Do EHRs and a Meaningful Use Initiative Affect Breaches of Patient Information? Information Systems Research, 30(4):1184–1202. Publisher: INFORMS.

Kwon, J. and Johnson, M. E. (2015). The Market Effect of Healthcare Security: Do Patients Care about Data Breaches? Workshop on the Economics of Information Security.

Larson, N. (2011). Network security. Working Paper.

Lim, K. (2018). Endogenous production networks and the business cycle. Working Paper.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies, 60(3):531–542. Publisher: [Oxford University Press, Review of Economic Studies, Ltd.].

Marti, C. (2024). Competition and cybercrime. Doctoral Dissertation.

McGlave, C. C., Neprash, H., and Nikpay, S. (2023). Hacked to Pieces? The Effects of Ransomware Attacks on Hospitals and Patients.

McLeod, A. and Dolezel, D. (2018). Cyber-analytics: Modeling factors associated with healthcare data breaches. Decision Support Systems, 108:57–68.

Mongey, S. (2021). Market structure and monetary non-neutrality. Working Paper.

Murciano-Goroff, R., Zhuo, R., and Greenstein, S. (2024). Navigating Software Vulnerabilities: Eighteen Years of Evidence from Medium and Large U.S. Organizations.

New York Times (2021). How a Cream Cheese Shortage Is Affecting N.Y.C. Bagel Shops. Ashley Wong, December 4th, 2021. Link.

Niculescu, M. F., Shin, H., and Whang, S. (2012). Underlying Consumer Heterogeneity in Markets for Subscription-Based IT Services with Network Effects. Information Systems Research. Publisher: INFORMS.

Oberfield, E. (2018). A theory of input–output architecture. Econometrica, 86(2):559–589.

O'Donnell, A. J. (2008). When malware attacks (anything but windows). IEEE Security & Privacy, 6(3):68–70.

Rossi-Hansberg, E., Sarte, P.-D., and Trachter, N. (2021). Diverging trends in national and local concentration. NBER Macroeconomics Annual, 35:115–150.

Rundle, J. (2024). Exclusive: Hackers Broke Into Change Healthcare's Systems Days Before Cyberattack. Wall Street Journal.

Rundle, J. and Stupp, C. (2024). Change Healthcare Hack: What You Need to Know. Wall Street Journal.

Schneier, B. (2008). Schneier on Security. Wiley.

Schneier, B. (2024). A Cyber Insurance Backstop?

Soo Hoo, K. (2000). How much is enough? a risk-management approach to computer security. Dissertation, Stanford Consortium for Research on Information Security and Policy.

Tirole, J. (1988). The Theory of Industrial Organization. MIT Press. Google-Books-ID: HIjsF0XONF8C.

Van Den Berg, G. J. (2001). Duration Models: Specification, Identification and Multiple Durations. In Heckman, J. J. and Leamer, E., editors, Handbook of Econometrics, volume 5, pages 3381–3460. Elsevier.

Wang, O. and Werning, I. (2020). Dynamic oligopoly and price stickiness. Working Paper.

Wisconsin State Farmer (2021). Schreiber foods hit with cyberattack; plants closed. Jan Shepel, October 26th, 2021. Link.