Chitra Marti
March 28, 2024
Competition and Cybercrime: An Application to Healthcare Digitization

---

Abstract

Cyberattacks pose a special kind of supply risk unlike anything we have seen before in two ways: first, attacks are *strategic* choices of malicious attackers, and thus possibly endogenous to market or network structure; second, cyberattacks *scale* at near-zero marginal cost as long as the targets have a common technology product and vulnerability. Both factors suggest that the structure of the market for technology products that facilitate cyberattacks – in particular, how concentrated it is, and the network structure between targets and the providers – affects the final security of the overall industry.

In this paper, I investigate how the technology market structure – both in terms of concentration in vendors as well as in networks – affects cybersecurity outcomes in the healthcare sector. Over the last decade, healthcare providers have embraced digital technologies and cybercriminals have embraced the sensitive health and financial information now at risk. I match two large datasets to form a clear picture of how data breaches are related to the technology products used by healthcare providers.

First, I show that healthcare providers do not appear to be making cybersecurity-driven choices about technology products, instead relying on their GPOs or local market leaders, without much of a response to any exposed vulnerabilities. Next, I show a hospital's choice of technology product does affect their cybersecurity risk in two ways. First, using a vulnerable product simply leads to more breaches. Second, and more subtly, I find a *negative network effect* of using the same technology vendor as other large providers. I show that the negative network effect may be the result of increased incentives for cybercriminals to develop an exploit and attack many providers at once with a scaled attack. I find the gains to hospital digitization are mitigated by the concentrated market that amplifies this negative network effect, which cannot be solved by minimum security policies or cyber-insurance alone. Instead, my findings suggest antitrust policies can play a significant role in securing the entire U.S. healthcare system against cyberattacks and keeping patients' data safe and secure.

Contents

1 Introduction **0**

2 Literature Review **2**
2.1 Literature: Non-Market Consequences of Market Power . . . 2
2.2 Literature: Economics and Cybercrime . . . 2
2.3 Literature: Effects of Healthcare Digitization . . . 3

3 Background: Hospital Digitization and the Regulatory Landscape **4**
3.1 The Beginning: Pre-Information Technology . . . 4
3.1.1 Frequent Errors . . . 4
3.1.2 High Cost of Information Sharing . . . 5
3.1.3 High Cost of Theft and Privacy Violations . . . 5
3.2 The Middle: HIPAA and Barriers to IT Adoption . . . 5
3.2.1 Barriers to the Mass Adoption of Healthcare IT . . . 5
3.2.2 The Passage of HIPAA . . . 6
The Privacy Rule . . . 6
The Security Rule . . . 7
Electronic Transaction and Code Set Standards . . . 7
3.3 The Acceleration: The HITECH Act . . . 8
3.3.1 Promoting Health Information Technology . . . 8
Stage 1: by 2015 . . . 8
Stage 2: by 2016 . . . 8
Stage 3: by 2018 . . . 9
Main Requirements . . . 9
3.3.2 Improving Privacy and Security Provisions . . . 9
3.3.3 Efficacy of the HITECH Act . . . 10
Adoption of EHRs . . . 10
Use of EHRs to Improve Quality and Lower Costs . . . 10
3.3.4 21st Century CURES Act . . . 11

4 Overview of Healthcare Technologies and Implementation Process **11**
4.1 Electronic Health Records . . . 11
4.2 Components of Electronic Health Records . . . 12
4.2.1 Health Information Exchanges . . . 14
4.2.2 Meaningful Use and Certification . . . 14
4.2.3 The EMR Developer Market . . . 15
4.3 The Cost of an EMR System . . . 15
4.4 Security Systems . . . 16
4.5 Other Digital Technologies . . . 16
4.6 Cybercrime in Healthcare . . . 16
4.6.1 Sensitive Information in the EMR . . . 16
4.6.2 Ransomware in Healthcare . . . 17

4.6.3 How Hackers Gain Access to Records . . . 17
Endpoint Attacks . . . 18
In-House Software Attacks . . . 18
On-Premises Server Attacks . . . 18
Scaling On-Premises Attacks . . . 19
Example: Phishing to On-Premises Server . . . 19
Cloud-Based Server Attacks . . . 19
Scaling Cloud-Based Attacks . . . 19
Example 2: A Scaled Cloud-Based Attack . . . 19

5 Data and Summary Statistics 20
5.1 Breaches: Health and Human Services' Office of Civil Rights . . . 21
5.1.1 Location of Breaches: Cyber vs. Physical . . . 23
5.1.2 Causes of Breaches: Crimes vs. Mistakes . . . 24
5.1.3 Examples of Breaches . . . 26
Example of a Cyber-Crime: . . . 26
Example of a Cyber-Mistake: . . . 26
Example of a Physical-Mistake: . . . 26
Example of a Physical-Crime: . . . 26
5.2 Technology: Healthcare Information Management Systems Society Analytics Legacy Database . . . 27
5.3 Matching Process . . . 31
5.3.1 Matching Technologies and AHA Data . . . 32

6 How Do Hospitals Choose To Digitize and Specific Vendors? 33
6.1 Digitization Over Time . . . 33
6.2 The EMR Market Over Time: More Adopters, More Concentration . . . 35
6.2.1 Year of EMR Adoption . . . 35
6.2.2 Drivers of EMR Vendor Choice over Time . . . 37
6.2.3 Importance of Hospital Characteristics . . . 38
6.2.4 Importance of Group Purchasing Organizations . . . 40
6.2.5 Importance of Neighbors . . . 41
6.2.6 Importance of History . . . 43
6.3 The Security Technology Market . . . 44

7 How Has Hospital Digitization Affected Data Breaches? 50
7.1 Does Having an EHR or Security System Predict or Prevent Breaches? . . . 51
7.1.1 Basic Correlations . . . 51
7.2 Duration Model: Time to Breaches . . . 53
7.2.1 Cyber vs. Physical Breaches . . . 54
7.2.2 Mistake Data Breaches . . . 54
7.2.3 Crime Data Breaches: . . . 54
7.2.4 Identification . . . 55
7.2.5 Results: Extensive Margin of Digitization . . . 56
Discussion of Control Variables . . . 58

7.3 Does the Specific EHR Vendor Predict Breaches? 59
7.3.1 Breaches in General: Some Vendors *are* Worse Than Others 59
7.3.2 Cyber vs. Physical Breaches 60
7.3.3 Crimes vs. Mistakes 60
7.4 How Do Hospitals React After a Breach? 61
7.4.1 Hospitals Implement Security Technologies After Breaches 63
7.4.2 Hospitals Do Not Switch EMR Vendors After a Breach 64

8 How Does the Market Structure for Healthcare Technology Affect Cybersecurity Outcomes? 65
8.1 Measuring the Positive Network Effect 66
8.2 Measuring the Negative Network Effect: Scaled Attacks 67
8.2.1 Across States, GPOs, and Other Non-Technology Networks 68
State-Level 68
GPO-Level 70
8.2.2 Through Health Information Exchanges 71
8.2.3 Through Common Technology Vendors and Software Monoculture 73
8.3 Robustness Checks: Instruments 75

9 Conclusions and Next Steps 78

10 References 79

1 Introduction

In 2021, President Biden warned that the next "real shooting war" the United States enters will "more than likely be a consequence of a cyber breach." The consequences of cyber breaches can be simply comically inconvenient: bagel shops in New York City were reported to have hawked tofu-based vegan cream "cheese" after a ransomware attack against the large dairy firm Schreiber Foods stopped regular production.[1] They can also be globally devastating: the 2017 NotPetya attack on Ukraine, allegedly by Russia, shut down ninety percent of Ukrainian firms and the radiation monitoring systems at Chernobyl.[2]

Digital technologies have unlocked a new world economy; to try and capture its effect in a few sentences is futile. However, what digital technology provides, cybercrime threatens to sabotage.

The economic causes and consequences of cyberattacks have been studied far less than its technical detail. Though the effect of a cyberattack is akin to what economists would call a supply shock – reduced production with spillover effects for firms up and down the supply chain – cybercrime is not quite like a meteor falling out of the sky and destroying a plant. Cybercriminals are **strategic** and have the ability to **scale** up their attacks. They specifically target large or well-connected firms with the explicit intention of seizing as much value as possible. Schreiber is the second-largest cream cheese manufacturer in the U.S. and was targeted just as its busy holiday season was to begin. The NotPetya attack was so destructive because it exploited a vulnerability in a tax software, M.E. Doc, that was being used by said ninety percent of Ukrainian firms. By focusing firms with high market share, and then scaling to attack many targets at once, cyberattacks present a new kind of risk. Further, Schreiber was never fined and seems to have not suffered in the market; M.E. Doc faced some charges that were eventually dropped. And, of course, the actual attackers in each case were either not found or not able to be held liable.

Technology markets are considered to be naturally concentrated for two reasons: first, digital technologies experience **economies of scale**: after the software has been developed, the marginal cost of adding one more consumer is near zero. Second, they experience **positive externalities** via network effects: adding a new consumer to the same technology can mean the value of the technology goes up thanks to better interoperability, support, and ability to communicate with more users. How might the (lack of) competition in technology markets affect the proliferation of cybercrime?

The exact characteristics of digital technology that make it uniquely valuable are also those that facilitate the destructive nature of cybercrime. First, cybercriminals also benefit from **(destructive) economies of scale**: after finding and developing an exploit for a particular technology, the attacker can often replicate the attack on new targets until the technology provider develops a patch. If the attacks go undetected, an attacker can cause massive destruction. Second, the addition of a new consumer to a technology creates a **negative externality**: the overall value to the attacker of investing in developing an exploit for a technology goes up the more final consumers there are to exploit, increasing risk for everyone. A concentrated, uncompetitive market creates exactly the conditions for destructive cyberattacks.

In this paper, I focus the discussion on the role **network effects** and **economies of scale** have played in the proliferation of cyberattacks in the **healthcare sector** in the United States.

[1] *New York Times* [62], *Wisconsin State Farmer* [74], *Bloomberg* [15]
[2] *BBC* [13]

The digitization of healthcare has intentions of, as observed in other sectors of the economy, unlocking productivity gains and reducing final costs to consumers. In the U.S. in particular, the the American Recovery and Reinvestment Act of 2009 incentivized doctors and healthcare networks to adopt electronic health record (EHR) systems, leading to near-ubiquity.[3] EHRs serve two purposes: to make data easily accessible to a wide variety of healthcare providers and patients, and to secure the data from malicious attackers.

Cybercriminals, on the other hand, have come to highly value individual healthcare data because – unlike a credit card number or an email address – most of the information it contains cannot be easily changed. For example, insurance information can be used in conjunction with diagnoses to file false claims to write and resell prescription medication; photos can be used for blackmail; and other PII can enable simple identity theft.

Hospital IT spending is at an all time high, with the sector only growing in importance. At the same time, data breaches in the in 2022 affected over 24 million individuals in the United States alone. A single data breach against a hospital has been estimated to cost an average entity $10 million or about $363 per record lost.[4] The conjoined growth of digitization and cybercrime means that as the sector completes its digital transition, without adequate mitigation, risks and losses from cybercrime threaten to devour any productivity gains.

As observed in other technology sectors, digital healthcare services is a concentrated market, subject to the same forces described above. In this paper I ask the following questions:

1. How has hospital digitization affected data breaches?
    1. Are hospitals choosing technology service providers on the basis of security?
    2. How do security outcomes vary with basic hospital characteristics?
    3. What technologies – either as categories or specific products – make hospitals more or less susceptible to breaches?
2. How does the structure of the market for healthcare technology affect cybersecurity outcomes?
    1. Do firms exercise market power to underprovide security?
    2. Is there a *negative network effect* of using the same technology as other hospitals?
3. How effective are commonly proposed cybersecurity policies when we take into account strategic attackers and scaled attacks?
    1. Do security investment minimums just redistribute risk from a strategic attacker?
    2. Can forced diversification in technology (i.e. antitrust) break the effect of scaled attacks?

    This paper, a chapter of my dissertation, addresses Questions 1 and 2. I leave Question 3 for the final chapter of my dissertation, which is excluded from this paper.

[3] Dranove et al. [30]

[4] IBM Cost of Data Breach Report 2023

2 Literature Review

The study of the economics of cybercrime rests 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 particular 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**.

2.1 Literature: Non-Market Consequences of Market Power

Much has been written in economics about increases in market power, markups, and the adverse consequences for consumers. Standard theory shows that market power leads to higher prices and lower output and possibly lower quality (Tirole [71]). While it seems clear that markets have become more concentrated, the effects have not always aligned with standard theory predicting negative effects. Ganapati [38] 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. Other papers, such as Autor et al. [9], 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. [14], Rossi-Hansberg et al. [67]). Many tech firms experience economies of scale thanks to growing networks, resulting in natural monopolies that are highly productive. In particular, Farboodi and Veldkamp [32] show the not necessarily adverse consequences of market power accrued through a data-heavy production process.

In this paper, 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. 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. [45], Wang and Werning [73], Mongey [60]). Geer et al. [39] 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 on each. Jamilov et al. [46] confirm that firm-level cyber risk is positively correlated with firm size.

2.2 Literature: Economics and Cybercrime

The question of how to best secure systems against malicious actors is no longer considered a purely technical question (Soo Hoo [70], Anderson [5], Schneier [68]) 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.

Many papers theoretically characterize the causes and effects **underinvestment in security** when attacks transmit through a network (Galeotti et al. [37], Goyal and Vigier [41], Larson [54], O'Donnell [63], Arce [6]). Notably, Acemoglu et al. [1] introduce the possibility of **overinvestment** by some firms as they attempt to shift the strategic attacker's focus to other firms. I discuss how risk redistribution may occur in my setting in the next chapter of my dissertation.

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. [26] 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 worsened the 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 (Section 6.2.6). In a "pre-mortem" analysis Eisenbach et al. [31] find that an attack on any one 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.

2.3 Literature: Effects of Healthcare Digitization

Hospitals and healthcare providers over the last two decades have transformed the health process by incorporating digital technologies. In doing so, digitization has affected patient outcomes, hospital costs, and, of course, cybersecurity outcomes. I discuss the potential and realized effects on patient outcomes and hospital costs in Section 3.

As expected, breaches adversely affect hospital outcomes (Choi et al. [23]). One might expect a mechanical increase in data breaches and cybersecurity incidents after hospitals digitize, simply because information to remote attacks as described in Section 3.1.3. However, many data breaches are the consequence of human error, either physical or digital. As I show in Section 5.1.1, many breaches are not actual crimes but simply mistakes due to provider errors (e.g. lost papers). McLeod and Dolezel [57] find using simple Logit regressions that that, broadly, increased digitization and connectivity are positively correlated with cyber-driven data breaches but do not address the baseline choice to digitize. Clement [24] 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 [53] 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 dissertation is perhaps most closely related to Kim and Kwon [51], 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. 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. First, mechanically, I use a slightly different measure of EHR adoption, looking within the HIMSS data rather than at the AHA data so I can maintain internal consistency with reports of other technologies use. I also use hospital characteristics from the AHA data, where they are more consistently reported. Second, more substantially, I focus on distinguishing breaches caused by weak products (strategic) from breaches caused by available opportunities (scaled) and offer policy recommendations. Using the HIMSS data allows me to focus on the actual product being used by the hospitals rather than just the extensive margin

of use-or-not. To my knowledge, no study has examined specifically examined the impact of the competitive market structure of healthcare IT services on data breach outcomes, nor estimated the possible impact of counterfactual policies such as security minimums and antitrust activity.

3 Background: Hospital Digitization and the Regulatory Landscape

The American healthcare system is, given the lives and profits at stake, naturally highly regulated. In this section, I describe the regulatory landscape under which healthcare providers choose their technology strategies. I show hospitals were driven to adopt digital technologies not necessarily by to internal cost-benefit analysis, but rather by external regulation and incentives that encouraged adoption. Furthermore, certification requirements and low competition resulted in many hospitals adopting from only a few technology service providers.

3.1 The Beginning: Pre-Information Technology

"Doctor's Handwriting" became a popular metaphor for illegibility for a reason: before widespread IT adoption, individual providers were tasked with taking extensive notes on a patient's condition that were then stored in physical folders. Physical records had two primary issues: first, bad handwriting and manual entries could lead to human errors. Second, the cost of transferring physical records is onerous and increases with the amount of information stored, limiting providers' ability to share patient information and coordinate care especially for the most complex cases.

Physical records did have one great benefit: they were relatively easy to keep safe. A locked cabinet or similarly simple security system meant those who wanted to access records without authorization – whether for malicious intent or for privacy violations – had to individually access each physical folder one at a time. Just like the cost of transfer, the cost of theft was also quite high.

3.1.1 Frequent Errors

The aforementioned "doctor's handwriting" means humans were tasked with note-taking and communication, resulting in simple errors with extreme consequences. A survey of a single large Spanish hospital found 15% of case studies written in the twentieth century were completely illegible (Rodriguez-Vera et al. [66]). A famous malpractice suit in 1999 found a doctor's handwriting responsible for the incorrect prescription that led to a patient's death (Charatan [20]).

Medical errors lead to a variety of adverse outcomes, including but not limited to omitted information, incorrect diagnoses, and adverse drug events.

The Harvard Medical Practice Study, whose results were first published in 1991, remains the preeminent source for understanding the origins of adverse events in hospitalized patients pre-mass IT adoption (Brennan et al. [18], Leape et al. [55]). The study found that 3.7% of inpatient visits at a large New York hospital had an adverse event, of which 27.6% were purely due to negligence. Older and sicker patients were disproportionately affected. The most common type of error was an adverse drug event, though the study does not specify if the issue was with transcription, communication, or otherwise.

In a separate study of such adverse drug events in an outpatient setting, Bates et al. [11] found 6.5% of patients experienced adverse drug events (more than in the inpatient setting), of which 56% were due to ordering mistakes and another 6% were simply due to transcription errors – exactly issues in information communication.

3.1.2 High Cost of Information Sharing

In other sectors, the adoption of information technology has facilitated quick and low-cost communication. One would therefore expect the same in healthcare.

Before mass IT adoption, medical knowledge was essentially stored within the provider. Providers – individually or in a team – were responsible to triaging patient care and checking for adverse consequences. Of course, the complexity of medical care created many related issues. information transmitted from the patient to the provider (e.g. allergies or existing conditions) may be incomplete and therefore not used in diagnosis or treatment. Second, it was costly to transfer information from provider to provider – either by sending physical records or by phone conversation – making care difficult to coordinate. Primary care providers acted as inefficient gatekeepers, resulting in redundant diagnostic tests (Bates et al. [10]). A task as seemingly simple as finding the patients affected by a drug recall would be near-impossible to complete in a timely fashion under purely physical records.

3.1.3 High Cost of Theft and Privacy Violations

In the early part of the twentieth century, before employer-sponsored health insurance was common, the inability of healthcare providers to gather and share information really only affected patients. Thefts of physical records were costly and not particularly lucrative – except perhaps in cases of blackmail – and therefore security was not of significant concern for providers nor regulators. However, as employer-sponsored heath insurance became more popular, a patient's information became newly valuable to parties beyond the patient and their provider: insurers looking to price patients based on their risk levels, and employers who paid much of that price. Patients were therefore understandably concerned that their health information could be used to affect their employment outcomes.

3.2 The Middle: HIPAA and Barriers to IT Adoption

In 1995, Congress passed the Health Insurance Portability and Accountability Act (HIPAA) to address the new role information played in healthcare. HIPAA introduced a variety of insurance mandates, offset with new rules and restrictions on information transmission and confidentiality (Atchinson and Fox [7]). At the same time, digital healthcare technologies were being developed but were not yet adopted *en masse*. HIPAA was enacted during this intermediate period of development, and the rules reflect the technologies' novelty and regulators' uncertainty.

3.2.1 Barriers to the Mass Adoption of Healthcare IT

As described above, digitization seemed to present a perfect solution to many of the common issues in healthcare: the low cost of storage and and speed communication are the comparative of digital technologies over their predecessors (Goldfarb and Tucker [40]).

In 2009 – by when smartphone and fitness tracker market were well-developed – however, only about 11% of hospitals reported basic use of electronic health records (Jha et al. [47]). Clearly, mass adoption by healthcare providers of IT faced barriers.

First, such technologies remained expensive, with significant start-up and maintenance costs that providers, especially in small or rural hospitals, might find unsurmountable. The price of the EHR system alone is negotiated individually between hospitals and the EHR provider and can range from tens of the thousands to tens of millions[5]. A concentrated market means individual hospitals also often do not have much choice. (Dranove et al. [30]) Although the subsidies described in Section 3.3 offset the cost partially, undertaking installation has always been a heavy investment for any healthcare provider.

Second, the structure of health insurance and provider competition in the U.S. means that a provider who uses technology to decrease prices or increase quality is not likely to see a direct strong response in volume to make up for investments (Cutler et al. [27]). Those who pay the cost are not those who see the benefits, which accrue to insurance firms and patients instead of the providers themselves.

Third, research continued to show mixed benefits of digital technologies on patient outcomes and provider cost savings (Buntin et al. [19], Himmelstein et al. [43], Dranove et al. [29], Freedman et al. [35]). The learning curve of new technology means provider productivity may actually decrease instead of increase in the short-term.

Finally, individual providers first needed other providers to also adopt digital technologies to realize the benefits of sharing information – a classical coordination problem for products with network effects. Furthermore, due to the nature of the software used for electronic health records, proper information sharing requires software developers to make records interoperable – which to the developer means lower switching costs and therefore lower market power.

Although such technologies were developed, available, and adopted by a few large flagship hospitals, the turn of the century – and the boom period of IT in other sectors across the economy – was not characterized by massive healthcare digitization.

3.2.2 The Passage of HIPAA

As described above, HIPAA was motivated by the risk of employer-sponsored health insurance and the associated information risks faced by patients, their employers, and insurance firms. While Title I created various insurance coverage mandates, Title II sought to simultaneously protect the confidentiality of patient information and increase insurer efficiency. To that end, the U.S. Department of Health and Human Services (HHS) designed the Privacy Rule, the Security Rule, and the Electronic Transaction and Code Set Standards.

The Privacy Rule The Privacy Rule was published in December 2000. The Privacy Rule defined, for the first time, "protected health information" (PHI) as any health-relevant information and/or personally identifiable information (e.g. a social security number). The rule also specified who was subject to the regulations in HIPAA: "Covered Entities" (CEs) are healthcare providers, health plans, and clearing houses. The "Business Associate" (BA) of a CE is also subject to the law if

[5]Based on proprietary data provided to me by Mordor Intelligence, a healthcare consulting firm

it transacts with PHI in any administrative or legal capacity. The Privacy Rule is enforced by the HHS Office of Civil Rights (OCR).

The main text of the Privacy Rule specified that CEs were only permitted to use and disclose PHI with the relevant individual for treatment, payment, or national priorities with authorization. Individuals also had a right to access their PHI. Additional guidance around privacy policy disclosure, record retention, and administrative requirements were provided to CEs. All CEs were given two years to comply, with an extra one-year period to small health plans. A penalty schedule for violations was also provided, with amounts ranging from $25,000 to $2,000,000 and subject to OCR discretion.

Notably the Privacy Rule did not explicitly distinguish rules for physical versus electronic records, treating all PHI as equal.

The Security Rule While the Privacy Rule did not distinguish the format of PHI, the Security Rule, published in February 2003, specifically took into account the burgeoning importance of electronic health information and sought to address possible issues quickly. The stated goal was to enable CEs to adopt new technologies with liberty while still complying with the basic provisions of the Privacy Rule.

The Security Rule defines "electronic protected health information" (ePHI) as any PHI that the CE uses in electronic form. The required security precautions are divided into three types: administrative, physical, and technical. I will focus here on the technical requirements, which are the primary way the Security Rule differs from the Privacy Rule.

Four "Technical Safeguards" are described in the text of the rule: access control, audit controls, integrity controls, and authentication procedures. It is notable that specific software features were not described in the Rule, as cybersecurity technologies were neither extremely well-developed nor of major concern at the time. Lawmakers were deferential to the CE and to EHR technology services, allowing it to determine how controls would be implemented. For example, the Rule prescribes encryption, but only when it is "deemed appropriate" by the CE. As we will see in the data, it seems CEs often did not deem encryption was appropriate very often.

Electronic Transaction and Code Set Standards The Electronic Transaction and Code Set Standards

The Privacy and Security Rule were designed to protect patient confidentiality and governed access to information. The Transaction and Code Set Standards, on the other hand, was designed to boost efficiency for insurers, creating cost savings that would balance out the coverage mandates of Title I. Specifically, the rule created a standard HIPAA format for electronic data interchange governing a wide variety of transactions (CMS, Social Security Administration). Any software, then, that involved ePHI had to be capable of receiving and transmitting information in the standard format (Report from HHS). The goal of the rule was to facilitate information sharing across providers – key for realizing the gains from digitization – and to encourage competition across technologies through the use of common standards

However, the rule stopped short of mandating true interoperability, e.g. a single file format – only that the information contained in the files be standardized. As we will see, that lack of perfect interoperability remains a key friction in the market for healthcare information technology.

3.3 The Acceleration: The HITECH Act

HIPAA calcified personal health information as private to the patient and the provider and therefore something to be protected. Because it was passed at an early stage of technological development, however, it was naturally limited in its ability to keep up with the changing landscape of both healthcare and health technology.

In 2009, the American Reinvestment and Recovery Act (ARRA) by the Obama administration sought to curb the Great Recession with heavy government investment. Included in the ARRA was the Health Information Technology for Economic and Clinical Health (HITECH) Act, which allocated over $100 billion dollars towards upgrading healthcare systems and subsidizing the adoption of healthcare technologies. As with HIPAA, the hope was that incentivizing adoption of healthcare technology could bring cost savings and efficiencies to the American healthcare system and provide relief to patients and insurers alike. The policy, whose relevant points I describe below, has generally been regarded as successful in accelerating the adoption of healthcare information technologies (Dranove et al. [30], Adler-Milstein et al. [2], Adler-Milstein and Jha [3]), though the outcome of such adoption remains questionable and is the subject of this paper.

3.3.1 Promoting Health Information Technology

The first three sections of the HITECH Act were dedicated to defining exactly how healthcare technologies should be used by healthcare providers to improve care outcomes (GovInfo). Specific uses were defined so that subsidies, grants, and loans could be awarded as healthcare providers met these uses. The most important of these conditions was the "Meaningful Use" (MU) standard for Electronic Health Records (EHRs). If physicians and hospitals could prove they met a set of MU standards, they were eligible for substantial extra Medicare and Medicaid payments. In 2015, those payments turned into penalties for providers who had *not* yet adopted (Blumenthal [16]).

Stage 1: by 2015 In Stage 1, providers and hospitals were required to meet 15[6] Core Objectives and select from 5 of 10 Menu Objectives. The full list of objectives was published widely by government agencies, medical associates, journals, and hospital organizations (e.g. AMA, NEJM). Most hospitals that adopted an EHR system after the HITECH Act created incentive payments did eventually meet the MU standards, suggesting there was not much variation in the exact characteristics of the EHR products nor in hospital discretion on what features to adopt (Botta and Cutler [17]).

The subsidies were significant: on average, a hospital could expect a 3% boost to its Medicare payment and a 2% boost to its Medicaid payment, which comes out to about $2.1 million per hospital per year, and about $5.3 million in total (Dranove et al. [30]).

Stage 2: by 2016 Stage 2 was modified in October 2015 into a combination of Stage 1 and Stage 2 to give providers more time and reduced perceived complexity. In addition to the Core Objectives, additional Menu Objectives and Public Health Objectives were mandated.

[6]The exact number of Core Objectives differed slightly for Hospitals and Individual Providers

Stage 3: by 2018 In 2016, the EHR Incentive Program was renamed the Promoting Interoperability Program (CMS). The list of objectives was again simplified, but no new objectives were added to hospital requirements. Furthermore, while the program ostensibly was meant to direct different EHR software programs to be able to send records easily back and forth, true interoperability remains elusive.

Main Requirements Although the exact wording and list changed over time and by provider type, meeting the following requirements generally meant a provider was able to certify "Meaningful Use" of the EHR System for 80% or more of patients:

1. Protect ePHI, Perform Security Risk Assessment
2. Use Clinical Decision Support
3. Use Computerized Provider Order Entry
4. Generate Prescriptions Electronically (eRx)
5. Participate in Health Information Exchange
6. Provide Patient-Specific Education Resources
7. Perform Medication Reconciliation
8. Provide Patient Electronic Access
9. Use Secure Electronic Messaging
10. Engage in Public Health Reporting

Some of the objectives listed here were initially only in the Menu Objectives, reflecting the program's staircase structure. In Section 4, I describe how EHR software firms and healthcare providers met each requirement. Because adoption of EHRs involves both significant one-time costs as well as ongoing maintenance costs, the subsidies were intended to smooth the transition for providers until they could themselves realize the gains of MU within their systems.

3.3.2 Improving Privacy and Security Provisions

Critical for this project, the HITECH Act also focused on maintaining the privacy and security of patient records.

First, the HITECH Act included in its initial MU list "Implement systems to protect privacy and security of patient data in the EHR," for which it prescribed the provider "Conduct or review a security risk analysis, implement security updates as necessary, and correct identified security deficiencies." As we will see in the data, the vague language nonetheless gave hospitals significant room for deficiency. Hospitals were also required to perform a Security Risk Assessment, which they often could contract out to EHR providers.

More importantly, however, it also for the first time at the national level mandated disclosure of data breaches. Previously, only a patchwork of state laws existed to inform the public about data

breaches involving their PHI. Under the HITECH Act, any Covered Entity – a provider, healthcare plan, clearinghouse, or business associate, as defined in Section 3.2.2 – that discovered a data breach involving more than 500 individuals had to report the breach within sixty days to the HHS' OCR, which would then publish details publicly and conduct an investigation to determine penalties.

Two points on measurement are worth noting here: first, discovery of a breach is not the same as incidence of a breach. As has always been the case in the economics of crime, a crime that goes undiscovered cannot be measured. Second, the penalties determined by the OCR were not required to be published, and therefore most penalties were kept private.

Note that under HITECH responsibility for the protection of patient data always lies *with the healthcare provider* – even if the actual cause was insecure software, for example. That is, services contracted by the healthcare are not considered directly responsible for the protection of data. Therefore, service providers are only incentivized to provide security through (a) the requirements of the software certification process, which over time did grow to include specific security provisions (see Section 4.2.2) and (b) market mechanisms as (at least in a full-information environment) hospitals would likely change contracts or pay less for insecure products.

3.3.3 Efficacy of the HITECH Act

The HITECH Act was generally regarded as quite successful at getting hospitals to adopt EHR systems and attest to their Meaningful Use. Whether or not they actual reaped benefits from EHR systems, however, remains up for debate.

Adoption of EHRs Data from the American Hospital Association (AHA) show that during the period in which the HITECH subsidies were active, meaningful use of the EHR systems among member hospitals did indeed significantly increase, from less than 50% in 2008[7] (before the HITECH Act) to 76% by 2011, to 94% in 2018. Dranove et al. [30] find that given the pre-HITECH trend, not all of the growth can be directly attributed to HITECH subsidies, though they estimate the 2011 level would have been about ten percentage points lower. In general, though, virtually all American hospitals are now in possession of and can attest to meaningful use of an EHR system.

Use of EHRs to Improve Quality and Lower Costs Dranove et al. [29] study the effect of EHR adoption on the eventual costs borne by hospitals. The HITECH Act counted on increased quality, network effects, and cost savings to justify its subsidies. As expected, due to high start-up costs, adopting an EHR system is associated with a rise in hospital costs. If the hospital has access to well-trained staff in an IT-intensive location, the costs were expected to come down after a few years; otherwise, costs remained high.

In one of the first empirical studies, Miller and Tucker [58] find that adopting Electronic Medical Records (EMRs) – the primary target of the 2009 HITECH Act – significantly reduces infant mortality, particularly when combined with other technologies relevant to obstetric outcomes. Ransbotham et al. [65] find use of EMRs reduces the resolution time for malpractice claims, as

[7]Note that the AHA does not include every American hospital, and members are likely to be urban, large, and sophisticated.

digitization smooths the legal discovery process. On the other hand, Parente and McCullough [64], on the other hand, find variation in quality gains by the type of healthcare facility as well as the type of technology; in particular, they find that EMRs alone have only a small effect on a few specific patient outcomes, suggesting the cost involved may not be entirely justified.

The effect of EHRs on quality outcomes remains mixed. Specific tools, such as Computerized Physician Order Entry (CPOE), have been shown to significantly decrease serious medication errors (Bates et al. [12]). A meta-study found that the most commonly studied tools, CPOE and Clinical Decision Support (CDS) found mixed evidence without clear mechanisms through which EHRs were sometimes useful and sometimes not (Jones et al. [48]). Given the level of investment and the comprehensive nature of the programs outlined in the HITECH Act, such mixed evidence would not be expected. For the reasons described in Section 4, full implementation and realization of cost savings and quality gains seems to still remain out of reach for the American healthcare system. Furthermore, new concerns around the exposure of ePHI to malicious actors via digital technologies threaten the few benefits of health IT that have been realized.

3.3.4 21st Century CURES Act

In 2016, the 21st Century CURES Act modestly updated provisions of the HITECH Act (Congress). It explicitly barred "information blocking" in a continued attempt to promote software interoperability. Further, the Act exempted EHRs from FDA approval, instead emphasizing the importance of the ONCHIT's own process. As a result, the certification process for EHRs was transferred away from private certifiers to the ONC. Through regional centers, ONC targets specific hospitals and providers to promote EHR use. Finally, the law banned "information blocking" or the practice of preventing information from being shared easily, though a wide variety of exceptions and an investigative process that required both reporting and an ONC investigation likely rendered the rule less effective. In 2019, the entire incentive payment system was overhauled and integrated with Medicaid quality payment programs.

4 Overview of Healthcare Technologies and Implementation Process

In this section I describe the various technologies healthcare providers have available to them and their purpose in either improving the quality or reducing the cost of healthcare. I use specific terms for technologies from the Health Information Management Systems Society Analytics Legacy Database, described further in Section 5.2.

4.1 Electronic Health Records

The primary technology specified and promoted by the HTIECH Act is a *Electronic Health Records* system. The phrase "Electronic Health Record" (EHR) generally refers to a patient's entire health history possibly across multiple providers, while "Electronic Medical Record" (EMR) typically refers to a single instance of access, e.g. an EMR generated from one appointment or at one single provider. Historically, the terms have been used essentially interachangeably. For consistency, I

use EMR to refer to the specific technology implemented by a single hospital throughout the rest of this paper.

In practice, an EMR software allows each healthcare provider to log into a database where information about the patient. There, information about the patient's medical and diagnostic history, along with personal details and insurance information, can be access and updated by the provider. The installation process therefore requires more than simply purchasing an off-the-shelf software:

- computer access for every relevant healthcare provider (e.g. doctors, nurses, physician's assistants)
- contracts with a EMR firm to install the software on each endpoint
- a server to store the information so it may be accessed from many different endpoints
- participation in a Health Information Exchange (HIE) to share the EMR with other providers and form a truly longitudinal EMR

and other technical components.

4.2 Components of Electronic Health Records

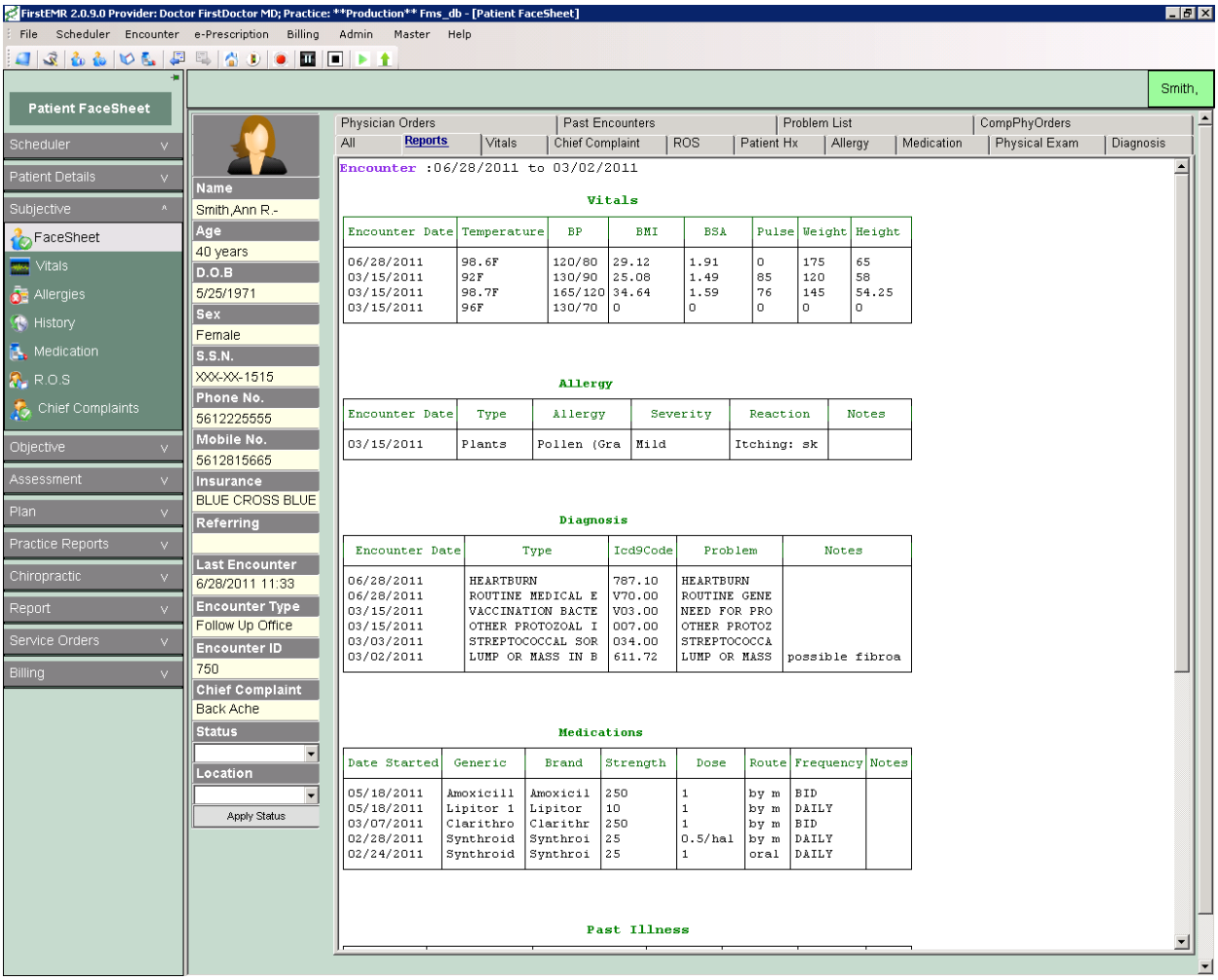
Basic EMR software contains three primary operations (Dranove et al. [29]):

1. Clinical Data Repository (CDR): the storage of medical information that a provider can access; the primary component shared in health information exchanges
2. Clinical Decision Support (CDS): a key use of EMRs, allowing the physician to enter patient information so the machine can search for possible diagnoses, adverse drug interactions, etc.
3. Order Entry: allows a physician to place basic orders.

More advanced components include:

1. Computerized Physician Order Entry (CPOE): a way for a physician to order tests and medications for the patient by communicating with providers at other stages of the healthcare process
2. Physician Documentation/Portal: a way for the physician to store her own information about patients and see information from many patients at once
3. Patient Portal: a way for the patient to access and perhaps even add or correct information in the EMR; patient reminder systems for follow-up appointments
4. Information Sharing: the ability to send information from the EMR to other hospitals, government agencies, insurance companies, pharmacies, laboratories, and more either via the software (with compatible providers) or via a Health Information Exchange
5. Exam Results: automatic addition of laboratory results to a patient's information
6. Secure Messaging: patient to physician communication

Figure 1: Example of a Electronic Medical Record Interface



4.2.1 Health Information Exchanges

A Health Information Exchange (HIE) is a collection of providers who hold healthcare data and agree to share the data across their organizations (ONCHIT). HIEs emerged as an alternative to the software interoperability that remains unprovided by the market. The members of the HIE agree to share data not only for the purpose of care continuity, but also often for public health objectives via aggregate analysis.

A lack of true national interoperability standards has led to the dominance of state-led HIEs. In the U.S., about six states have statewide exchanges with near-full participation of providers in the state (HIMSS). Furthermore,

The cost of infrastructure, the lack of interoperability, and cybersecurity concerns mean in practice the records are not centralized but rather that providers agree to exchange individual records when relevant. Individuals are meant to own their own health record under HIPAA, which again holds back the ability of providers to exchange records without full consent of the patient. In practice, few HIEs have been targeted for cyberattacks, perhaps because attacking individual providers is simpler and contains the same information.

4.2.2 Meaningful Use and Certification

As described in Section 3.3, national policy sought to not only encourage providers to pay the upfront costs of purchasing and installing an EHR system but also their "meaningful use," defining a variety of criteria through which the EMR software could be considered certified and meaningfully deployed by the provider.

Certification had to be proven by the EMR software developer. Developers had to prove they showed the capacity to meet providers' needs as described in Section 4.2 through a variety of tests, including "real world applications." The software needed to be able to store data in a structured manner and meet basic privacy and security safeguards. First, the HHS's Office of the National Coordinator of Health Information Technology (ONCHIT) contracted with a private organization to certify Electronic Health Records systems at the software firm level. The private certification process was replaced by an ONCHIT-specific process by 2014 that was partially outsourced to three firms. The full list of certified EMR providers is made public in a portal made for providers to easily search and compare the characteristics of each EMR system (CHPL).

Meaningful Use, on the other hand, had to be proven by the provider. Providers first had to select a certified EMR software[8], and then prove their meaningful use in three separate stages, though the policies were modified along the way. Meaningful Use required not just use and implementation of the EMR system, but also specific targets to ensure that the provider had fully integrated IT into its workflow. For example, by 2018 Stage 3 requirements included (AAFP):

- transmit more than 60% of prescriptions electronically
- recording more than 60% of medication orders using CPOE
- providing more than 80% of patients electronic access to a portal

[8]In the January 2024 selection shown on the CHPL website, providers had 624 options. However, while many products are available, the developer market is quite concentrated, with the top four firms commanding over 80% of the market share according to calculations shown in Section 4.2.3. Further, not every option is availabel in every location or to every type of provider. The specifics of how providers choose an EMR vendor is discussed in Section 6.

- implement five specific CDS interventions
- retrieve an electronic summary of care for more than 40% of new patient transcriptions/referrals
- submit immunization data to a public health agency

and more. Meaningful Use is meant to incentivize a hospital to truly use HIT in its daily processes. As a result, providers who sought the Medicaid and Medicare subsidies were forced to learn a new technology quickly, leaving, as we will see, ample room for error and incorrect use.

4.2.3 The EMR Developer Market

The market structure of healthcare IT may also influence its efficiency. Dranove et al. [30] find the subsidies contained in the HITECH Act did drive faster EMR adoption; however, because the market is concentrated, much of the subsidies may have simply increased price and consequently firm profits rather than total provision. Hersh and Wright [42] find a large workforce is necessary for proper EMR adoption, suggesting the gains in adoption may be concentrated in large or well-resourced hospitals. Miller et al. [59] use case studies to argue the benefits to small or solo providers are lower.

At the time of this writing, the EMR market is dominated by two players: Epic and Cerner. Cerner was recently acquired by Oracle Corporation, but maintains its products under the Cerner brand. I discuss the demand side of the market in Section 6

4.3 The Cost of an EMR System

Healthcare providers face both significant fixed and variable costs when implementing an EMR system. There is no systematic data on the prices that hospitals pay for EMR systems, for many reasons, such as:

- firms may avoid publishing prices to preclude being undercut by competitors
- prices are often negotiated individually between the technology firm and the healthcare provider (or a group purchasing organization)
- products are also often customized to the needs of the healthcare provider, especially for large hospitals and systems
  - on-premises (or licensed) software typically has a one-time fee paid upfront with more customization
  - software-as-a-service (or cloud) software typically has an ongoing subscription fee with less customization
- both startup and variable costs may depend on the number of providers or the cost of associated labor, which varies across regions and over time

The ONC-HIT CHPL does seek to promote price transparency for healthcare providers to understand the product characteristics of their EMR systems. However, specific cost information is not required to be posted. Although providers were mandated to disclose all additional fees they might charge as part of the certification process, in practice they only required to disclose the *types* of costs and fees – not the actual levels. In a few cases, firms disclose those prices (e.g. Epic) but most simply offer a list of possible additional service fees with levels hidden for the reasons above.

The ONC also provides the "Vendor Pricing Template" is meant for healthcare providers to share with potential contractors to understand total costs. For example, Vendors are asked to specify if they are operating a Licensing Model or a Cloud Model, and then provide estimates across 42 separate possible categories of fees, including implementation and training fees across each specific component – the portal, lab interface, HIE connection, State Immunization Registry, e-prescribing capacities, hosting, patient education, and more. Costs are typically per provider (i.e. per physician in a hospital), and also per month in the Cloud model. Again, levels are not provided – the spreadsheet is only a tool to help hospitals uncover otherwise hidden fees.

4.4 Security Systems

Both the Certification and the Meaningful Use process each contained privacy and security requirements following the HIPAA Privacy and Security Rules (Section 3.2.2). The specifics were typically left up to the software, with encryption suggested but, as we will see in the data, seemingly not enforced. An update to the certification process in 2015 required that end devices be encrypted and that the software have multi-factor authentication available to the provider.

However, to be considered Meaningful Use, the provider only had to ensure data was encrypted, meaning not every provider implemented the full range of security controls that were actually available in the software. For example, the hospital had to independently install firewalls, spam filters for its email systems, its own server backups, mobile device management, and anti-virus software.

4.5 Other Digital Technologies

Of course, other machines and tests are run that enable healthcare providers to collect and share information. Most information generated by a machine is translated to a workable format using the codes described in Section 3.2.2 and then loaded into the patient's EMR. Such machinery, especially when internet-enabled to communicate with EMR systems, is primarily subject to ransomware-style attacks; however such attacks do not appear in my data as they do not often contain or relate to patient data on their own.

4.6 Cybercrime in Healthcare

4.6.1 Sensitive Information in the EMR

The EMR, as given by the simple example in Figure 1, contains a wide variety of information that a malicious attacker might find valuable:

- sensitive health information: possible to blackmail, either *en masse* with a general threat or specifically against prominent individuals

- prescription information: paired with individual names can be used to write fake prescriptions for controlled substances, which are then filled and resold
- insurance information: paired with individual information
- personally identifiable information: e.g. a social security number, used for common identity theft
- phone numbers and email addresses: used for follow-on attacks against individuals

4.6.2 Ransomware in Healthcare

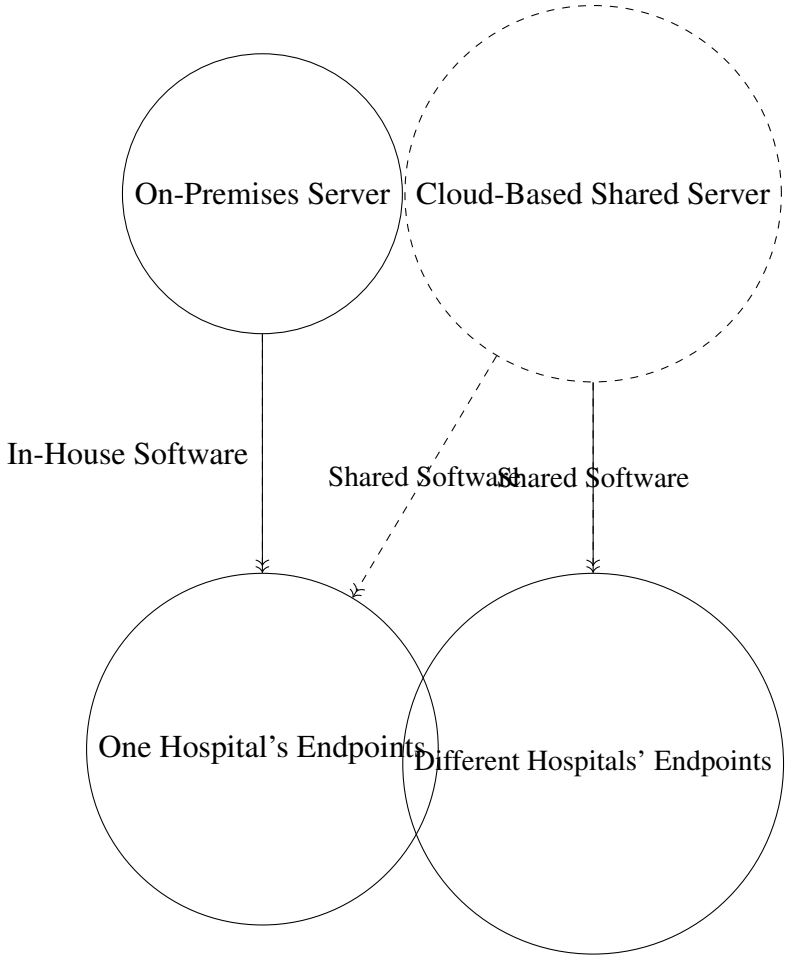
Furthermore, as a crucial piece of software used by providers in every patient interaction, access to the software itself is valuable. A cyberattacker who remotely accesses the network server of the provider can (threaten to) shut down access to the software, corrupt or delete data, or simply publish everything and demand a ransom.

4.6.3 How Hackers Gain Access to Records

In general, providers access a patient's records through some kind of **Endpoint**: a physical computer that they log in to which allows them to access the records stored in a server. To access the records, they require *Software*

There are two main kinds of servers: on-premises and cloud-based. If the hospital has contracted with an *on-premises* provider, then the data are stored also in an on-premises server, and the hospital is fully responsible for the security of that server.

For cloud-based software, the Endpoint is again how providers access the record. However, the software also provides a connection to the server, stored remotely as the responsibility of the software developer instead. The ONC Handbook describes cloud-based EMRs as having "shared responsibility" for data security between the software provider and the client (i.e. the hospital). On-premises servers, meanwhile, must be secured exclusively by the end client.



An attacker therefore has multiple options when seeking to access electronic medical records. If they are stored in an On-Premises server, the hacker can seek to gain access to an Endpoint to gain access to the server.

Endpoint Attacks Attacking an endpoint simply requires somehow acquiring the authentication of the system that would allow a physician to access the records and extracting large amounts of information quickly. Staff might have reused passwords compromised in other hacks. **Phishing** emails sent to particular physicians or support staff can lead to malware that traces keystrokes and discerns passwords. **Physical** attacks are always possible and entail simply stealing a laptop or a desktop computer.

In-House Software Attacks Once the attacker has access to the endpoint, vulnerabilities in the Software enable the attacker to access the on-premises server. For example, if a physician's password alone is enough because the software does not have multi-factor authentication or because the physician can access way more records than would be necessary for business. Most Endpoint attacks can be prevented with *properly secured software*.

On-Premises Server Attacks After using the software to get from the endpoint to the server, the attacker can then choose from a number of malicious actions: extracting records, corrupting

the data, or installing ransomware are the most common. Aside from the physical attack on an endpoint, attacks can be done **remotely**, as described in Section 5.1.1.

Scaling On-Premises Attacks On-Premises attacks scale when the software connecting the endpoint to the server has vulnerabilities that can be repeatedly exploited across providers. That is, the attacker selects a software, and then attacks the providers one-by-one, e.g. by sending phishing emails to multiple organizations. The cost of the second attack is presumably lower as the vulnerabilities have already been found and exploits already developed. Here, *multiple* attacks are required, but with low marginal costs.

In Section 7 I discuss the consequences scale on the basis of common software providers: providers may not know that an attack on a different entity makes them more vulnerable. As a result, the software and the provider remain open to *new attacks* on *other* providers who happen to be using the same software.

Example: Phishing to On-Premises Server *On June 3, 2019, Rosenbaum Dental Group, the covered entity (CE), discovered ransomware on its computer system that encrypted the CE's electronic medical record system. The breach affected 1,208 individuals protected health information (PHI), including patients' first and last names, dates of birth, chart notes, financial information, insurance information, pharmacy numbers, and the names of other family members who had received services from the CE. The CE provided breach notification to HHS, affected individuals, and the media. The CE investigated and determined that the breach was caused by malware that was activated when an employee clicked on a fraudulent computer pop-up window. In response to the breach, the CE improved its process for saving backups of its computer system, removed internet access from computers located in the patient treatment area, changed malware detection vendors, developed and revised its policies and procedures, and trained staff members on the new policies and procedures. The CE also trained employees identifying malicious computer pop-up windows and phishing emails. OCR obtained assurances that the CE implemented the corrective actions listed above.*

Cloud-Based Server Attacks As with On-Premises software, the physician access data through an endpoint and various authentication schemes. The difference is storage: the data is not stored "locally" but rather in a remote location. Furthermore, the same software provider can serve multiple providers at once and store all their data in the same cloud-based server. As a result, the a hospital's data are exposed not just from attacks on endpoints in their own systems but also in attacks on other systems or on the cloud provider itself.

Scaling Cloud-Based Attacks Cloud-Based attacks scale well as the acquired target is not an individual provider's records but could contain the records of multiple organizations – anyone using the same software provider is at risk of having their records breached. As a result, a *single* attack can breach multiple providers. The costs of the attack are ultimately borne by the individual provider. In Section 7 I develop a methodology for detecting scaled attacks in partial data.

Example 2: A Scaled Cloud-Based Attack It is important to note here that the individual provider might not realize the issue is with the *software provider's ability* to protect the server. To

the provider, it may just appear to access is blocked. For example, an attack on Eye Care Leaders, an EMR firm specializing in eye care practices, experienced an attack on its cloud-based server with which it served every client (Alder [4]). However, the reports from individual providers who experienced attacks do not all name Eye Care Leaders nor cite the cloud-based server attack as the issue:

Summit Eye Associates, the covered entity (CE), reported that its business associate (BA) experienced a cyber-attack that compromised the protected health information (PHI) of 53,818 individuals. The PHI involved included names, dates of birth, medical record numbers, health insurance information, Social Security numbers, and other treatment information. The CE notified HHS, affected individuals, the media, and posted substitute notice on its website. Following the breach, the CE terminated its business relationship with the BA. In its mitigation efforts, the CE established a call center and provided complimentary credit monitoring and identity protection services to those affected.

5 Data and Summary Statistics

The study of data breaches in healthcare is facilitated by the extensive data available. Commonly, empirical studies of cybercrime suffer from 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. [34]) or for fear of follow-on or copycat attacks (Choi et al. [21]). 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

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 a survey of the technology vendors used by the breached entity, I can plausibly discern the cause of the attack and move towards assigning fault – key for my counterfactual analyses.

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*.

Empirical studies of cybersecurity are therefore deeply 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 [69]). The challenge of this paper is then drawing inferences on how and why hospitals are experiencing data breaches when explicit data is not available.

To do so, I combine two datasets that each provide difference sides of the equation. First, Section 5.1 describes the *breach* information, the outcome of cyberattacks. The data described mitigate *Disclosure Bias* due to the national mandate.

Next, Section 5.2 describes the 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).

5.1 Breaches: Health and Human Services' Office of Civil Rights

Comprehensive data on breaches of healthcare information comes from from the U.S. Department of Health and Human Services' Office of Civil Rights. This administrative dataset contains every breach reported to the OCR from 2009 to the present[9], which under the HITECH Act is mandatory for any breach affecting more than 500 individual records (as described in Section 3.3.2). The full data contain over 5,000 breaches across healthcare providers, insurance firms, clearinghouses, and auxiliary providers. I focus on breaches affecting hospitals and doctors offices, excluding dentists, pharmacies, rehabilitation centers, and other non-hospital providers.

Often, breaches are reported not because records have actively been breached but because there is a *possibility* of a breach. Further, breaches may sometimes only be discovered after law enforcement or a patient notice a separate issue. Therefore, while Disclosure Bias is mitigated, it is not *eliminated* – hospitals still have to know they were breached.

The data contain the following pieces of information:

- affected entity (by name)
- entity name and state
- breach type: hacking/IT incident, theft, loss, unauthorized access or disclosure, etc.
- location of breach: network server, desktop, laptop, paper, electronic record, etc.
- number of individuals affected
- description in words (not always given)

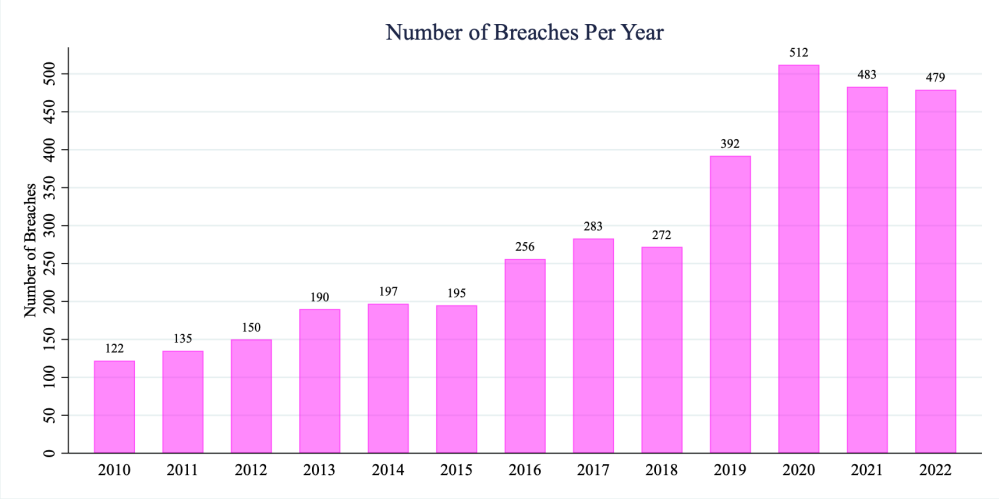
[9] I use the full calendar years only in this paper, 2010-2022.

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

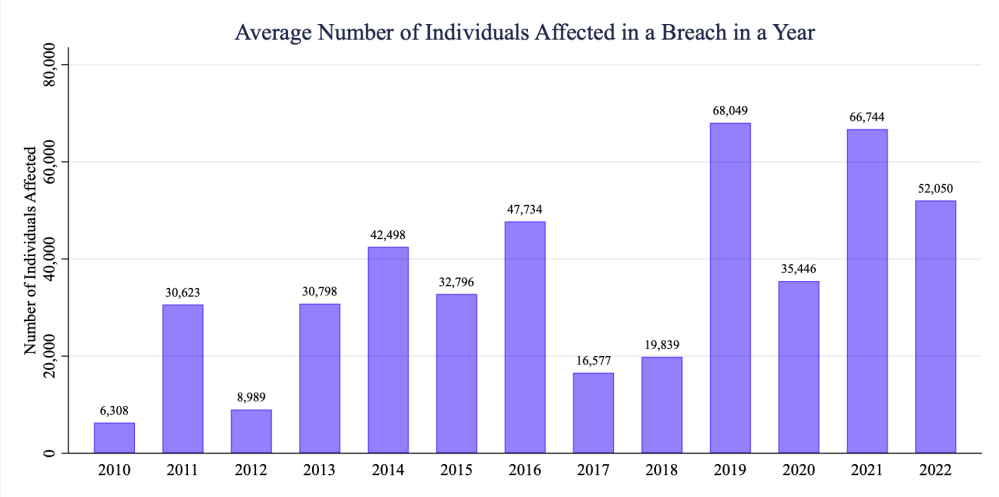
Source: Full Health and Human Services Office of Civil Rights List of Breaches as of August 2023. The data used in this table cover all breaches across any healthcare provider. The data used in the analysis later removes some breaches, a process described in Section 5.3

Figure 2: Total Breaches Over Time



]

Figure 3: Total Breaches Over Time



Figures 2 and 3 show that not only has the number of breaches increased over time, but so has the number of individuals affected. Table 2 shows the average number of individuals per breach has also increased broadly over time, though not always linearly. In particular, some outlier breaches – such as the Anthem 2015 data breach, the largest in the dataset –

5.1.1 Location of Breaches: Cyber vs. Physical

The raw data lists seven possible "locations" of the breach:

1. Desktop Computer
2. Electronic Medical Record
3. Email
4. Laptop
5. Network Server
6. Paper/Films
7. Other Portable Electronic Device

I use the description in conjunction with the reported location to determine if a breach was 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," "Electronic

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 [51] rely on the same manual labeling process).

5.1.2 Causes of Breaches: Crimes vs. Mistakes

The raw data lists seven "types" of breaches, with the possibility for multiple types:

1. Hacking/IT Incident
2. Improper Disposal
3. Loss
4. Theft
5. Unauthorized Access/Disclosure Unknown
6. Other

I also use the description in conjunction with the listed "type" of the breach to determine if a breach was 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 cybercriminal 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.

Figures 5, 4, 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: Number of Crimes vs. Mistakes in the HHS Data

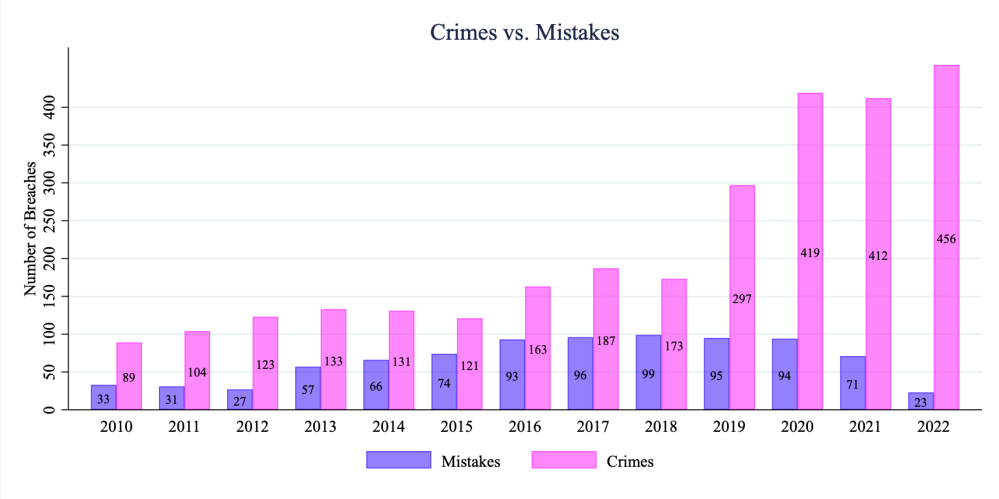


Figure 5: Number of Cyber vs. Physical Breaches in the HHS Data

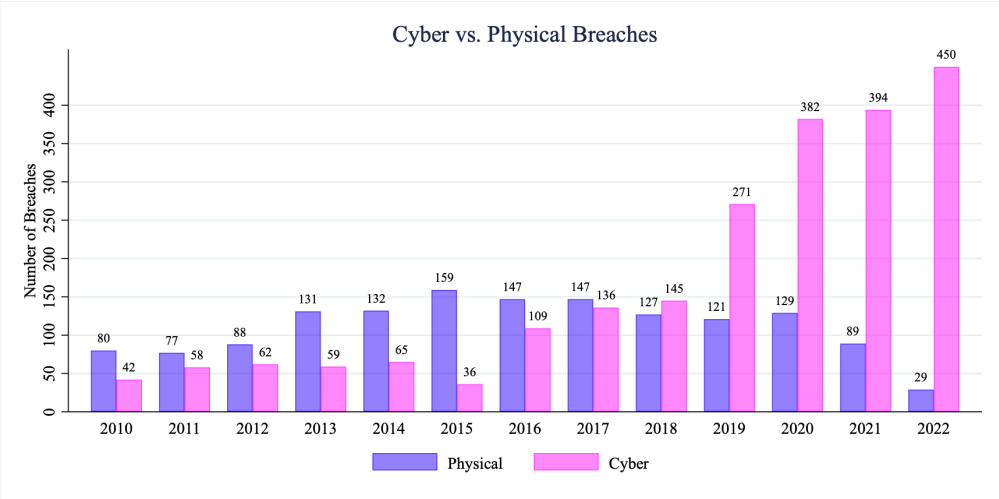
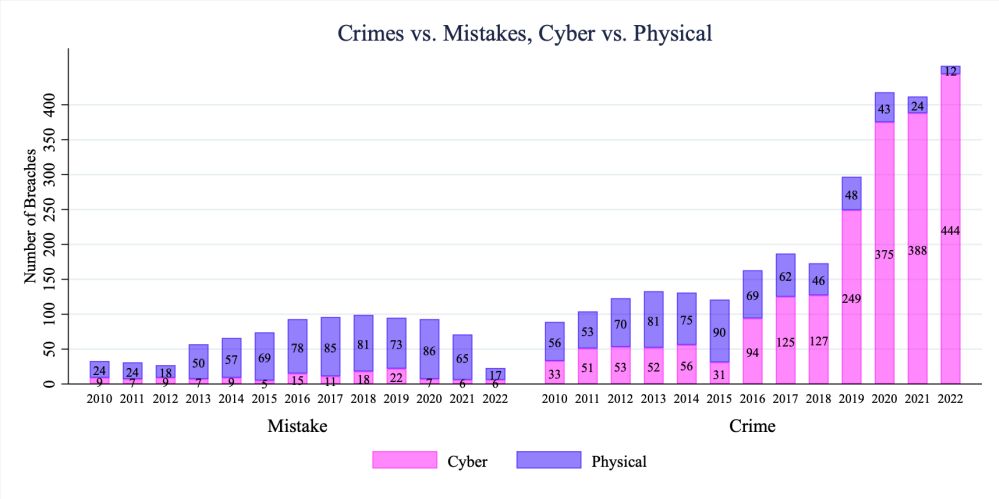


Figure 6: Number of Crimes vs. Mistakes *and* remote vs. local Breaches in the HHS Data



5.1.3 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."*

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."*

5.2 Technology: Healthcare Information Management Systems Society Analytics Legacy Database

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[10] 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.

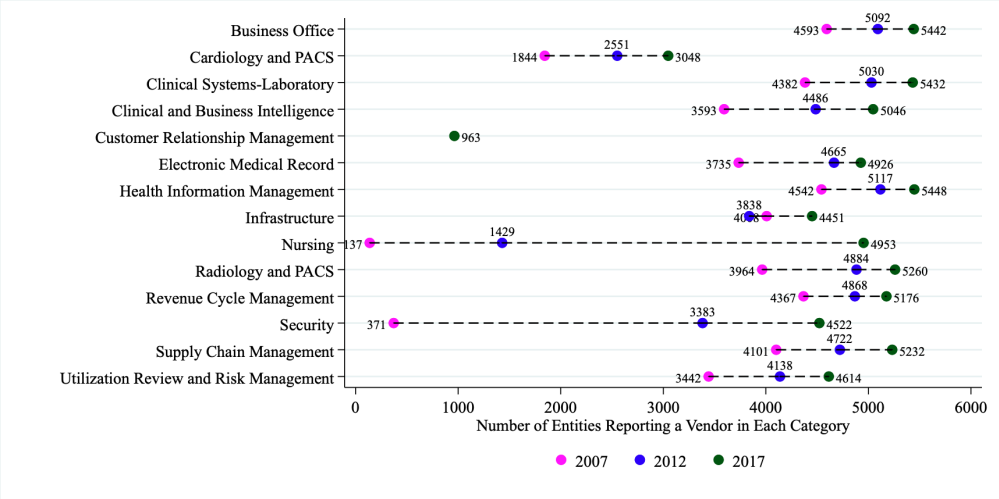
Information about a hospital's technology choices will help us to better *Assign Fault* in the case of breaches. The next several figures illustrate the type and the level of detail of data available. In each, I show the number of hospitals that reported using a digital technology in the relevant category over time.

First, Figure 7 shows that more hospitals are reporting using more technologies in more categories over time – i.e., digitization is increasing. Note that the increase is not uniform across categories; some categories, such as Telemedicine or Health Information Exchange, saw a greater increase than others.[11]

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

[11]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



I focus the majority of this paper on three specific categories: Electronic Medical Record, Health Information Management, and Security. Figures 8, 9, and 10 to 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.

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

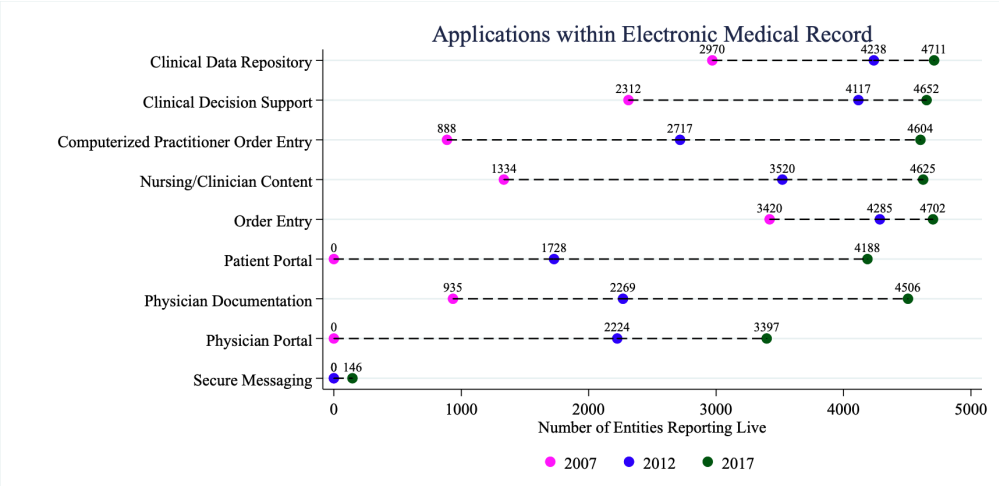


Figure 9: Within a Category, Specific Applications: Health Information Exchange

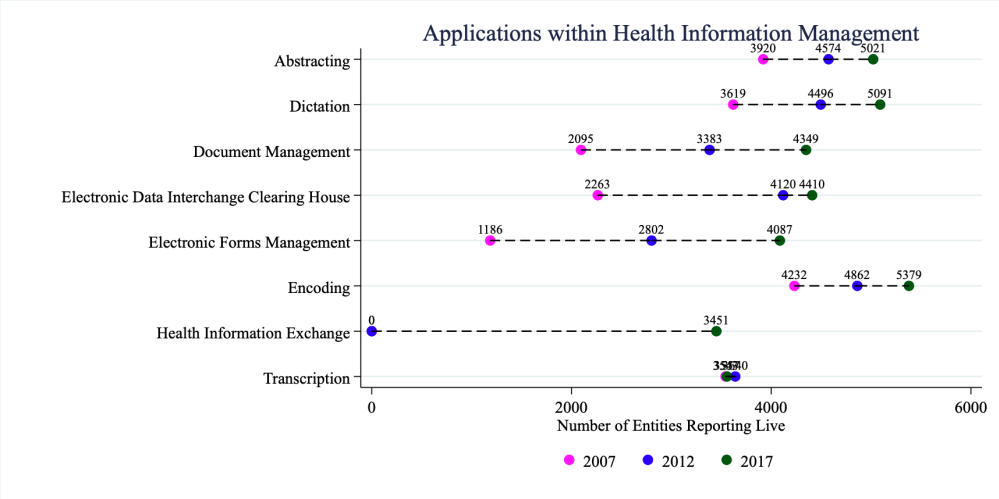
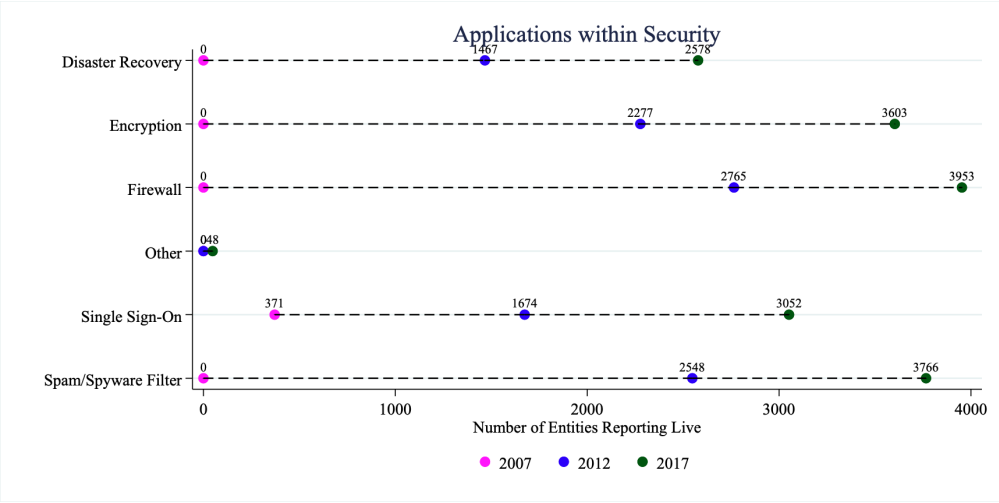


Figure 10: Within a Category, Specific Applications: Security



Next, the survey data 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 across providers, within each application.[12] There is variation both within years and across 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.

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

Figure 11: Vendors within Application: Clinical Data Repository

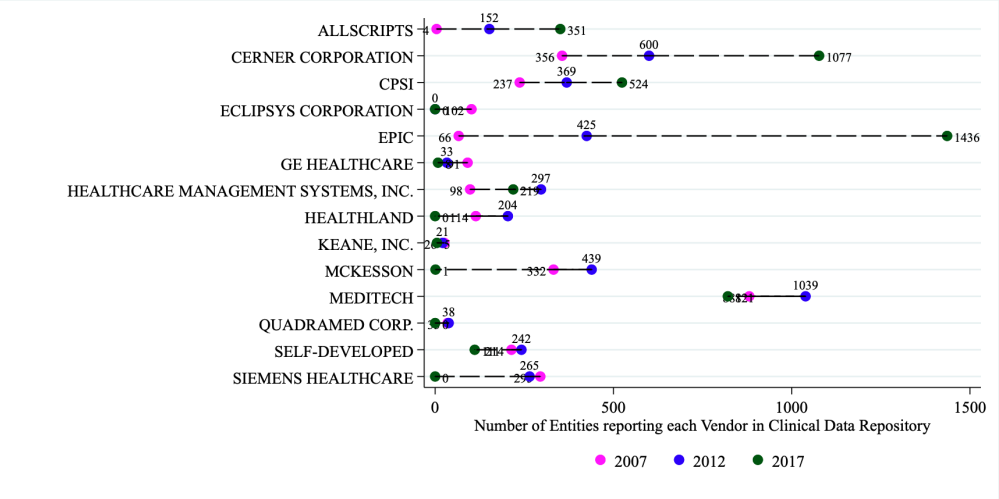
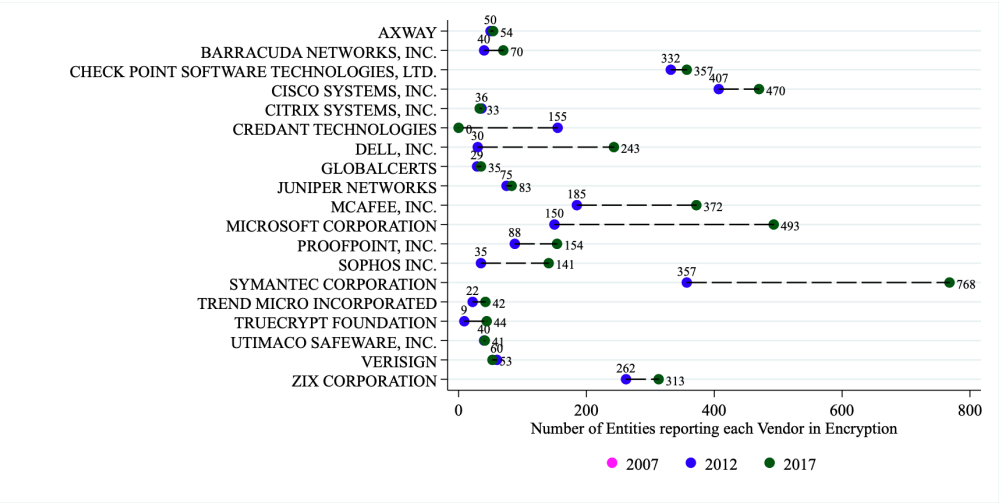
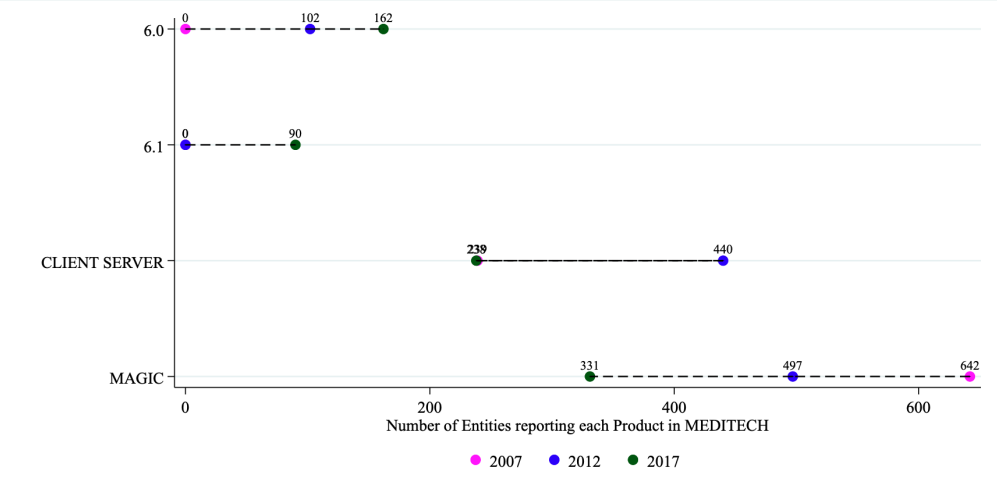


Figure 12: Vendors within Application: Encryption



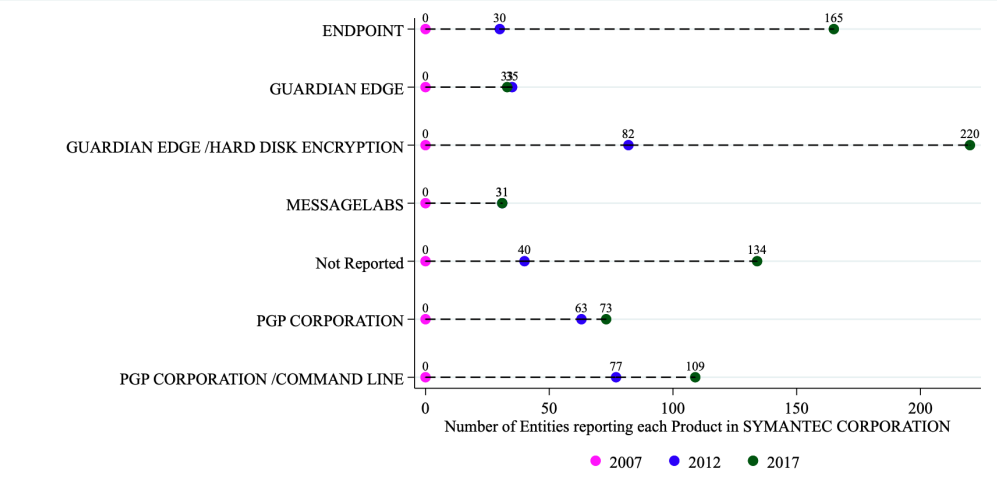
The data contain even more specific details. Within each vendor, we can see which *products* are chosen by specific healthcare providers. For example, Figure 13 provided by Vendor *Meditech* within Category *Electronic Medical Record* and Application *Clinical Data Repository*. Over time, Magic and Client Server have been replaced by 6.0 and 6.1.

Figure 13: Products within Meditech, Clinical Data Repository



Similarly, Figure 14 shows the products provided by Vendor *Symantec* within Category *Security* and Application *Encryption*. As we think about common vulnerabilities between users of the same vendor or product, such detailed information will help us construct the network structure of each technology market.

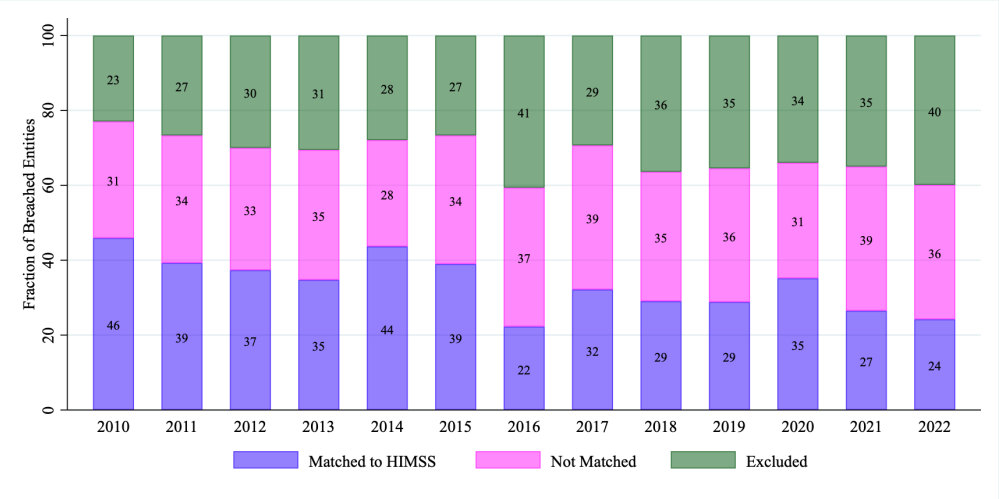
Figure 14: Products within Symatec, Encryption



5.3 Matching Process

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

Figure 15: HHS to HIMSS Data: Fractions Matched and Excluded



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 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.

5.3.1 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.[13]

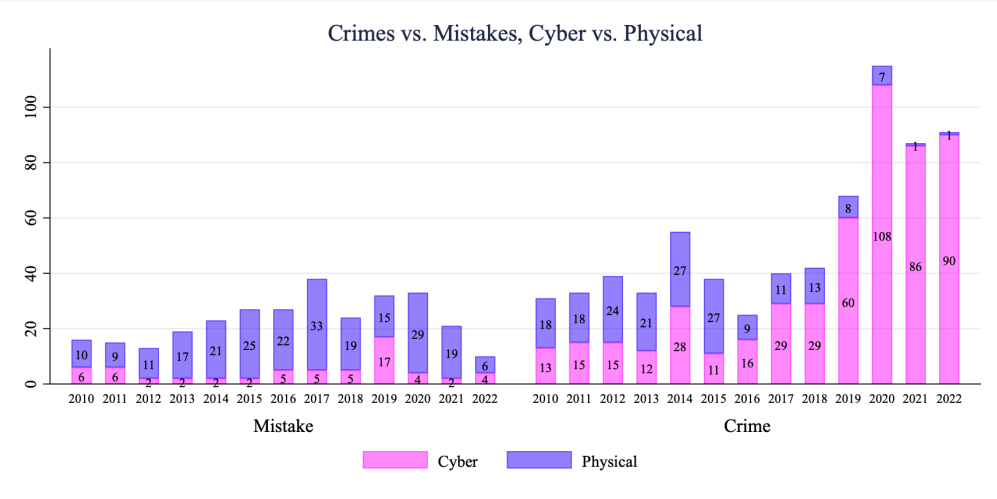
Further, I also only focus on hospitals who report their technologies to the HIMSS Survey. The survey is voluntary and hospitals may elect to only respond to part of it. I only include hospitals that report having at least one Vendor in both the EMR and the Security application categories. Although this does exclude hospitals that do not adopt either, note that separate estimates from the AHA data suggest that 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.

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

All HIMSS	All AHA	Both
4968	5136	4214

[13]The authors of Gabriel et al. [36] provided me with a crosswalk matching part of the AHA and HIMSS data, for which I thank them.

Figure 16: HHS to HIMSS Data: Matched Crimes vs. Mistakes, Cyber vs. Physical



6 How Do Hospitals Choose To Digitize and Specific Vendors?

I begin this analysis with a detailed description and taxonomy of *how hospitals choose whether or not to digitize and their specific technology products and vendors.* As described in Section 3, hospitals' choices are constrained by the variety of regulations involved. After the HITECH Act, providers received reimbursement only if they used a Certified EHR from a pre-specified list. Furthermore, hospitals are often members in Group Purchasing Organizations (GPOs), where groups of hospitals act collectively to negotiate prices with sellers of hospital equipment, including technologies. Finally, the markets for these technologies are themselves concentrated, limiting choices.

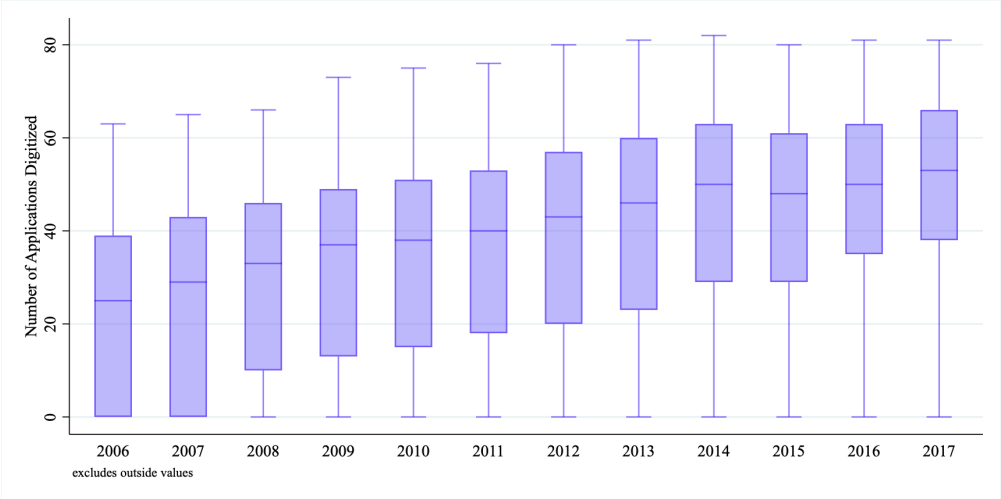
In this section, I show that regulation, peer influence, invariant hospital characteristics, and concentrated markets drive hospital choices of technology vendors, rather than specific product characteristics, cybersecurity concerns or evaluations. That is, hospitals do not seem to *explicitly* considering security when making their vendor choices, motivating the next section on how their choices nonetheless do affect cybersecurity outcomes.

6.1 Digitization Over Time

First and most obviously, I show that hospitals have increased their digital capabilities over time. Figure 17 shows the number of applications that each hospital in the final sample has digitized in each year. An application is defined at the level in the HIMSS data shown in Figure 8, e.g. "Clinical Data Repository" within the broader category Electronic Medical Record.

More interestingly, there remains significant variation over time: some hospitals are heavily digitized early on in the sample period while others remain relatively un-digitized (in number of applications) even by 2017, well after the HITECH subsidies turned into penalties.

Figure 17: Number of Applications



Looking at specific technologies covered by HITECH, we see the same pattern of increases over time in Figures 18 and 19.

Figure 18: Number of Applications: Electronic Medical Records

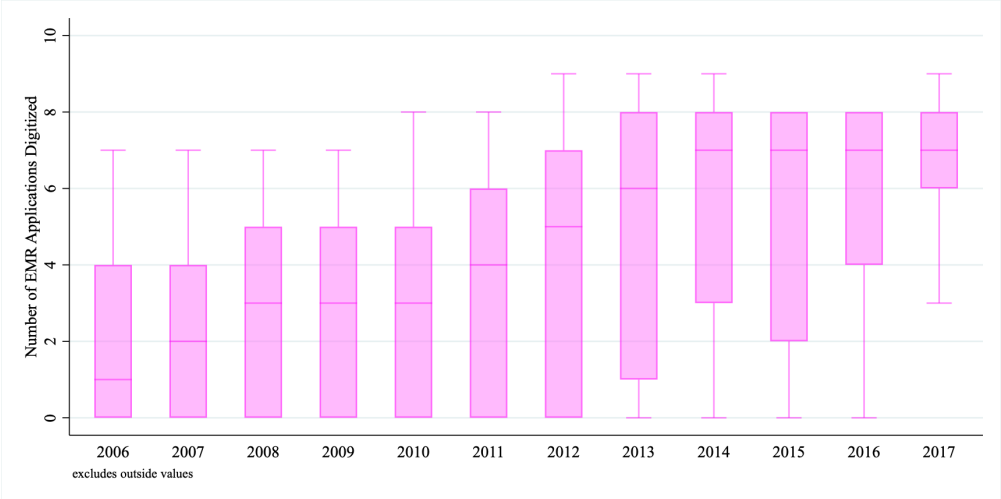
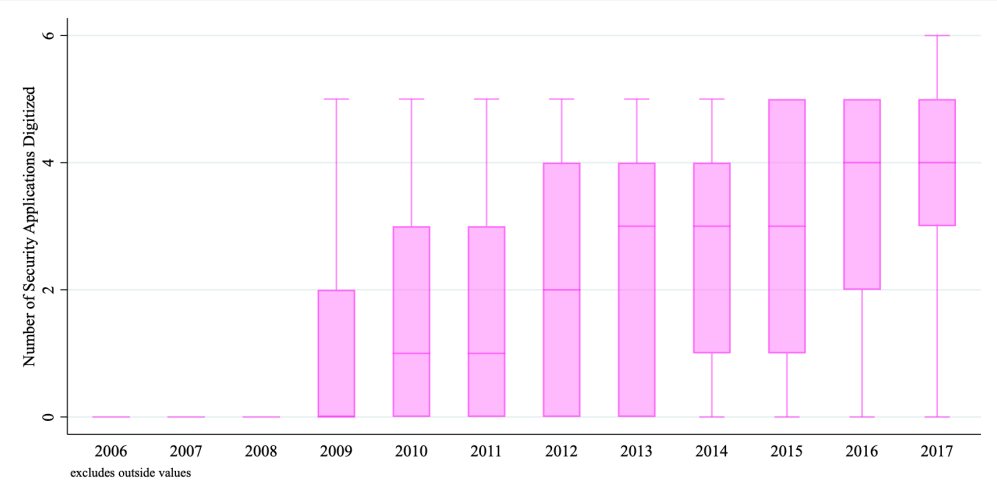


Figure 19: Number of Applications: Security



6.2 The EMR Market Over Time: More Adopters, More Concentration

In this section I show how the adoption of new technologies, measured as the first year a contract is signed (or a survey response is recorded) indicating the hospital has a vendor providing a particular application. I begin, as is standard, with a discussion of EMR adoption (Dranove et al. [30], Kim and Kwon [51]), and then – new in this study – move to a more general discussion of technology adoption, including security technologies.

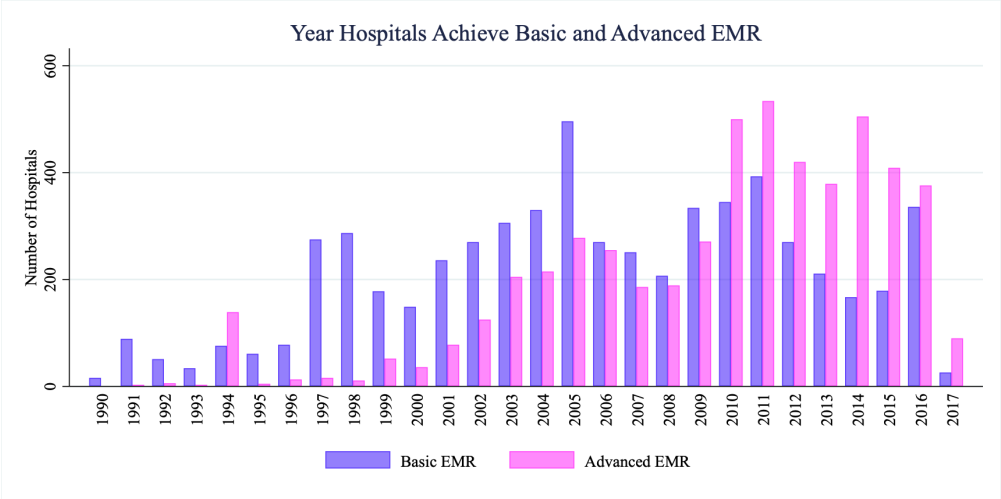
6.2.1 Year of EMR Adoption

I first separate out specific EMR functions into "Basic" and "Advanced" functions (Dranove et al. [30], Kim and Kwon [51], Dranove et al. [29]). Following these papers, I label a hospital has having a Basic EMR if it has at least two of the three out of Clinical Data Repository, Clinical Decision Support, and Order Entry. A hospital has an Advanced EMR if it has either a Computerized Practitioner Order Entry or Physician Documentation listed in the HIMSS data. Everything else I consider "Extra."

Figure 20 shows the year hospitals in the HIMSS data achieve their first Basic EMR and Advanced EMR status. As Dranove et al. [30] found, the trend of EMR adoption preceded the HITECH subsidies and may have reached high levels even in the absence of subsidies.[14]

[14]Note that I only allow hospitals that don't report a Contract Year or Implementation Year to have 2006 as their first Survey Year, since the survey I use starts in 2005.

Figure 20: Year of EMR Adoption: Basic vs. Advanced



Within each Phase, Basic and Advanced, the vast majority of hospitals have most functions implemented seemingly jointly – we do not see significant differences between the bars within each year (Figures 21 and 22). That is, hospitals are probably implementing many EMR functions at once. The key takeaway is that hospitals simply digitize over time, rather than monolithically or all at once.

Figure 21: Year of Basic EMR Function Adoption

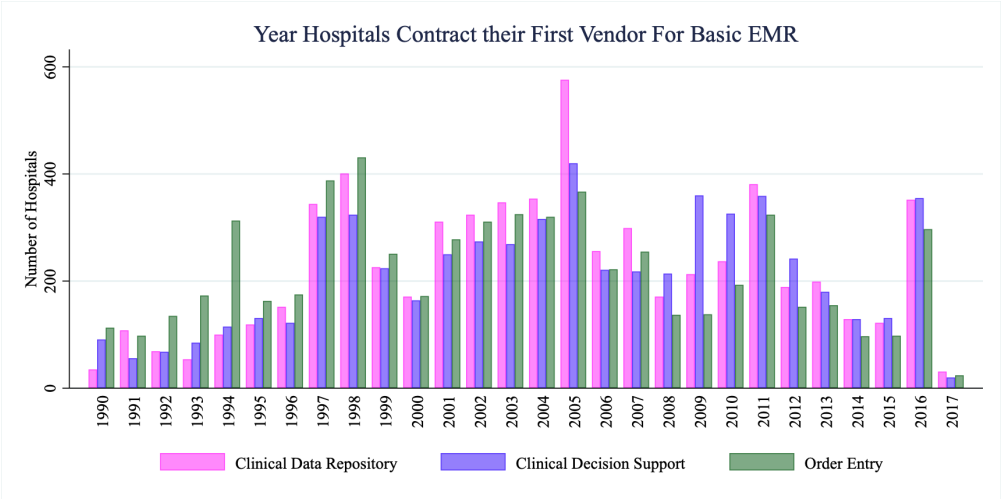
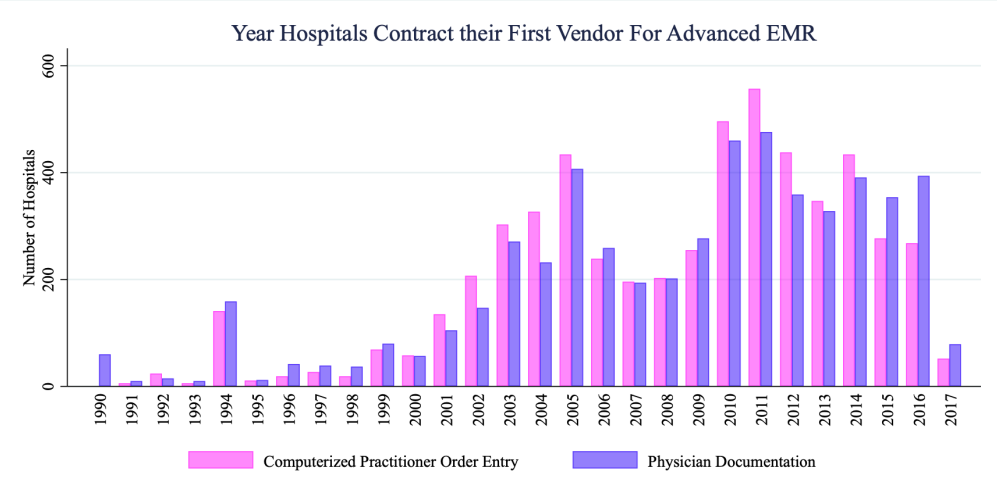


Figure 22: Year of Advanced EMR Function Adoption



6.2.2 Drivers of EMR Vendor Choice over Time

Kim and Kwon [51] focus on the extensive margin EMR adoption status as a predictor for a hospital experiencing a breach. I focus on the intensive margin – the identity of the EMR software used by breached hospitals – and on other technologies that may be involved in a data breach.

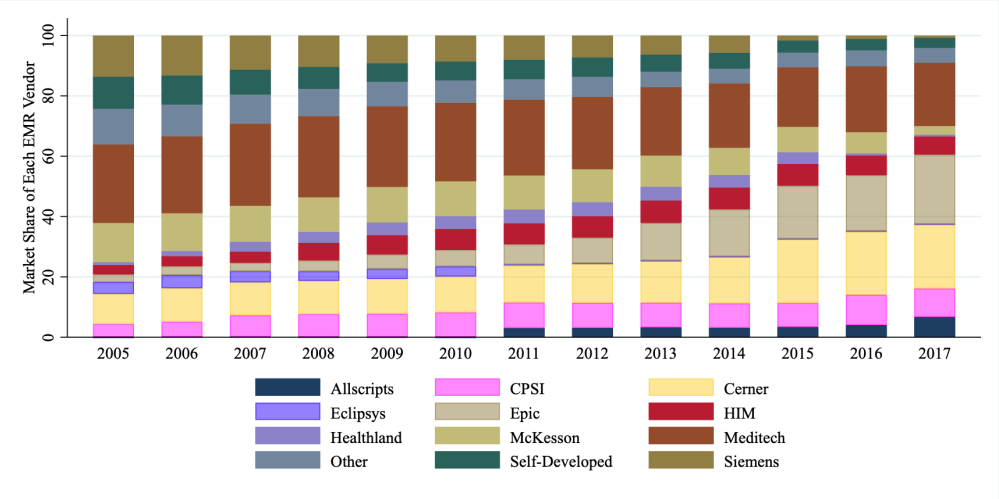
For the reasons I discuss in Section 4.3, there are no systemic data on the prices hospitals pay for their EHR systems, nor for any of the other technologies in the HIMSS data. I thus instead focus on non-price drivers of hospital choice for EHRs: without the price of the system, can we still identify patterns in hospital choice? As I show in Section 7, the identity of the EHR vendor maters for the ultimate security outcomes of the hospital.

Figure 23 shows the market share of the main EMR vendors over time. First, note that these firms are (in the sample period[15]) primarily EMR vendors – not necessarily large conglomerates that offer general purpose software but ones offering health-specific products. Third, many hospital originally reported using self-developed software, but by the end of the sample period, very few do.[16]

[15]Cerner was acquired by Oracle in 2022, which is outside the time period of interest in this paper.

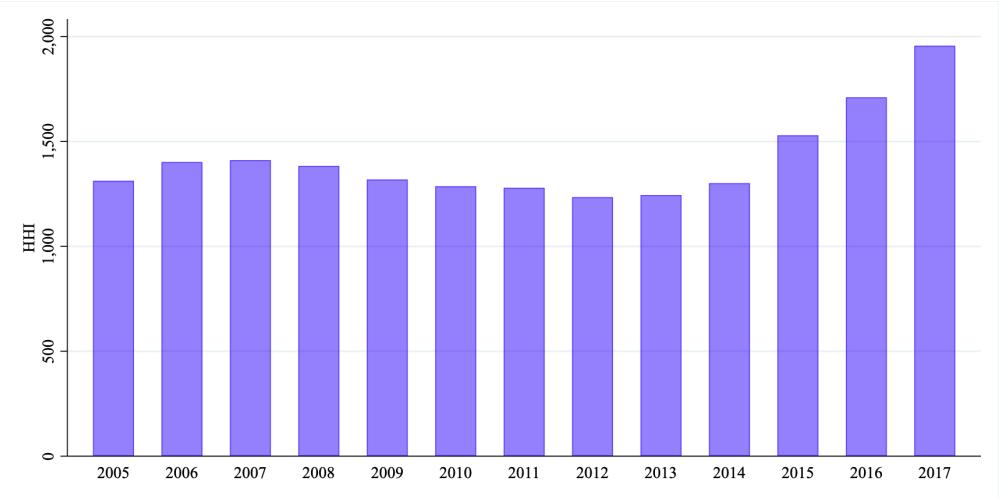
[16]When a hospital reports having more than one contract within the broad category "Electronic Medical Record" I choose the one that provides the most of its specific applications, e.g. if Epic is listed as providing CDR, CDSS, and CPOE, while Meditech provides the Patient Portal, I select Epic. About 20% of hospitals report more than one vendor in the broad Electronic Medical Record category, and less than 2% report more than one vendor in a specific application.

Figure 23: Market Share of EMR Vendors



The market has grown more concentrated over time. Figure 24 shows how the HHI of the most basic application, "Clinical Data Repository" increased during the sample period.[17] Concentration takes off especially in the post-HITECH period, around 2013.

Figure 24: HHI of Vendors within EMR Application "Clinical Data Repository"



6.2.3 Importance of Hospital Characteristics

As might be expected, Figures 25 and 26 shows that larger hospitals adopted both Basic and Advanced EHR functionality before smaller hospitals, in most cases even switching ahead of HITECH

[17]The pattern holds for other applications, including the Advanced applications CPOE and Physician Documentation. I show the CDR only for simplicity. When the hospital reports more than one vendor for the specific application, which occurs in about 1% of cases, I choose the one with the higher market share. The results are robust to choosing randomly or the one with the smaller market share, simply because there are so few cases of multiple vendors.

deadlines. Adler-Milstein et al. [2] study the distribution of EHR adoption behavior on types of hospitals in more detail, and find that larger hospitals, teaching hospitals, not-for-profit, and urban hospitals were all more likely to have a comprehensive (i.e. Advanced) EHR system in place by 2014, well after the HITECH deadlines. Hospitals that adopted EHR systems earlier were more likely to either visualize the greater benefits (as a large hospital might) or have a better ability to weather the costs (e.g. by virtue of being located in a technology center, as Dranove et al. [29] show). When considering the impact of digitization on other hospital outcomes, like data breaches, it will therefore be important to account for the fact that such systems are likely to have different costs based on hospital characteristics.

Figure 25: Year Hospitals Met Basic EHR Criteria, by Size

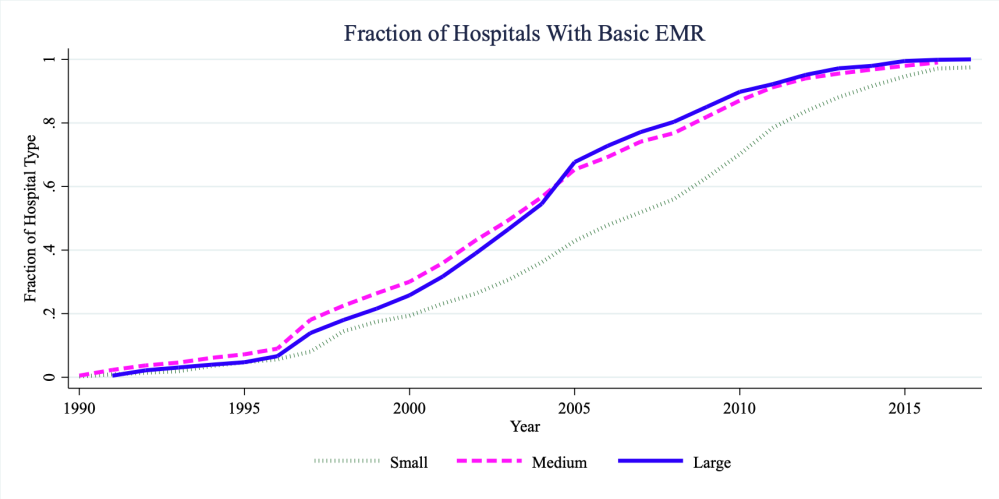
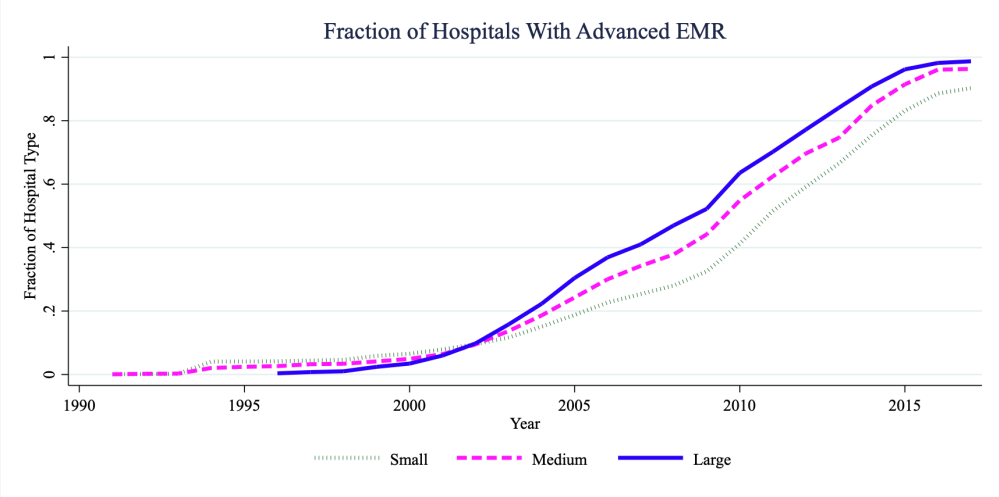


Figure 26: Year Hospitals Met Advanced EHR Criteria, by Size



6.2.4 Importance of Group Purchasing Organizations

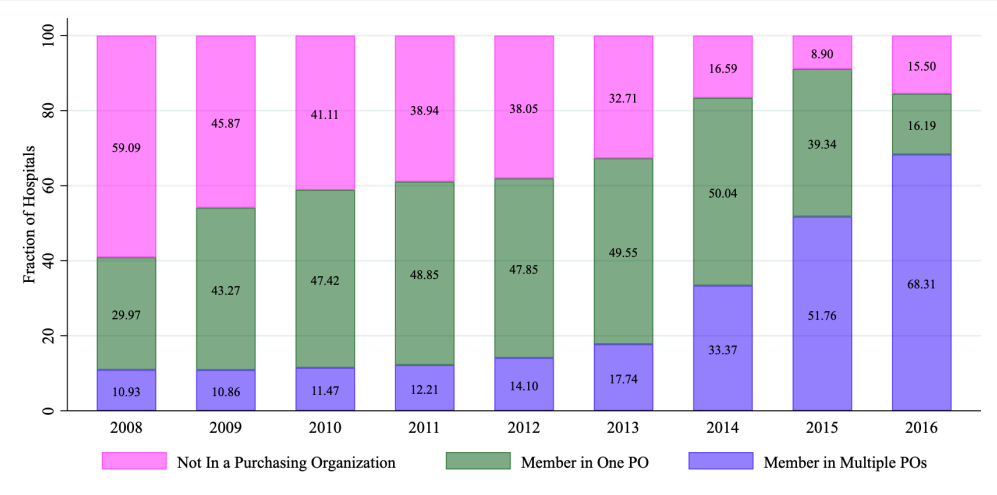
Healthcare providers in the U.S. are often members of *Group Purchasing Organizations*, which allow hospitals to work together and contract with vendors collectively at possibly favorable prices. If a GPO is not involved in a particular purchase, the hospital may be visited by vendors, e.g. for their EMR/EHR system, who can offer personalized pitches and pricing to cater to a hospital's specific needs. Such transactions are often highly localized – vendors may specialize in geographic areas near their headquarters – and dependent on hidden negotiations. For that reason, it has been historically difficult to estimate a demand model in a typical industrial organization style for how hospitals choose technologies within *any* category, let alone *all* categories.

GPOs have become more popular in the last decade. The HIMSS data show, for 2008-2017, which Group Purchasing Organizations hospitals joined. Hospitals may be in multiple GPOs; the mean number in 2016 is 2.612 and the median is 2.0. Figure 27 shows how more hospitals have joined GPOs, and even more have joined more than just one at a time.

Notably, GPOs do not necessarily *mandate* that hospitals all use the same vendors in any given category. Rather, they negotiate contracts on behalf of a large group of hospitals. These negotiations may then implicitly favor one vendor over the other if the GPO is better able to secure, e.g., low prices with one vendor over another. We thus do not expect to see an HHI of 10,000 within each GPO, but might expect some concentration as hospitals find some GPO options better than others.

Finally, note that security may in fact be one of the factors over which the collective GPO negotiates, the actual choice of the hospital would still only be *indirectly* affected by security. That is, while we might have endogeneity at the GPO level, endogeneity at the hospital level is still unlikely.

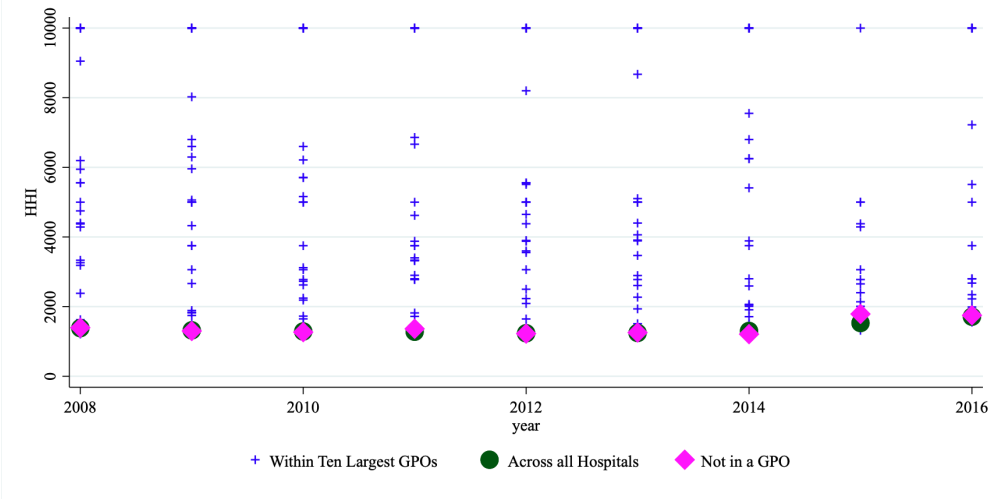
Figure 27: Fraction of Hospitals in Group Purchasing Organizations



Next, Figure 28 shows that hospitals within GPOs do generally tend to use the same software when compared with the general market (pink square) or organizations not in GPOs at all (blue diamond). GPOs, acting with buyer power in a market where the sellers of EHR software may

exercise market power (evidenced by regulatory barriers to entry and subsequent increases in concentration) would be expected to then boost concentration further. GPOs often require compliance or exclusivity by their members leading to the greater concentration of hospitals in the same product.

Figure 28: HHI of the EHR Market Within Group Purchasing Organizations



6.2.5 Importance of Neighbors

In the data, we see not only state-level concentration but also regional. Figure 29 shows that, as industry reports also discuss, Epic dominates in the Midwest and Western United States, while Cerner seemingly dominated the Mountain states in that period (Becker 2016). There is of course significant variation not only regionally but also in the definition of "dominance," as in some places the market share of the top EHR provider is not actually that high (Figure 30) – some have nearly 70% (dark blue) while others are around 20%.

Figure 29: Most Common EHR Provider in Each State

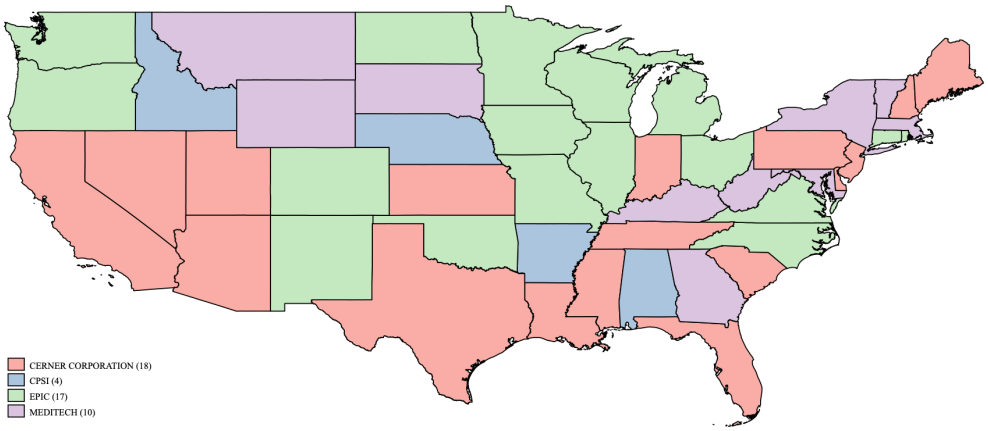
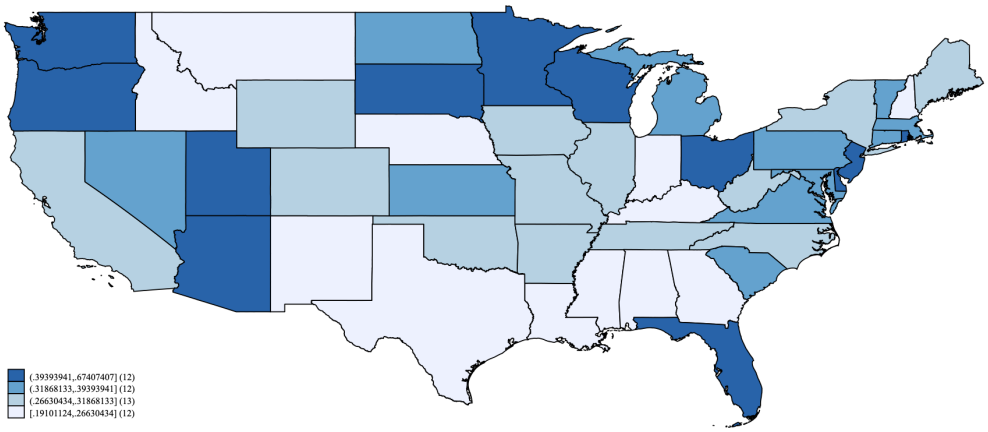
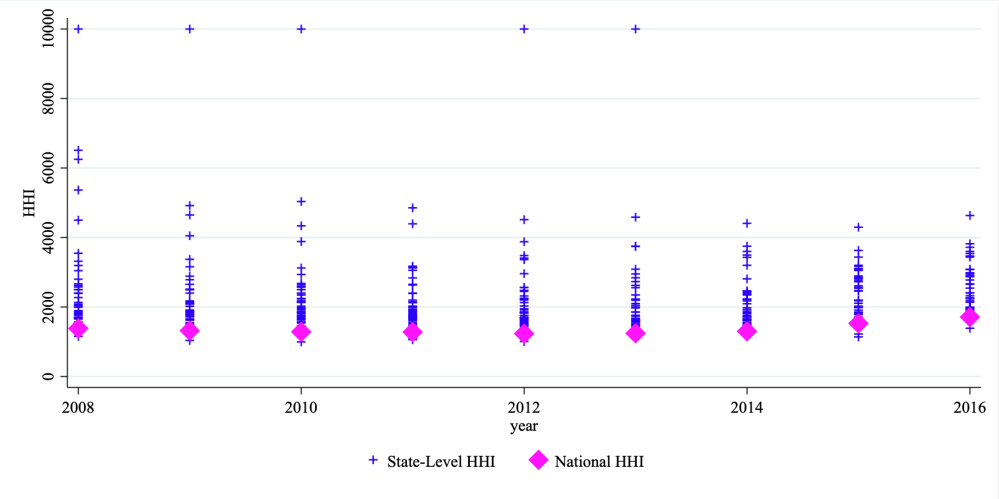


Figure 30: Market Share of the Most Common EHR Provider in Each State



A low national concentration might obscure high local concentration if each locality actually concentrates in a different product (Neiman and Vavra [61]). That is, the local market power of the EHR providers may be higher than their national market power. Hospitals are highly influenced by their neighbors in adoption, and supply chains and local headquarters lead to local specialization of EHR contracts. Figure 31 shows that indeed, national concentration is lower than the state-level, suggesting local-specialization may be taking place.

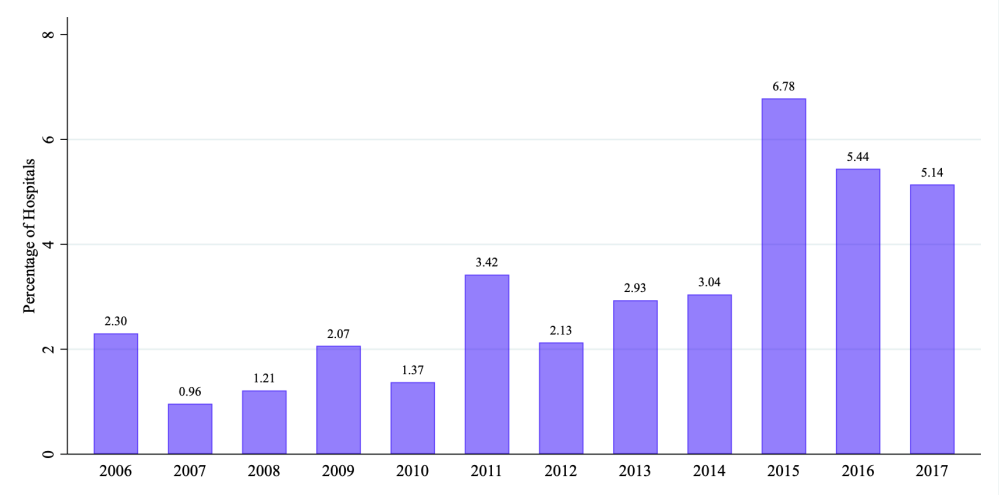
Figure 31: HHI of the EHR Market at the State-Level



6.2.6 Importance of History

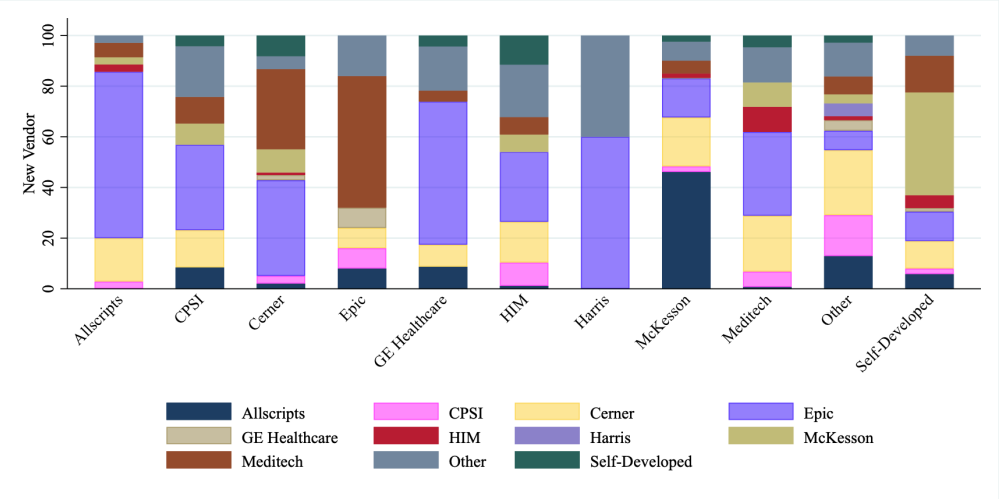
Hospitals do not *frequently* switch EMR Vendors: Figure 32 shows that 2-7% of hospitals switch vendors each year. Switching costs are certainly high, due to the startup costs and the time for implementation and training discussed in Section 4.3.

Figure 32: Fraction of Hospitals who Switch Their EMR Vendor



Those who do switch are generally switching from smaller vendors or self-developed software to, in about 30% of cases, to either Cerner or Epic, the two market leaders at the end of the sample.

Figure 33: Which Vendors Hospitals Switch From and Into



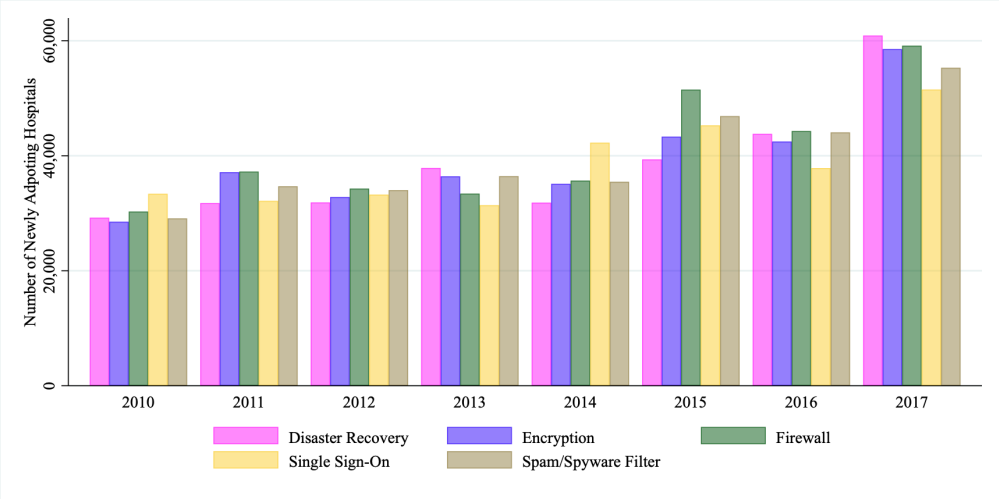
6.3 The Security Technology Market

The security technologies reported in HIMSS are reported in a somewhat more haphazard manner than the EHR firms reported, likely because each hospital within a system, or even each division, may have its own processes. Further, cybersecurity technologies often also require vigilance on the part of the user, as described in 4.4. Technology design, nonetheless, can help mitigate the consequences of a cyber attack. For example, using a single-sign on technology makes it easier for users (i.e. doctors and nurses) to sign into secure systems quickly, but also requires the user to choose a sophisticated enough password (both a technology choice and a user action) to withstand common password-spray attacks (Florencio and Herley [33]). The level of access granted from a single authentication is a technology choice. The password sophistication is a user choice, though the technology may provide some baseline (e.g. mix of characters).

In Figure 10, we saw that hospitals increased their contracts with security vendors across all categories reported in the HIMSS data. Figure 19 showed that hospitals listed contracts in more of each of applications out of the seven possible over time. That is, the overall security (extensive margin of all hospitals) and within-hospital security (intensive margin of which applications) increased over the sample period.

In Figure 34 I show the number of hospitals that report, for the first time, having a vendor contracted for each technology. I begin the graph 2010 because most security technologies were not explicitly asked about until 2009, creating an artificial spike in 2009. As such, we only have security technology data in the post-HITECH period. As I discuss in Section 3.3.2, the laws s written did not require any particular technologies but did make suggestions. The certification process also required some standard security provisions in the EHRs, but those would be counted as characteristics within the EHR contract rather than the standalone security contracts shown in Figure 34. We should therefore interpret these technologies as additional, voluntary security measures taken by the hospital to avoid the [costs associated with] data breaches.

Figure 34: Year of Security Technology Adoption: Specific Applications



The markets for each technology are fairly concentrated as shown in Figure 35, though each category of technology is dominated by a different firm as shown in Figures 36 through 40.

Figure 35: HHI of Specific Security Applications Markets

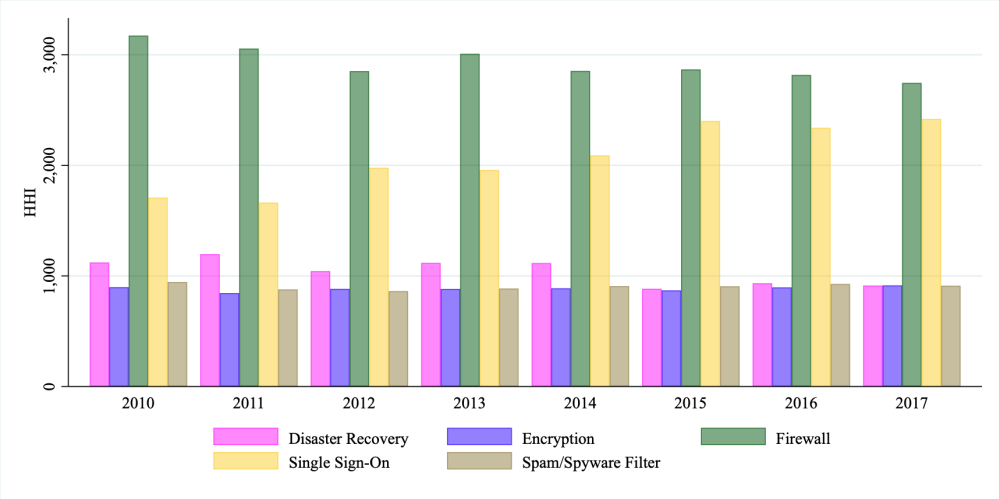


Figure 36: Market Shares: Firewall Software

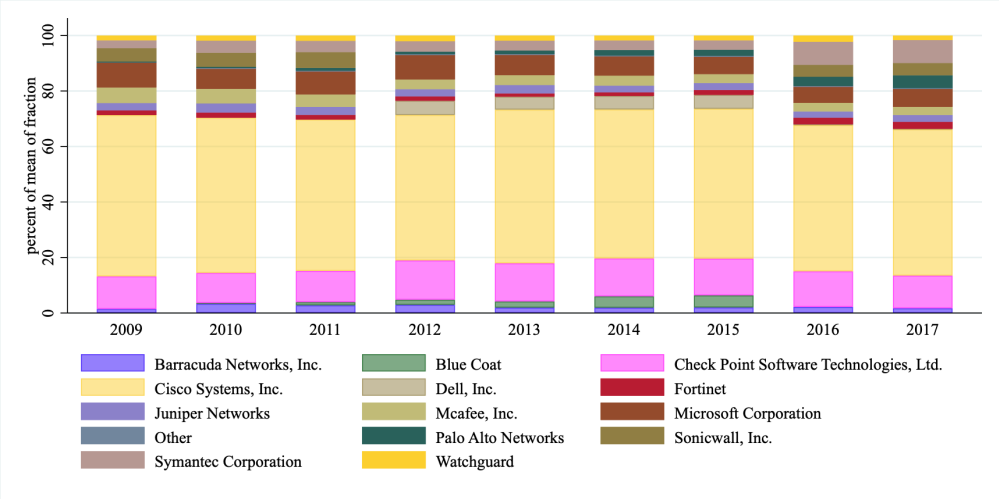


Figure 37: Market Shares: Encryption Software

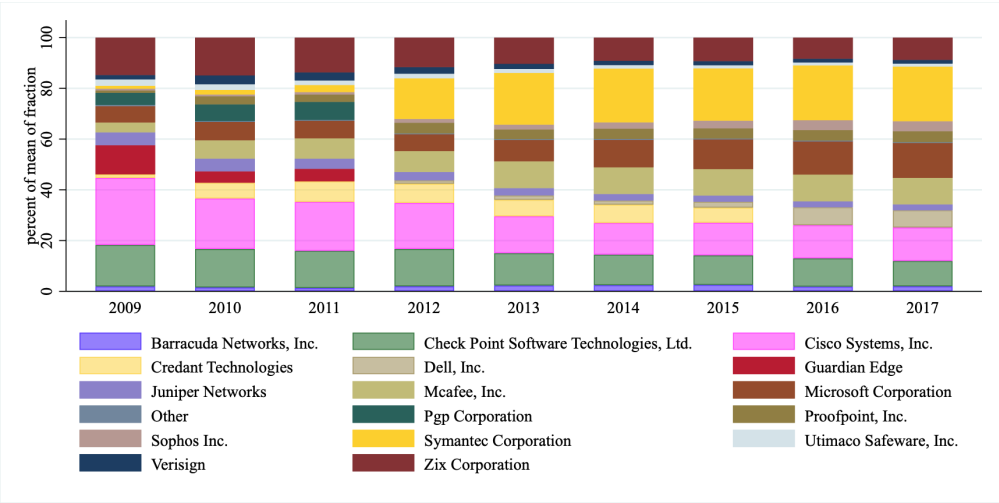
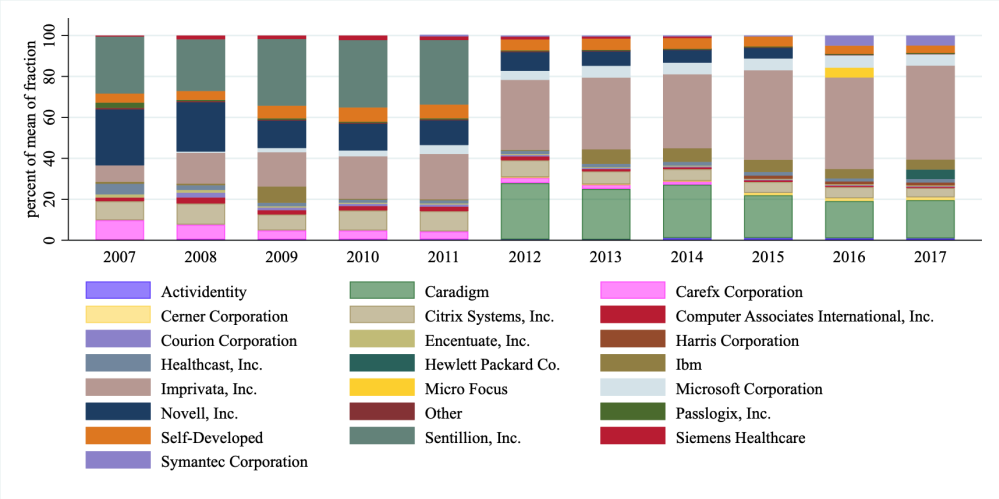


Figure 38: Market Shares: Single Sign-On Software



Sentillion, Inc. was acquired by Microsoft in 2010; Caradigm was acquired by Imprivata in 2017. I have accounted for mergers in the years after they occur, and assigned the original vendors when available (rather than the post-merger entity).

Figure 39: Market Shares: Spam/Spyware Filter Software

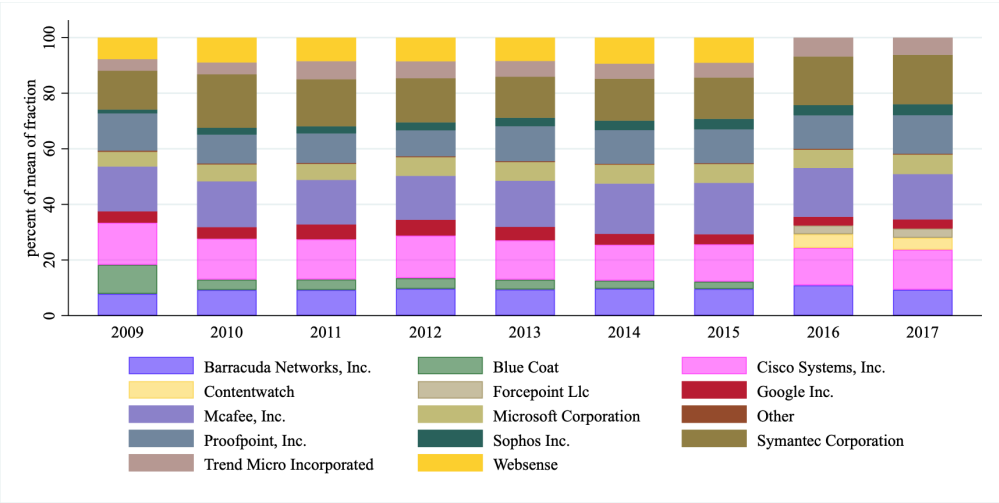
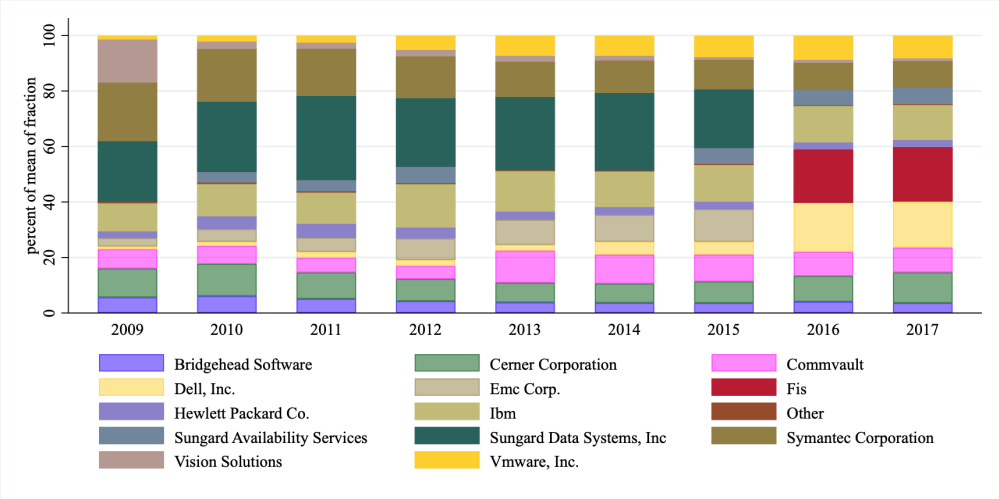


Figure 40: Market Shares: Disaster Recovery Software



These markets are more difficult to delineate from one other – e.g. Cisco provides Spam/Spyware Filter software as well as Encryption and Firewall software. I will instead consider these markets altogether, as a grand sum of "Security" that can be provided with different products. Still, I will regard each as separate: whether or not the hospital has a firewall, encryption, etc., as these are likely to actually be different products and therefore have different characteristics and sets of competitors.

As with EMRs, it does appear that large hospitals adopted *earlier* and that by the end of the sample a larger fraction of large hospitals have adopted. The differences are smaller, mostly because the data only start in 2009 and end in 2017, covering eight years of possible new adopters. In all categories, the total is far from 100% adoption. The technology with the largest size gap is Single Sign-On (Figure 43, which perhaps makes sense since it is a technology that facilitates signing in easily across different physical systems – more useful at large providers.

Figure 41: Year Hospitals Adopted a Encryption, by Size

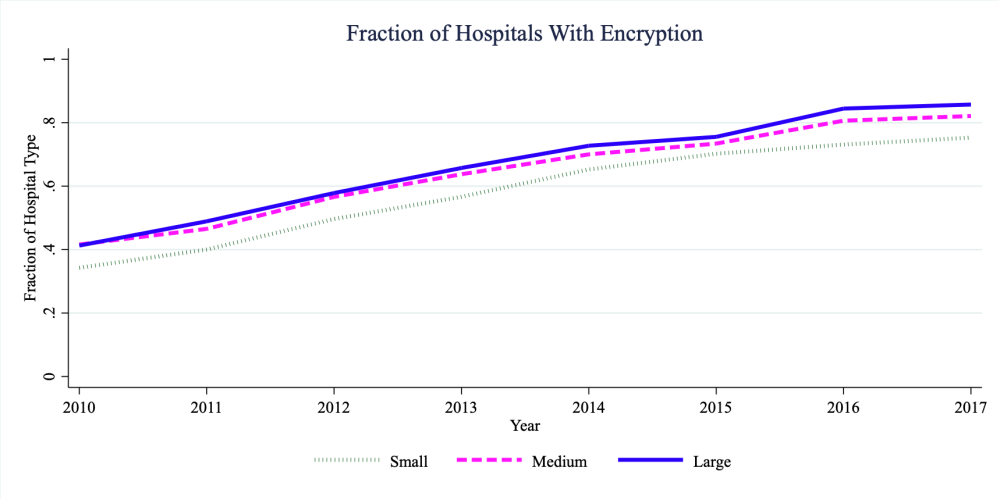


Figure 42: Year Hospitals Adopted a Firewall, by Size

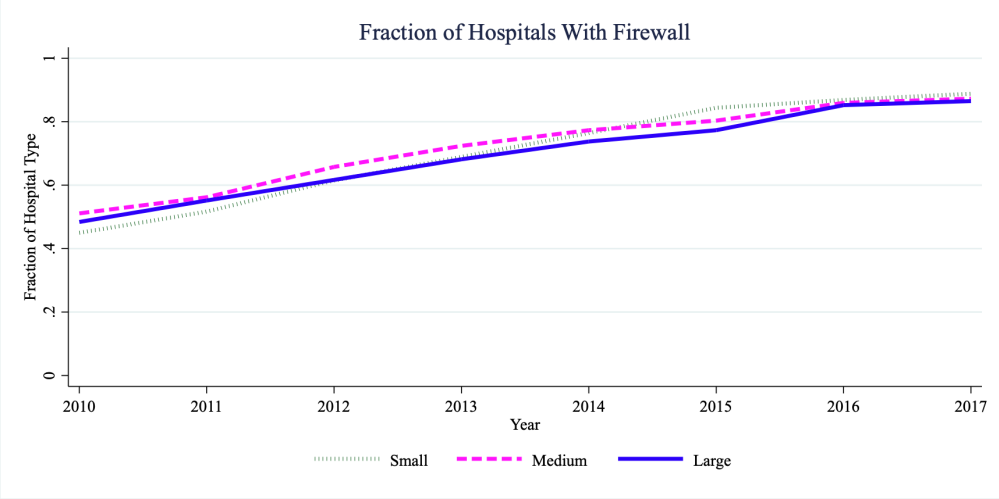


Figure 43: Year Hospitals Adopted a Single Sign-On Software, by Size

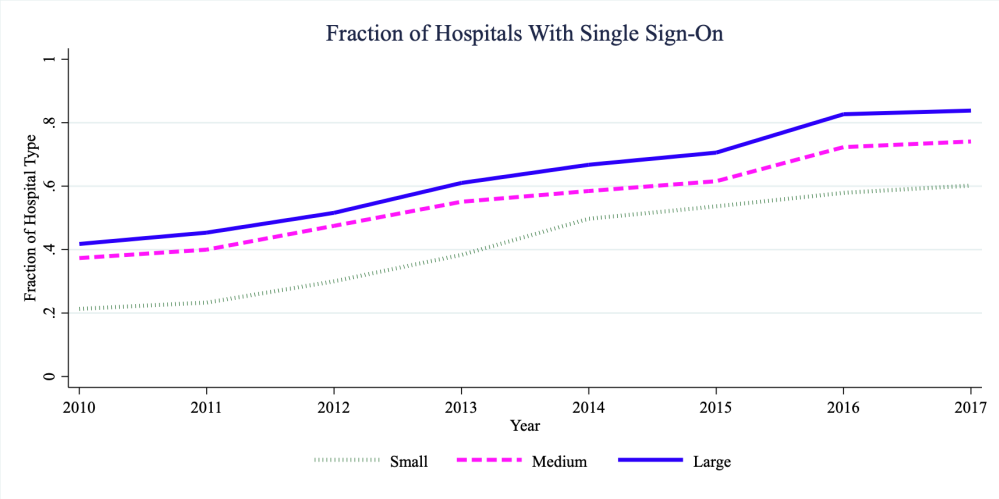
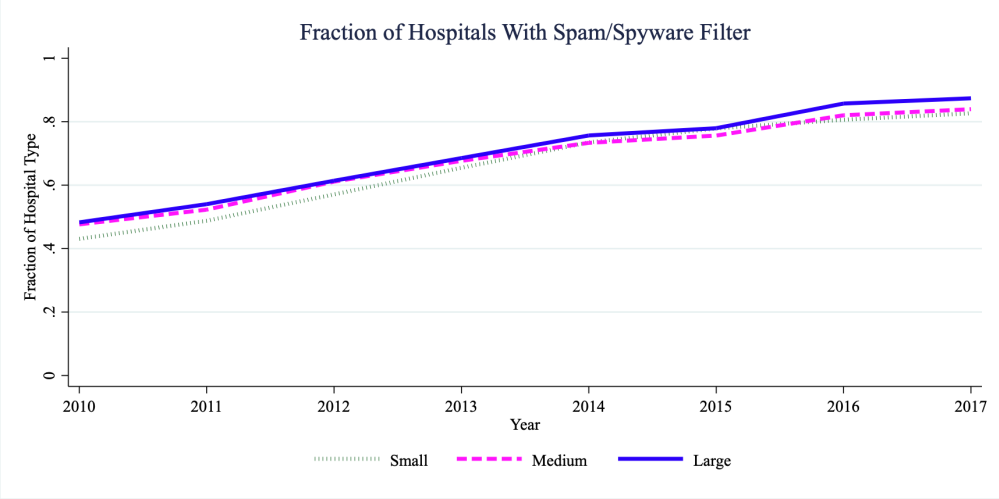


Figure 44: Year Hospitals Adopted a Spam/Spyware Filter, by Size



7 How Has Hospital Digitization Affected Data Breaches?

In this section, I analyze how a hospital's experience of data breaches changes due to digitization. Mechanically, we might expect that as hospitals digitize their operations, they simply become exposed to cyberattacks in a way they were not with physical records. They may also expect fewer physical breaches. However, if digital data is actually better secured than physical data, then we might expect overall breaches to go down, with perhaps only a compositional shift. Finally, hospitals that digitize may be fundamentally different from those that do not, and hospitals that

experience data breaches may also be targeted for some characteristics outside of their digitization behavior.

7.1 Does Having an EHR or Security System Predict or Prevent Breaches?

I first investigate whether the switch to digitization of healthcare records in EHR systems leads to greater breaches. I follow the common strategy in the literature of separating EHR systems into Basic and Advanced. I focus on the presence of cyber breaches and physical breaches, as described in Section 5.1.1, which would be expected to increase relative to physical breaches after digitization.

7.1.1 Basic Correlations

Figure 45 shows that in the years they were breached, hospitals were more likely to have an Advanced EHR system than in the rest of the sample period. Each observation is a hospital-year combination; the Breached bar only includes the hospital-years that have an actual breach. Of course, the figure masks much heterogeneity in *which* and *when* hospitals digitize, motivating the next analyses. The difference intensifies when we look at cyber breaches specifically

Figure 45: Data Breaches by EHR Attainment

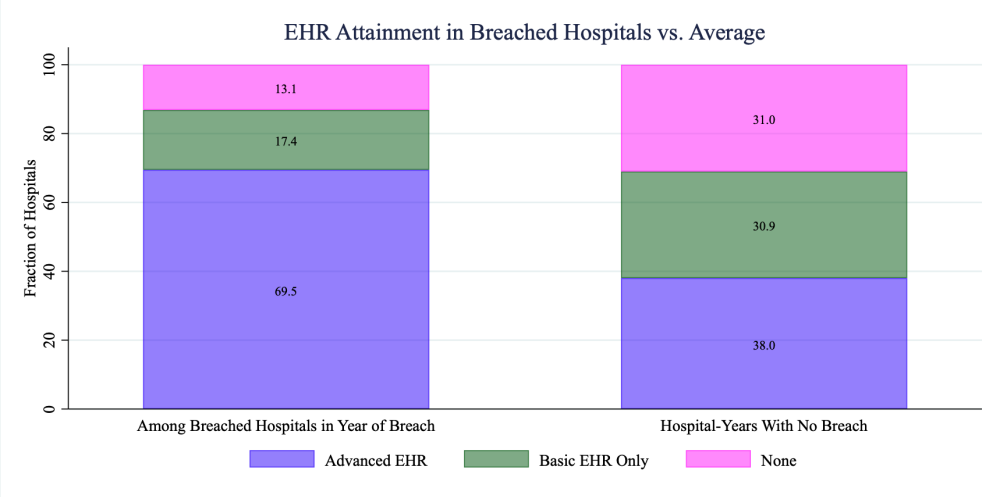
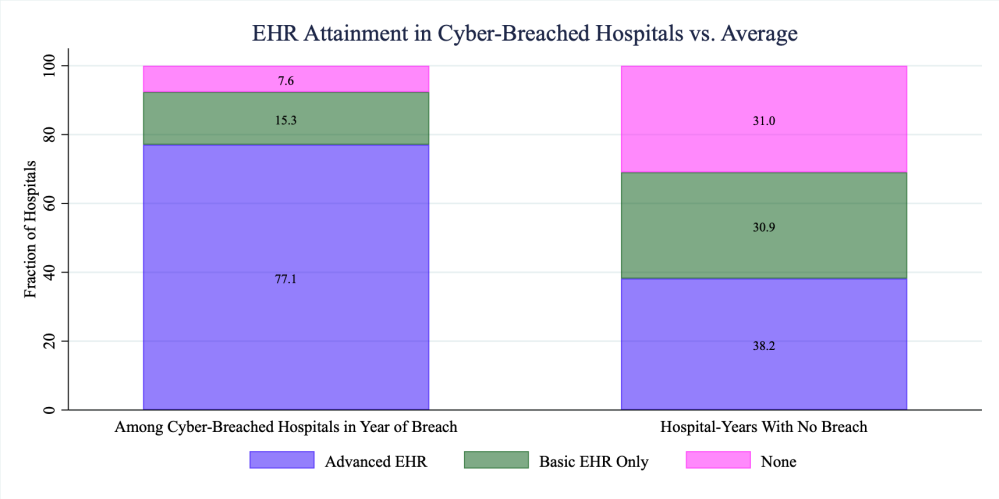


Figure 46: Cyber-Only Data Breaches by EHR Attainment



The endogeneity problem becomes more obvious when we look at security technologies. Having the three main security technologies – Firewall, Encryption, and Spam/Spyware Filter – is associated with having *more* breaches. Of course, hospitals that have security are different from those that do not; my apartment has less security than Fort Knox because it contains nothing of value and therefore I expect fewer attacks.

Figure 47: Data Breaches by Security Technologies

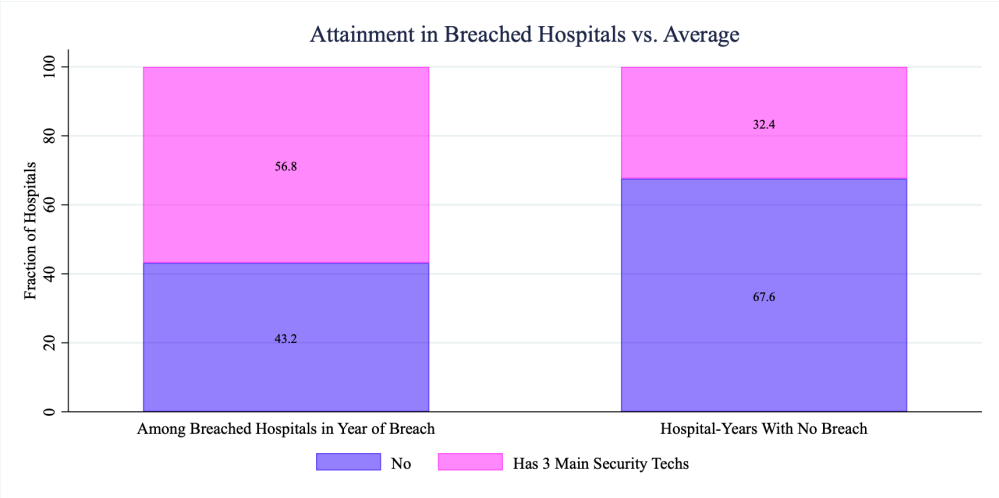
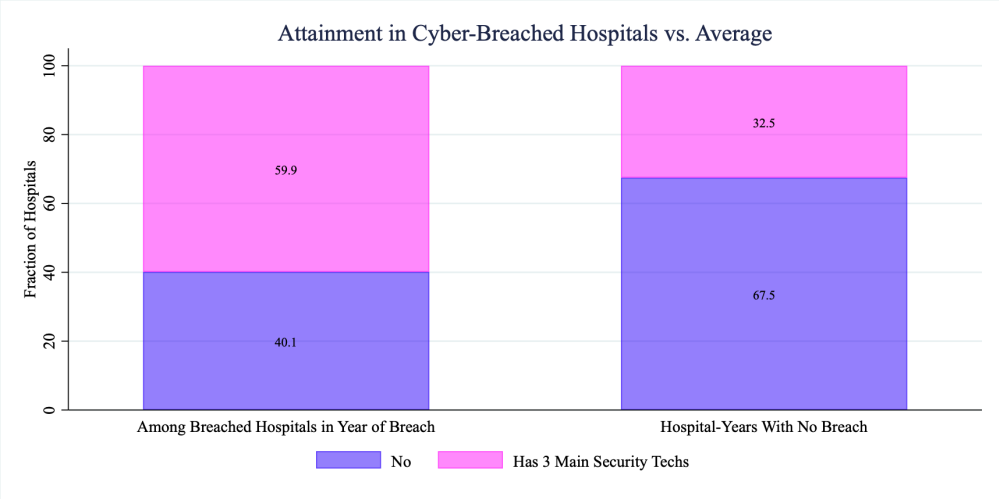


Figure 48: Cyber-Only Data Breaches by Security Technologies



7.2 Duration Model: Time to Breaches

My primary methodology for evaluating the role various technologies play in cyberattacks is the *duration model* (Van Den Berg [72]). 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-sectional 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 does not "survive" unbreached longer 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 [72]). 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.

7.2.1 Cyber vs. Physical Breaches

One would expect a priori that after records are digitized, there would be more cyber attacks and fewer physical attacks. On the other hand, physical attacks may simply hold steady while cyber attacks increase, leading to 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 cyber attacks, or if other factors simply mean more breaches in general.

7.2.2 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 5.1.3). 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 [25]):

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

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.

7.2.3 Crime Data Breaches:

The full theory model is outside the scope of this particular paper. For exposition purposes, I present some basics here.

Let us assume for some basic structure around the probability of a criminal data breach. 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 7.2.2, 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 not just one 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.

7.2.4 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 [51]). However, my exact argument in Section 7.2.3 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 literature of hospitals' choice to digitize, which shows that hospitals' choice to adopt EHR systems is primarily driven by the perceived costs and benefits. These costs and benefits change over time thanks to the HITECH subsidies (Dranove et al. [30], Adler-Milstein and Jha [3]), might vary by location (Dranove et al. [29]), or basic hospital characteristics (Adler-Milstein et al. [2]).

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 [51] 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. [22], joining an HIE is possibly related to data breach risk. I describe the methodology through which joining an HIE might affect data breach risk in Section 8.2.2. 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 and discuss the network effect of an EDI in Section 8.2.2. 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.[18] By 2017, 89% of hospitals in the final sample have a radiology information system.

Finally, in Section 6.2.6, I showed that hospitals do not switch their vendors nor particularly shift their digitization behavior around a breach. Further, given the length of time it takes to contract with and install an EHR system, hospitals' choice to digitize may precede actual digitization by years, introducing a time wedge between the choice to digitize (the possibly endogenous choice) and the years or months later when system is implemented and therefore accessible to attackers. I therefore simply include hospital characteristics and use of other technologies to capture the forces that drive EHR adoption, and allow for the effect of digitization to vary over time.

7.2.5 Results: Extensive Margin of Digitization

In the Cox Proportional Hazards model, the coefficients act multiplicatively on the hazard rate of a breach. A coefficient above one increases the hazard ratio while a coefficient below one decreases it.

First, Table 3 evaluates the relationship between digitization, of both EMRs and security technologies, and breaches. I find that having a Basic EMR system at first appears to reduce the likelihood of a breach, while having an Advanced EMR system does the opposite; however, as discussed above, basic hospital characteristics correlate with digitization behavior. I add in the main hospital characteristics: size, state, teaching status, residency status, control type, and service type. I display the results for the size, which can give the reader an understanding of the magnitude of the digitization effects in relation to a well-established fact, that larger hospitals experience more breaches.

When we include hospital characteristics, the relationships disappear, and notably the directions are exchanged: the hazard ratio of having a Basic EMR system is now above one and the Advanced coefficient drops to below one.

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

Table 3: Hazard Ratios, Breached: Extensive Margin

	Breached	Breached
Basic EMR	0.669*	0.887
	(0.150)	(0.215)
Advanced EMR	1.461**	0.916
	(0.226)	(0.144)
Firewall	0.776	1.021
	(0.131)	(0.193)
Spam/Spyware Filter	0.839	0.767
	(0.129)	(0.136)
Encryption	1.276	1.155
	(0.191)	(0.181)
Single Sign-On	1.539***	1.232*
	(0.183)	(0.146)
HIE	1.373***	0.832
	(0.162)	(0.108)
EDI	1.101	1.100
	(0.173)	(0.180)
RIS	1.795**	1.606*
	(0.467)	(0.454)
Medium		2.171***
		(0.366)
Large		3.899***
		(0.716)
Pseudo-R-Squared	0.01	0.07
Hospital Count	4212	4212
Breach Type Count	381	381

I then look at how such technologies affect cyber vs. physical breaches, and crimes vs. mistakes. As described in Section 7.2.3, we expect digitization to have different effects on each type of breach that may not be captured when we look at all breaches together.

Table 4 shows explicitly that having a basic EMR system corresponds to an *increase* in the hazard rate for cyber-specific breaches, significantly and high in magnitude – on par with being a Medium instead of a Small hospital – while it weakly *decreases* the hazard rate for physical-specific breaches. The same pattern holds for Advanced EMR systems: the hazard ratio is above one for cyber breaches but below one for physical breaches.

Similarly, the coefficients on Security technologies – Firewall, Spam/Spyware Filter, and Encryption – are weakly related to decreases in cyber breaches, but not with physical breaches. However, these coefficients lack statistical significance.

The crime vs. mistake coefficients are all statistically insignificant, suggesting that the extensive margin – whether or not the hospital has an EMR or particular security technologies – may not directly inform the type of breaches it experiences.

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

	Breached	Cyber	Physical	Crime	Mistake
Basic EMR	0.887	2.596	0.674	1.367	0.457**
	(0.215)	(1.801)	(0.167)	(0.470)	(0.155)
Advanced EMR	0.916	1.049	0.897	0.976	0.902
	(0.144)	(0.283)	(0.171)	(0.184)	(0.247)
Firewall	1.021	1.495	0.892	0.944	1.213
	(0.193)	(0.451)	(0.202)	(0.222)	(0.344)
Spam/Spyware Filter	0.767	0.687	0.783	0.755	0.766
	(0.136)	(0.175)	(0.175)	(0.173)	(0.200)
Encryption	1.155	0.922	1.281	1.335	0.859
	(0.181)	(0.232)	(0.243)	(0.262)	(0.205)
Single Sign-On	1.232*	1.167	1.258	1.173	1.382*
	(0.146)	(0.237)	(0.181)	(0.174)	(0.259)
HIE	0.832	1.017	0.808	0.752*	1.094
	(0.108)	(0.241)	(0.127)	(0.123)	(0.239)
EDI	1.100	1.247	1.091	1.041	1.303
	(0.180)	(0.371)	(0.220)	(0.208)	(0.375)
RIS	1.606*	0.810	1.990**	1.083	2.765***
	(0.454)	(0.511)	(0.571)	(0.401)	(1.091)
Medium	2.171***	3.443***	1.798***	2.416***	1.899***
	(0.366)	(1.125)	(0.356)	(0.558)	(0.472)
Large	3.899***	5.635***	3.608***	4.327***	3.998***
	(0.716)	(2.017)	(0.759)	(1.080)	(1.075)
Pseudo-R-Squared	0.07	0.12	0.07	0.10	0.08
Hospital Count	4212	4212	4212	4212	4212
Breach Type Count	381	138	255	239	155

Note that because I collapse the set of breaches into years, if a single hospital experienced both a cyber breach and a physical breach in the same year, it will be listed in both samples. There are four such cases.

Discussion of Control Variables I take the time here to discuss a few control variables. First, Choi et al. [22] discusses the possibly amplifying role joining an HIE may have on a hospital's probability of a data breach. I find a positive effect without controls and a negative one with controls. With controls, it is also not statistically significant in any breach category. There are many reasons for our differences. First, I use different data: Choi et al. [22] get their information from the AHA while I use HIMSS' definition of an HIE. Hospitals in my data only report joining HIEs in 2012 or later. Second, and more importantly, it may be that the effect of joining an HIE has more to do with the technologies that enable such information exchange. That is, the effect of an HIE is actually the effect of having an EMR technology, or even a specific vendor who facilitates the HIE. I investigate the specific roles an HIE may play later in Section 8.2.2.

7.3 Does the Specific EHR Vendor Predict Breaches?

Next, I use the specificity of the HIMSS data combined with the HHS data to understand if certain technologies are actually correlated with more breaches – i.e. if one might be able to discern if some vendors' products simply have *worse* security. 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 5.1.1 I use the information to attribute an attack to a "crime" vs. "mistake" and a "cyber" vs. "physical" attack, but the description is not detailed enough to get more specific.

In this section, I take a backwards-induction-style approach, looking at breach *outcomes* to infer something about breach *sources*. Using the model of Section 4.6.3, 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 of 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, we might investigate 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 6, 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, Section 7.4 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 7.2, 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 or not 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.

7.3.1 Breaches in General: Some Vendors *are* Worse Than Others

In Table 5 I present the basic results for *all* breaches. All coefficients on specific vendors use the "Other" category as a baseline, which covers all other vendors not in the top ten vendors of any given year. We can think of the "other" category as "small" vendors.

To see the power of the intensive vs. the extensive margins, Column 2 repeats the same extensive-margin regression from Table 3. Here, we see no real change in the R-squared. No real change is expected given that using an EMR Vendor in the first place basically incorporates the extensive margin implicitly – these hospitals are nearly all meeting the Basic criteria almost by definition We see heterogeneity in the hazard rate of particular vendors: while having a Basic EMR had almost no relation to being breached (first column), the second column shows higher point estimates for Allscripts and Self-Developed software, and lower estimates for HMS and Healthland.

These results are 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 [28]), and Self-Developed software simply seems unlikely to be as secure as a commercial offering. We can therefore in a sense *uncover* possible common vulnerabilities even when official descriptions exclude them.

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.

Section 8.2.3 examines exactly how and why the specific vendor might matter more than just the extensive margin. In particular, it may not be which vendor *you* use, but which vendor *everyone else uses*.

7.3.2 Cyber vs. Physical Breaches

Next, I repeat the exercise comparing Cyber and Physical breaches once more in Table 6. Hospitals should specifically be concerned with an increase in cyber breaches following technology implementation as discussed in Section 7.2.1. A correlation with physical breaches would not necessarily be attributable to a particular vendor, while cyber breaches may be.

As in Section 7.1, most of the coefficients on specific technology vendors once again positively multiply the hazard ratio for Cyber breaches (second column) but negatively multiply the hazard ratio for Physical Breaches. In particular, the Allscripts point estimate is above one for Cyber breaches, but close to one for physical breaches, as is the case for nearly all other vendors. The association of vendors with cyber breaches again makes sense as these may be common vulnerabilities not discovered by the defense but nonetheless exploited. We would not necessarily expect a digital technology to have an effect on physical breaches.

In all cases, I control for the extensive margin of whether the EMR system is used in a Basic or Advanced manner. Further, adding vendor identity simply explains more of the variation in cyber breach outcomes, evidenced by the Pseudo-R-Squared, than it does for Physical breaches.

7.3.3 Crimes vs. Mistakes

Finally, I make use of the crime vs. mistake classification and begin to look at how specific vendors might be associated with either type of breach in Table 7. Hospitals specifically are concerned with the possibility of crime breaches, which cannot be necessarily prevented by the actions of hospital employees but are the consequence of bad technology and a third-party seeking to do harm. I do not here distinguish between cyber-crimes and physical-crimes.

Table 5: Hazard Ratios, Breached: Intensive Margin

	Breached		Breached	
Basic EMR	0.887	(0.215)	0.817	(0.206)
Advanced EMR	0.916	(0.144)	0.885	(0.141)
Firewall	1.021	(0.193)	0.988	(0.191)
Spam/Spyware Filter	0.767	(0.136)	0.809	(0.147)
Encryption	1.155	(0.181)	1.130	(0.177)
Single Sign-On	1.232*	(0.146)	1.250*	(0.149)
HIE	0.832	(0.108)	0.806	(0.110)
EDI	1.100	(0.180)	1.102	(0.180)
RIS	1.606*	(0.454)	1.708*	(0.488)
Medium	2.171***	(0.366)	1.886***	(0.323)
Large	3.899***	(0.716)	3.245***	(0.600)
Allscripts			1.537	(0.475)
CPSI			0.825	(0.346)
Cerner			1.207	(0.329)
Epic			1.286	(0.364)
HIM			0.551	(0.323)
Healthland			0.273	(0.279)
McKesson			1.466	(0.443)
Meditech			0.957	(0.268)
Other			1.000	(.)
Self-Developed			2.235**	(0.755)
Siemens			1.123	(0.387)
Pseudo-R-Squared	0.07		0.07	
Hospital Count	4212		4211	
Breach Type Count	381		381	

Some vendors do seem more associated with crime breaches than physical ones, with an even clearer-cut "above one for crime" and "below one for mistake" dichotomy evident. Distinguishing crimes from mistakes allows us to understand whether the fault may lie with the *technology itself* – for even indirectly facilitating a breach[19] In particular, the Allscripts vulnerability possibly facilitating Crime breaches becomes obvious in the analysis.

7.4 How Do Hospitals React After a Breach?

In this section, I show that both the hospital's choice to digitize or of specific do not appear to respond to the occurrence of data breaches.

What do hospitals *do* after they experience a data breach? Choi et al. [23] and Lee and Choi [56] find that hospital quality, measured by two healthcare outcomes, goes down for some time,

[19]For example, the Change healthcare hack is rumored to be caused by social engineering of a high-level administrator who had extremely broad access to systems and data

Table 6: Hazard Ratios, Total vs. Cyber vs. Physical Breaches

	Breached		Cyber		Physical	
Basic EMR	0.817	(0.206)	2.345	(1.730)	0.613*	(0.159)
Advanced EMR	0.885	(0.141)	1.050	(0.282)	0.830	(0.161)
Firewall	0.988	(0.191)	1.480	(0.457)	0.868	(0.203)
Spam/Spyware Filter	0.809	(0.147)	0.729	(0.192)	0.827	(0.191)
Encryption	1.130	(0.177)	0.904	(0.225)	1.231	(0.234)
Single Sign-On	1.250*	(0.149)	1.217	(0.248)	1.267	(0.184)
HIE	0.806	(0.110)	0.988	(0.250)	0.769	(0.126)
EDI	1.102	(0.180)	1.239	(0.367)	1.096	(0.222)
RIS	1.708*	(0.488)	0.885	(0.578)	2.130***	(0.620)
Medium	1.886***	(0.323)	3.347***	(1.156)	1.513**	(0.301)
Large	3.245***	(0.600)	5.211***	(1.961)	2.918***	(0.612)
Allscripts	1.537	(0.475)	3.350	(2.665)	1.522	(0.545)
CPSI	0.825	(0.346)	2.717	(2.405)	0.637	(0.331)
Cerner	1.207	(0.329)	1.987	(1.509)	1.195	(0.377)
Epic	1.286	(0.364)	2.497	(1.942)	1.279	(0.418)
HIM	0.551	(0.323)	2.646	(2.592)	0.188	(0.196)
Healthland	0.273	(0.279)	0.000	(.)	0.290	(0.300)
McKesson	1.466	(0.443)	3.622*	(2.817)	1.176	(0.418)
Meditech	0.957	(0.268)	1.808	(1.356)	0.938	(0.304)
Other	1.000	(.)	1.000	(.)	1.000	(.)
Self-Developed	2.235**	(0.755)	5.295*	(4.507)	2.035*	(0.831)
Siemens	1.123	(0.387)	1.831	(1.596)	1.064	(0.417)
Pseudo-R-Squared	0.07		0.12		0.07	
Hospital Count	4211		4211		4211	
Breach Type Count	381		138		255	

while productivity measured as value added does not appear to change. Kwon and Johnson [53] find that patients do not respond significantly to data breaches by switching hospitals, especially if they live in an area with few alternative options; Hydari et al. [44] find hospital competition might encourage security investment as a differentiator, though results are small in magnitude. In general, measuring the impact of a data breach on a hospital will depend on many hospital and breach-specific factors. Hospitals may themselves be highly dependent on digital processes, which could interact with a comprehensive cyberattack that shuts down access to have a massive impact on costs and outcomes; on the other hand, papers scattered in the wind may not have much of an impact beyond the direct fines imposed by HHS. Differentiating the impact requires highly detailed information about both the hospital and the breach that is not systematically available.

For the purposes of this study, the relevant dimension is whether hospitals change their *technology* choices in respond to a breach, which would mean they may are actually making security-based decisions about technologies. In that case, security would be a relevant dimension of competition for the EHR providers. I investigate the following:

Table 7: Hazard Ratios, Total vs. Crime vs. Mistake Breaches

	Breached		Crime		Mistake	
Basic EMR	0.817	(0.206)	1.205	(0.435)	0.466**	(0.165)
Advanced EMR	0.885	(0.141)	0.941	(0.180)	0.856	(0.233)
Firewall	0.988	(0.191)	0.916	(0.217)	1.205	(0.348)
Spam/Spyware Filter	0.809	(0.147)	0.808	(0.186)	0.769	(0.215)
Encryption	1.130	(0.177)	1.326	(0.261)	0.829	(0.196)
Single Sign-On	1.250*	(0.149)	1.210	(0.181)	1.362	(0.256)
HIE	0.806	(0.110)	0.724*	(0.127)	1.053	(0.231)
EDI	1.102	(0.180)	1.061	(0.210)	1.302	(0.384)
RIS	1.708*	(0.488)	1.199	(0.458)	2.738**	(1.077)
Medium	1.886***	(0.323)	2.139***	(0.492)	1.648*	(0.426)
Large	3.245***	(0.600)	3.671***	(0.926)	3.337***	(0.904)
Allscripts	1.537	(0.475)	1.971*	(0.811)	0.983	(0.474)
CPSI	0.825	(0.346)	1.211	(0.666)	0.505	(0.323)
Cerner	1.207	(0.329)	1.318	(0.510)	0.962	(0.355)
Epic	1.286	(0.364)	1.665	(0.672)	0.841	(0.320)
HIM	0.551	(0.323)	1.249	(0.805)	0.000	(.)
Healthland	0.273	(0.279)	0.000	(.)	0.383	(0.412)
McKesson	1.466	(0.443)	1.712	(0.723)	1.035	(0.436)
Meditech	0.957	(0.268)	1.348	(0.521)	0.586	(0.232)
Other	1.000	(.)	1.000	(.)	1.000	(.)
Self-Developed	2.235**	(0.755)	3.519***	(1.624)	0.980	(0.597)
Siemens	1.123	(0.387)	1.194	(0.567)	0.930	(0.462)
Pseudo-R-Squared	0.07		0.10		0.08	
Hospital Count	4211		4211		4211	
Breach Type Count	381		239		155	

1. Do hospitals implement security technologies they previously lacked?

2. Do hospitals switch vendors for technologies they have already implemented?

Answering the above helps understand if the *network structure* of which hospitals use which vendors is *endogenous* to the security outcomes of those technologies in a direct way.

7.4.1 Hospitals Implement Security Technologies After Breaches

Table 8 shows the results of a regression comparing hospitals' technology choices on the extensive margin before and after they experience a breach. The regression is the following at the hospital h and year t level:

$$\text{Has Technology}_{h,t} \sim t \text{ is After } h\text{'s Breach}_{h,t} + \text{EverBreached}_h + \text{Hospital Controls}_h$$

I run the above as a Logit regression since the outcome variable is binary. In line with results of Section 7.2, hospitals that were ever breached are less likely to have the technologies; they are however more likely to have them *after* experiencing a breach, which may help them prevent subsequent breaches. In general, however, implementing security technologies *in response* to a breach suggests the a story of *reactive* rather than *proactive* security investment suggested by Kwon and Johnson [52]. If hospitals anticipated their breaches, we might expect no particular effect on *After Breach*.

Table 8: Implementing Technologies Post-Breach

	Firewall	Spam/Spyware Filter	Encryption	Single Sign-On
main				
After Breach	0.475***	0.375***	0.552***	0.731***
	(0.086)	(0.083)	(0.086)	(0.084)
Ever Breached?	-0.294***	-0.379***	-0.240***	-0.155***
	(0.057)	(0.057)	(0.057)	(0.057)
Observations	33696	33696	33696	33696

7.4.2 Hospitals Do Not Switch EMR Vendors After a Breach

On the intensive margin, hospitals do not appear to be switching vendors more frequently after experiencing a breach. About 37.9% of hospitals switched their EMR vendor at some point in the sample period. Were those switchers more likely to be breached hospitals switching after experiencing a breach? What about for other technologies?

Table 9 runs the same specification as Table 8, but with the outcome variable of Hospital h Switched Vendor in year t – the intensive margin of *which* vendor a hospital chooses rather than the extensive margin of *having* the technology. I include the EMR vendor here since although hospitals seem to adopt EMR vendors based on policies and incentives, they may still switch vendors based on cybersecurity outcomes.

I find that no coefficient on *After Breach* is statistically significant, meaning hospitals that experienced breaches are *not* more likely to subsequently switch their breaches than they were before their breaches; the coefficient on *Ever Breached* suggests that hospitals that were breached at all are not switching more often than those who were never breached as a baseline. Note that each regression is conditional on the hospital *having* the technology in the first place, which changes the sample from Table 8 and resulting in much fewer observations.

Table 9: Switching Vendors Post-Breach

	EMR	Firewall	Encryption	Single Sign-On	Spam/Spyware Filter
main					
After Breach	0.135	-0.185	0.195	-0.172	0.051
	(0.186)	(0.262)	(0.235)	(0.245)	(0.227)
Ever Breached?	0.114	0.070	-0.064	-0.088	-0.074
	(0.133)	(0.176)	(0.181)	(0.176)	(0.164)
Observations	27766	23307	19819	16019	22072

8 How Does the Market Structure for Healthcare Technology Affect Cybersecurity Outcomes?

In this section I examine how the structure of the market for healthcare technology interacts with the technical aspects of cybercrime to shape cybersecurity outcomes.

As I discussed in Section 1, cybercrime is distinguished by (i) the presence of a strategic attacker and (ii) the fact that attacks can scale costlessly to multiple targets when a common vulnerability is exploited. At the same time, the technology at hand benefits from *economies of scale*: software has near-zero marginal costs. In the case of EHRs, they also benefit from *positive network effects*. Imperfect interoperability means using the same software vendor as other hospitals has actual benefits to any hospital looking for a new contract.

In this section, I investigate both the positive and negative network effects that arise in the market for healthcare technologies. On the positive side, we cannot directly observe hospital costs and benefits of using the same EHR as its peers, but we can observe implicitly if hospitals are making choices based on *the possibility* of positive network effects.

Negative network effects – which we can also call *contagion externalities* – arise because of the increased incentive for cyberattacks to develop exploits for widely used technologies, especially if they are used by larger hospitals. Is a hospital more likely to be breached if one of its peers was breached? The outcome variable here is not a hospital cost of benefit but the incidence of a cyberattack: is a hospital more likely to experience a cyberattack if it uses a common technology, especially if that technology may have been exploited before?

Using the technique of Section 7.3, of backing out the security of a vendor by looking at whether and how many breaches it is associated with, I examine how breaches pass through a network via insecure technology vendors. I am therefore able to measure a cybersecurity "cost" of using a common vendor, which stands opposed to the "benefits" of positive network effects and economies of scale.

As described in Section 5 I do not have prices or explicit characteristics of products that determine why a hospital might choose a particular vendor. I use the arguments of Section 6.3 to instead treat the hospital's choice of EHR vendor as orthogonal to its security posture, as other factors (like the certification process, GPO negotiations, and regional dominance) are more salient. Section 8.3 presents robustness checks that account for possible endogeneity.

8.1 Measuring the Positive Network Effect

As I discussed in Section 4.2.1, the HITECH Act sought to encourage interoperability, which was seen as key to the social benefits of EHR adoption. Unless providers can actually exchange records with each other and share relevant information with authorities, only within-provider benefits could be realized. In practice, however, EHR software is still subject to within-software network effects, and perfect interoperability remains elusive.

Further, there are other non-interoperability network effects at play on the supply side. Software has a low marginal cost; once it has been developed it can be deployed with economies of scale. Administrative overheard and other fixed costs are spread across many consumers, giving large firms a natural price advantage. Technical support systems, formal (through the software firm) and informal (by asking peers at other hospitals, for example, or by maintaining a presence at industry conferences) , make using the dominant software more attractive.

A market with positive network effects may be a natural monopoly, as the monopolist maximizes the positive externality resulting from different users being able to interact with each other (Katz and Shapiro [49], Katz and Shapiro [50]). Further, late adopters of the technology – e.g. those who value the inherent technology less than the early adopters – may be incentivized to adopt only after sufficient other users have adopted and the network effects are high enough (August et al. [8]). Early market concentration, and the subsequently high network effects, may then be a predictor of future adoption.

I first test for the presence of network effects in adopting the same EHR vendor as other hospitals. I specifically look to see if early market share predicts later hospital choices, an indicator of the presence of network effects. I estimate the following for each hospital h in state s, year t, when the market leader is firm j_1 and the second-best is j_2, and so on.

$$\begin{aligned}\Pr\left(\text{Hospital } h \text{ Chooses Market Leader } j_1 | \text{Newly Digitizing}\right)_{s,t} \sim\ & \text{Market Share}_{j_1,s,t-1} \\ & + \left[\text{Market Share}_{j_1} - \text{Market Share}_{j_2}\right] \\ & + \text{Hosp. Characteristics}_{h,s,t-1} + f_{j_1}\end{aligned}$$

The above answers the following:

1. Is a higher market share for the leading EHR firm in a state correlated with higher future market share for that leader?

2. Does the size of the lead (relative to the second-place firm) correlate with higher future market share?

If both are answered with "yes," then the market may exhibit network effects and/or economies of scale, where choosing the leader is preferable specifically *because* they are the leader.

Table 10 shows that indeed, among the sample of hospitals that are picking their vendor for the first time, they are more likely to choose their state's market leader, and more so if that market leader has a wider lead on the second best. When we add in state fixed effects, the magnitude of the effect decreases, suggesting these network effects may be concentrated at the state and therefore captured in state-specific qualities. As a check, I instead include a fixed effect for the firm that is the market leader – correlated with region as we saw in Figure 29, and capturing unobserved firm

quality unrelated with its status as the leader – and see the same result. The result is maintained. Further, using the share of *beds* instead of the share of *hospitals* – to account for some vendors specializing in very large hospitals – maintains the result.

Positive network effects therefore do seem to push this market towards concentration.

Table 10: Probability of Adopting Market Leader When Digitizing

	Hospital Share	Hospital Share	Beds Share	Beds Share
Adopted Leader?				
Leader Market Share	9.443***	8.779***	2.203	3.980*
	(1.792)	(2.407)	(1.400)	(2.143)
Difference with Second Place	-1.172	0.368	1.313	1.629
	(1.284)	(1.690)	(1.152)	(1.743)
Hospital Controls?	Yes	Yes	Yes	Yes
State FE?	Yes	Yes	Yes	Yes
Vendor FE?	No	Yes	No	Yes

8.2 Measuring the Negative Network Effect: Scaled Attacks

In this section, I use the same Cox Proportional Hazard Model to see how breaches at other hospitals multiple an individual hospital's hazard rate. If there are indeed negative network effects, i.e. contagion externalities, then we would expect to see a very particular pattern of contagion externlities.

To judge the presence of negative networks, I exploit the characteristic differences between breaches that are *crimes* and those that are *mistakes*. As with the cyber vs. physical distinction, we expect that breaches that are *crimes* are those that are possibly subject to negative network effects. An attacker who discovers an exploit in a technology may seek to deploy it across multiple hospitals. A repeat offender may try to exfiltrate records from as many different hospitals as possible, especially if they want to create a large corpus for secondary resale. The negative network effect will therefore also be concentrated at the *technology-level*: rather than contagion across states, we would specifically expect the *technology vendor* to be the conduit for the attack.

On the other hand, in my classification, mistakes are breaches that rely on the internal operations of the hospital. These are not expected to scale, as each hospital can take actions on its own to prevent them. I focus here on the crime vs. mistake precisely because the scaled attack relies on the presence of a third party, which a general cyber breach (which includes mistakes such as incorrect emails) does not. Mistakes are not expected to scale in the same way, and a mistake at another hospital is not expected to put another hospital at risk. I also focus on the difference between cyber and physical breaches, which follow a similar dichotomy but again may not scale if a malicious third party is not responsible for the cyber breach (e.g. an incorrectly addressed email).

I define the "network effect" as the coefficient on the *fraction of peer group hospitals who were*

also breached. In each of the analyses that follow, the main dependent variable is constructed as:

$$\text{Perc. Group Breached}_{g,t} = \frac{\sum_{h \in g} Breached_{h,t}}{\#\text{Number of Hospitals in Group } g_t}$$

That is, in the year t, how many of the other hospitals in group g – importantly, excluding the *self* from both the denominator and the numerator to avoid an obvious reverse causality.

The results displayed here are also for *contemporaneous* breaches. Because the data are aggregated to the annual level, we cannot look at exactly when the previous breaches took place. Further, we expect crimes to take place at scale, affecting many hospitals at once, making looking at current-year variables more reasonable than lags. The results are however robust to using the lagged set of breaches.

I use the same set of controls here as in Section 7.2. I further include the two EMR variables (Has Basic and Has Advanced) and the presence of security technologies (Firewall, Encryption, Spam/Spyware Filter, and Single Sign-On). I present all coefficients to emphasize that here we are *adding* to the extensive margin and intensive margin analysis by looking at the network margin. However, for each analysis I omit any group fixed effects, if applicable, to preserve the variation required for the analysis. Because I consider the technology choice exogenous, the non-time-varying component is both important to the variation here (scaled attacks exploit vulnerabilities in products that are not immediately corrected) and also not important for the hospital's *individual* breach probability. I discuss robustness checks in Section 8.3.

8.2.1 Across States, GPOs, and Other Non-Technology Networks

I first look at non-technology networks that exist between hospitals: are attacks possibly contagious at a geographic or a coalition level?

State-Level First, Table 11 shows the hazard ratio on "one additional breach in another state in the same year." I use the same extensive-margin technology controls from Section 7.2 – the new explanatory variable is simply the count of breaches that took place in *other hospitals* in the state of in each year.

Table 11: Hazard Ratios: State Network Effects, OLS

	Breached	Crime	Mistake	Cyber	Physical
Perc. State Breached	1.013				
	(0.030)				
Perc. State Crime		1.053			
		(0.040)			
Perc. State Mistake			0.661***		
			(0.087)		
Perc. State Cyber				0.932	
				(0.076)	
Perc. State Physical					0.903*
					(0.054)
Firewall	1.040	0.935	1.407	1.586	0.871
	(0.192)	(0.212)	(0.404)	(0.466)	(0.196)
Spam/Spyware Filter	0.769	0.744	0.751	0.700	0.774
	(0.135)	(0.169)	(0.191)	(0.185)	(0.170)
Encryption	1.112	1.307	0.821	0.899	1.263
	(0.171)	(0.255)	(0.192)	(0.226)	(0.239)
Single Sign-On	1.287**	1.241	1.428*	1.206	1.326**
	(0.149)	(0.180)	(0.264)	(0.238)	(0.188)
HIE	0.846	0.762*	1.166	0.990	0.852
	(0.107)	(0.124)	(0.249)	(0.230)	(0.132)
EDI	1.117	1.052	1.277	1.346	1.067
	(0.180)	(0.208)	(0.352)	(0.394)	(0.211)
RIS	1.564	1.081	2.666**	0.738	2.014**
	(0.437)	(0.396)	(1.047)	(0.457)	(0.577)
Basic EMR	0.865	1.301	0.444**	2.441	0.657*
	(0.212)	(0.446)	(0.149)	(1.692)	(0.164)
Advanced EMR	0.987	1.034	1.036	1.191	0.957
	(0.154)	(0.191)	(0.279)	(0.307)	(0.182)
Pseudo-R-Squared	0.06	0.07	0.06	0.08	0.06
Hospital Count	4212	4212	4212	4212	4212
Breach Type Count	381	239	155	138	255

The first column shows the relationship between all breaches at other hospitals in the state and breaches at a single hospital. The coefficient is not statistically significant from one, which is "no effect" in the Cox model. When we break down breaches into crimes and mistakes, however, we can see that the effect is muted for mistakes and is statistically significant. For crimes on the other hand, an additional crime at a hospital in the same state multiplies the hazard rate with a coefficient greater than one and is not statistically significant. That is, we cannot say that crimes at other hospital influence crimes at one's own hospital within the same state.

The findings also suggest that the "learning" channel is perhaps salient in the negative effect

that mistakes at other hospitals seem to have: perhaps a hospital learns from the mistakes of its literal neighbors. However, no evidence of scaled attacks.

GPO-Level Next, I look at hospitals that may not be in the same location but might have some commonalities due to being in the same Group Purchasing Organization. It is likely these hospitals may contract with the same vendor, or have other informal ties that attackers may seek to exploit. Does an additional breach in the same GPO lead to more breaches at one's own hospital?

First, note that GPOs are often as big as if not larger than states and include more hospitals across states; the largest has over 2000 hospitals, which puts it at the size of Florida. Second, the analysis conditions on hospitals being in a GPO, which results in larger and more technologically equipped hospitals – i.e. hospitals that are more likely to have any breaches at all. However, while hospitals in the same GPO might be more likely to have more vendors in common, they won't have all of their vendors in common, as we saw in Figure 28. I do include state-level fixed effects (not included in Table 11) to account for the overlap between GPOs and states.

Here, we get a negative contagion externality point estimate but without statistical significance, meaning that if another hospital is breached, another in the same GPO is less likely to be breached in the same year.

Table 12: Hazard Ratios: GPO Network Effects, OLS

	Breached	Crime	Mistake	Cyber	Physical
Perc. GPO Breached	0.969				
	(0.040)				
Perc. GPO Crime		1.003			
		(0.030)			
Perc. GPO Mistake			0.985*		
			(0.008)		
Perc. GPO Cyber				0.806	
				(0.241)	
Perc. GPO Physical					0.980
					(0.045)
Firewall	0.968	0.893	1.206	1.535	0.839
	(0.208)	(0.226)	(0.407)	(0.540)	(0.213)
Spam/Spyware Filter	0.871	0.894	0.757	0.818	0.909
	(0.174)	(0.218)	(0.242)	(0.250)	(0.220)
Encryption	1.113	1.169	0.964	0.774	1.264
	(0.195)	(0.251)	(0.255)	(0.209)	(0.260)
Single Sign-On	1.120	0.975	1.382	1.012	1.106
	(0.148)	(0.162)	(0.283)	(0.232)	(0.175)
HIE	0.704**	0.620***	0.928	0.821	0.680**
	(0.096)	(0.108)	(0.204)	(0.199)	(0.110)
EDI	1.182	1.173	1.446	1.388	1.154
	(0.233)	(0.287)	(0.507)	(0.493)	(0.276)
RIS	0.908	0.551	1.931	0.322**	1.450
	(0.295)	(0.231)	(0.847)	(0.182)	(0.456)
Basic EMR	1.385	2.797**	0.581	7.702**	0.902
	(0.444)	(1.327)	(0.240)	(7.226)	(0.265)
Advanced EMR	0.819	0.889	0.734	1.025	0.738
	(0.155)	(0.200)	(0.240)	(0.345)	(0.164)
Pseudo-R-Squared	0.07	0.10	0.09	0.13	0.07
Hospital Count	4191	4192	4192	4192	4192
Breach Type Count	296	183	124	108	199

8.2.2 Through Health Information Exchanges

Next, I look at Health Information Exchanges. Choi et al. [22] found that hospitals were more likely to experience breaches after joining an HIE, but did not offer the exact mechanism. We might expect joining an HIE to result in either more crimes, as each member hospital now has the potential to access many more records. However, we should not necessarily expect more mistakes. I use different data, however, and only can track hospitals that join HIEs in 2012. In Section 7.2 I treated joining an HIE as a control variable.

Table 13 shows the results of the HIE-level network effect. Here, I do find results in line with the scaled-attack story: for crimes and cyber breaches, the coefficient is positive and statistically significant. I again control for state-level network effects as HIEs sometimes are at the state level.

Note some caveats to the analysis: first, hospitals must be in an HIE, which cuts the sample by about a third and limits it to 2012 onwards (excluding two years of earlier breaches). We saw that joining an HIE itself has mixed effects on the hazard rate of a breach. Because HIEs share literal information, a single attack at one member hospital may implicitly affect the information at other hospitals: the channel not being a scaled attack but literally the information located in a different hospital – a hypothesis supported by the statistically significant effect in *physical* breaches as well. Further, because I condition on hospitals *being in an HIE at all* – which we found to be important in Section 7 – selection may result in hospitals that are in HIEs making common mistakes prompted by new HIE systems. One final possibility is that joining an HIE mechanically leads to more information transmitted across hospitals, and breaches occur through mistakes in that sharing process (e.g. sending an incorrect email counts as a cyber-mistake breach) and that mistakes get amplified by hospitals' requests for data from each other.

Table 13: Hazard Ratios: HIE Network Effects, OLS

	Breached	Crime	Mistake	Cyber	Physical
Perc. HIE Breached	1.036***				
	(0.012)				
Perc. HIE Crime		1.022***			
		(0.005)			
Perc. HIE Mistake			1.023		
			(0.030)		
Perc. HIE Cyber				1.046***	
				(0.011)	
Perc. HIE Physical					1.025
					(0.025)
Firewall	1.001	1.249	0.751	1.042	1.110
	(0.290)	(0.523)	(0.274)	(0.530)	(0.357)
Spam/Spyware Filter	0.902	0.838	0.998	1.511	0.757
	(0.262)	(0.321)	(0.405)	(0.767)	(0.251)
Encryption	1.177	1.310	0.910	0.868	1.209
	(0.325)	(0.455)	(0.364)	(0.352)	(0.385)
Single Sign-On	1.112	1.025	1.456	0.929	1.225
	(0.204)	(0.245)	(0.412)	(0.282)	(0.284)
HIE	1.000	1.000	1.000	1.000	1.000
	(.)	(.)	(.)	(.)	(.)
EDI	2.030**	2.087	1.965	1.995	2.113*
	(0.705)	(1.112)	(0.939)	(1.180)	(0.891)
RIS	4.623***	2.130	3.313*	1.887	3.016*
	(2.517)	(2.377)	(2.098)	(3.332)	(1.765)
Basic EMR	0.463	1.723	0.258	5.163	0.342
	(0.254)	(1.774)	(0.226)	(9.237)	(0.245)
Advanced EMR	0.975	1.002	1.249	0.610	1.238
	(0.435)	(0.556)	(0.945)	(0.369)	(0.747)
Pseudo-R-Squared	0.08	0.13	0.09	0.17	0.08
Hospital Count	2930	2939	2942	2940	2943
Breach Type Count	174	97	83	66	113

8.2.3 Through Common Technology Vendors and Software Monoculture

I next look at what is likely to be the most *direct* vector of a scaled attack: common technology vendors that result in supply-chain attacks. In the data, such supply chain attacks are not identified, but anecdotally still exist. The Allscripts and Change technologies, for example, were known to have vulnerabilities that attackers exploited multiple times. I keep the control for "Has Advanced" but remove the "Has Basic" as virtually every hospital that reports using an EMR reports having fulfilled two out of the three basic functions.

The methodology I present here allows us to detect attacks that on the surface might seem unrelated but are in fact all designed to exploit the same underlying technology. A phishing email sent to many hospital administrators, for example, might contain a virus that exploits a particular vulnerability in Allscripts software, and so the attack works if someone clicks on the email *and* the technology matches.

Table 14: Hazard Ratios: EMR Network Effects, OLS

	Breached	Crime	Mistake	Cyber	Physical
Perc. EMR Breached	1.052				
	(0.060)				
Perc. EMR Crime		1.125**			
		(0.056)			
Perc. EMR Mistake			1.054		
			(0.086)		
Perc. EMR Cyber				1.123	
				(0.195)	
Perc. EMR Physical					1.030
					(0.074)
Firewall	1.023	0.959	1.217	1.523	0.885
	(0.195)	(0.230)	(0.350)	(0.464)	(0.203)
Spam/Spyware Filter	0.765	0.767	0.742	0.690	0.770
	(0.137)	(0.179)	(0.196)	(0.178)	(0.173)
Encryption	1.153	1.322	0.875	0.915	1.300
	(0.182)	(0.259)	(0.213)	(0.229)	(0.249)
Single Sign-On	1.229*	1.166	1.392*	1.170	1.261
	(0.145)	(0.173)	(0.260)	(0.236)	(0.181)
HIE	0.819	0.748*	1.080	0.996	0.804
	(0.105)	(0.121)	(0.236)	(0.229)	(0.126)
EDI	1.077	1.055	1.153	1.322	1.037
	(0.174)	(0.211)	(0.312)	(0.393)	(0.206)
RIS	1.506*	1.278	1.945*	1.235	1.631*
	(0.357)	(0.382)	(0.728)	(0.594)	(0.431)
Advanced EMR	0.832	0.991	0.626**	1.246	0.740*
	(0.118)	(0.178)	(0.139)	(0.353)	(0.119)
Pseudo-R-Squared	0.07	0.10	0.07	0.12	0.07
Hospital Count	4211	4211	4211	4211	4211
Breach Type Count	381	239	155	138	255

Indeed, I do find that for the cases of *crimes* and *cyber* attacks there is a large effect of a peer breach on one's own breaches. That is, a 1% increase in the fraction of peers who are crime-breached is associated with multiplying the hazard rate by 1.125 of one's own crime-breach, with a similarly large effect on cyber breaches. In a crude way, we can think of that as a 12% increase in

breach probability (holding all else equal). The effect is muted for mistakes and physical breaches, and is not statistically significant for non-crime breaches.

That lack of statistical significance outside of the crime breaches aligns with the story of Section 7.2.3 – that the network effect is relevant for crime breaches in a way it simply cannot be for a mistake.

8.3 Robustness Checks

8.3.1 Instrumental Approach

The omitted variable here is the basic quality of the EMR, which may influence both its security properties and its breach proliferation. High-quality EMRs may be widely spread due to their better security properties, and also be more breached simply because they serve more hospitals. In that case, we would be *underestimating* the coefficient: EMR quality is negatively correlated with both the outcome variable (breached) and positively correlated with the explanatory variable (fraction of hospitals breached, which is related to the market share). The obvious instrument is the market share of the EMR, which should account for the quality of the EMR in that year relative to its peers normalized by its price.

Table 15 shows the results for all breaches comparing the OLS regression of the previous section with its Instrumental Variables counterpart.

The standard errors will be inconsistent due to the nonlinearity of the model, but the point estimates reflect the power of the instrument and the mechanism at hand. To address consistency, I include the residuals in the estimation, following The identifying assumption is that the market share of an EMR does not affect whether a *particular* hospital is breached *except* through the same channels that would lead to breaches at other hospitals and be therefore captured in the percentage that use the same that were also breached. That is, individual hospitals do not have some special relationship with an EMR in the dimension of security. The assumption seems reasonable if we believe the security characteristics of the products are broadly the same across hospitals, and as long as extra security is captured in the hospitals' reported security technologies.

Table 15: Hazard Ratios: EMR Network Effects, Instrumented

	Breached	Breached, IV
Perc. EMR Breached	1.052	
	(0.060)	
Perc. EMR Breached, IV		2.714*
		(1.403)
Firewall	1.023	0.986
	(0.195)	(0.190)
Spam/Spyware Filter	0.765	0.849
	(0.137)	(0.162)
Encryption	1.153	1.091
	(0.182)	(0.177)
Single Sign-On	1.229*	1.226*
	(0.145)	(0.145)
HIE	0.819	0.619**
	(0.105)	(0.120)
EDI	1.077	1.059
	(0.174)	(0.172)
RIS	1.506*	1.573*
	(0.357)	(0.370)
Advanced EMR	0.832	0.620**
	(0.118)	(0.125)
Pseudo-R-Squared	0.07	0.07
Hospital Count	4211	4211
Breach Type Count	381	381

To put it simply, results get rather unbelievable when we expand from looking at all breaches to specific categories. Specifically, the mistake and cyber coefficients – there are less than 200 breaches of each kind in the sample – become rather unbeliveably large. I attribute the issue to small sample sizes, especially those that arise when the number of positive hits – i.e. the number of cyber breaches we are actually trying to predict – is very low. Nonetheless, the IV version maintains the same pattern observed earlier, and confirms that the negative network effect I find in this section is, at worst, an *underestimation* of the true effect.

Table 16: Hazard Ratios: EMR Network Effects, Instrumented

	Breached, IV	Crime, IV	Mistake, IV	Cyber, IV	Physical, IV
Perc. EMR Breached, IV	2.714*				
	(1.403)				
Perc. EMR Crime, IV		2.211			
		(1.521)			
Perc. EMR Mistake, IV			132.745		
			(676.851)		
Perc. EMR Cyber, IV				4376.369	
				(25583.563)	
Perc. EMR Physical, IV					2.159
					(1.492)
Firewall	0.986	0.934	1.220	1.150	0.878
	(0.190)	(0.227)	(0.350)	(0.413)	(0.202)
Spam/Spyware Filter	0.849	0.829	0.770	1.174	0.807
	(0.162)	(0.210)	(0.207)	(0.574)	(0.187)
Encryption	1.091	1.273	0.861	0.760	1.258
	(0.177)	(0.259)	(0.210)	(0.232)	(0.245)
Single Sign-On	1.226*	1.175	1.258	1.141	1.261
	(0.145)	(0.175)	(0.262)	(0.232)	(0.181)
HIE	0.619**	0.729*	0.307	0.398	0.690*
	(0.120)	(0.119)	(0.412)	(0.267)	(0.141)
EDI	1.059	1.042	1.105	1.173	1.034
	(0.172)	(0.209)	(0.305)	(0.360)	(0.206)
RIS	1.573*	1.291	2.705**	1.134	1.697**
	(0.370)	(0.385)	(1.267)	(0.551)	(0.442)
Advanced EMR	0.620**	0.870	0.329	0.435	0.635**
	(0.125)	(0.180)	(0.233)	(0.345)	(0.128)
Pseudo-R-Squared	0.07	0.10	0.07	0.12	0.07
Hospital Count	4211	4211	4211	4211	4211
Breach Type Count	381	239	155	138	255

Further, we have also analyzed alternative networks of hospitals – their states, their GPOs, and their HIEs – with the finding that only HIEs and EMRs follow the scaled-attack pattern. I therefore conclude that there does exist a negative network effect – a contagion externality – specific to using the same technology as other hospitals on cybersecurity outcomes.

9 Conclusions and Next Steps

This paper contains a description of the windy yet seemingly inevitable path that the digitization of healthcare in the United States has taken, and its consequences on patient privacy and cybersecurity

outcomes. I find that the fundamental characteristics of digital technology – its scalability and the low cost of information transmission – are exactly those characteristics that enable both a concentrated market and incentivize devastating cybersecurity attacks.

I find that individual hospitals themselves do not necessarily internalize the security characteristics of the technologies they choose, focusing instead on who the regional dominant players are, the choices their GPO has negotiated with, and simply sticking with whichever vendors they contracted with in their first year of digitization. Yet it is exactly those choices which them lead to adverse cybersecurity outcomes.

Furthermore, and new to the literature, I document a *negative network effect*, also known as a *contagion externality* of using the same technology as other hospitals on one's own probability of

Further empirical analyses are forthcoming. In this paper I have not yet made us of the *individuals affected* information available for data breaches. One may expect that those vulnerabilities which are used to simply affect more individuals may be those that scale extremely well, and therefore are more likely to scale *across* hospitals as well. I also have not interacted the type of breach (crime vs. mistake) and the location (cyber vs. physical) though I do discuss above the two in conjuction.

In the next chapter of my dissertation in analyses excluded from this paper, I describe theoretically how the contagion externality transmits through hospitals' use of a common technology, using the empirical findings of this one to describe the consequences of the negative network effect on the overall security of patient records in the United States.

Two plausible interventions arise. First, simply mandating that all hospitals implement particular security technologies, or requiring EMR vendors, would seem to reduce the overall success rate of attackers. However, as Acemoglu et al. [1] describe, when multiple targets are available, increasing security of one hospital and/or software may simply *redistribute* risk to other, weaker hospitals. Furthermore, the allocation may be less efficient according to hospitals' preferences over security vs. usability and interoperability. After all, they are using their current vendor for a reason.

Alternatively, an antitrust-style policy whereby the market for hospital technologies becomes more competitive, with more options for hospitals, may serve to *break the negative network effect*: an attacker simply wouldn't be able to scale u the attack enough to get a lot of value from multiple hospitals. The tradeoff there is with the *positive network effects* of interoperability that lead hospitals to generally choose the market leader in the first place.

Future research will use the empirical findings of this paper to evaluate these two policy interventions and propose solutions for improving patient privacy and security.

10 References

References

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

[2] 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.

[3] 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.

[4] Alder, S. (2022). Eye Care Leaders Hack Impacts Millions of Patients.

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

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

[7] Atchinson, B. K. and Fox, D. M. (1997). From The Field: The Politics Of The Health Insurance Portability And Accountability Act. Health Affairs, 16(3):146–150. Publisher: Health Affairs.

[8] 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.

[9] 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.

[10] Bates, D. W., Boyle, D. L., Rittenberg, E., Kuperman, G. J., Ma'Luf, N., Menkin, V., Winkelman, J. W., and Tanasijevic, M. J. (1998a). What Proportion of Common Diagnostic Tests Appear Redundant? The American Journal of Medicine, 104(4):361–368.

[11] Bates, D. W., Cullen, D. J., Laird, N., Petersen, L. A., Small, S. D., Servi, D., Laffel, G., Sweitzer, B. J., Shea, B. F., and Hallisey, R. (1995). Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. JAMA, 274(1):29–34.

[12] Bates, D. W., Leape, L. L., Cullen, D. J., Laird, N., Petersen, L. A., Teich, J. M., Burdick, E., Hickey, M., Kleefield, S., Shea, B., Vander Vliet, M., and Seger, D. L. (1998b). Effect of Computerized Physician Order Entry and a Team Intervention on Prevention of Serious Medication Errors. JAMA, 280(15):1311–1316.

[13] *BBC* (2017). Global ransomware attack causes turmoil. June 28th, 2017. Link.

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

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

[16] Blumenthal, D. (2009). Stimulating the Adoption of Health Information Technology. New England Journal of Medicine, 360(15):1477–1479. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMp0901592.

[17] Botta, M. D. and Cutler, D. M. (2014). Meaningful use: Floor or ceiling? Healthcare, 2(1):48–52.

[18] Brennan, T. A., Leape, L. L., Laird, N. M., Hebert, L., Localio, A. R., Lawthers, A. G., Newhouse, J. P., Weiler, P. C., and Hiatt, H. H. (1991). Incidence of Adverse Events and Negligence in Hospitalized Patients. New England Journal of Medicine, 324(6):370–376. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJM199102073240604.

[19] Buntin, M. B., Burke, M. F., Hoaglin, M. C., and Blumenthal, D. (2011). The Benefits Of Health Information Technology: A Review Of The Recent Literature Shows Predominantly Positive Results. Health Affairs, 30(3):464–471. Publisher: Health Affairs.

[20] Charatan, F. (1999). Family compensated for death after illegible prescription. BMJ : British Medical Journal, 319(7223):1456.

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

[26] 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.

[27] Cutler, D. M., Huckman, R. S., and Landrum, M. B. (2004). The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery. American Economic Review, 94(2):342–346.

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

[29] 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.

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

[31] 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.

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

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

[34] 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.

[35] Freedman, S., Lin, H., and Prince, J. (2018). Information Technology and Patient Health: Analyzing Outcomes, Populations, and Mechanisms. American Journal of Health Economics, 4(1):51–79. Publisher: The University of Chicago Press.

[36] 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.

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

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

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

[40] Goldfarb, A. and Tucker, C. (2019). Digital Economics. Journal of Economic Literature, 57(1):3–43.

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

[42] Hersh, W. and Wright, A. (2008). What Workforce is Needed to Implement the Health Information Technology Agenda? Analysis from the HIMSS Analytics™ Database. AMIA Annual Symposium Proceedings, 2008:303–307.

[43] Himmelstein, D. U., Wright, A., and Woolhandler, S. (2010). Hospital Computing and the Costs and Quality of Care: A National Study. The American Journal of Medicine, 123(1):40–46.

[44] 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.

[45] 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.

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

[47] Jha, A. K., DesRoches, C. M., Kralovec, P. D., and Joshi, M. S. (2010). A Progress Report On Electronic Health Records In U.S. Hospitals. Health Affairs, 29(10):1951–1957. Publisher: Health Affairs.

[48] Jones, S. S., Rudin, R. S., Perry, T., and Shekelle, P. G. (2014). Health Information Technology: An Updated Systematic Review With a Focus on Meaningful Use. Annals of Internal Medicine, 160(1):48–54. Publisher: American College of Physicians.

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

[50] Katz, M. L. and Shapiro, C. (1986). Technology Adoption in the Presence of Network Externalities. Journal of Political Economy, 94(4):822–841. Publisher: The University of Chicago Press.

[51] 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.

[52] Kwon, J. and Johnson, M. E. (2014). Proactive Versus Reactive Security Investments in the Healthcare Sector. MIS Quarterly, 38(2):451–471.

[53] 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.

[54] Larson, N. (2011). Network security. Working Paper.

[55] Leape, L. L., Brennan, T. A., Laird, N., Lawthers, A. G., Localio, A. R., Barnes, B. A., Hebert, L., Newhouse, J. P., Weiler, P. C., and Hiatt, H. (1991). The Nature of Adverse Events in Hospitalized Patients. New England Journal of Medicine, 324(6):377–384. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJM199102073240605.

[56] Lee, J. and Choi, S. J. (2021). Hospital Productivity After Data Breaches: Difference-in-Differences Analysis. Journal of Medical Internet Research, 23(7):e26157.

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

[58] Miller, A. R. and Tucker, C. E. (2011). Can Healthcare IT Save Babies? SSRN Scholarly Paper 1080262, Social Science Research Network, Rochester, NY.

[59] Miller, R. H., West, C., Brown, T. M., Sim, I., and Ganchoff, C. (2005). The Value Of Electronic Health Records In Solo Or Small Group Practices. Health Affairs, 24(5):1127–1137. Publisher: Health Affairs.

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

[61] Neiman, B. and Vavra, J. (2023). The Rise of Niche Consumption. American Economic Journal: Macroeconomics, 15(3):224–264.

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

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

[64] Parente, S. T. and McCullough, J. S. (2009). Health Information Technology And Patient Safety: Evidence From Panel Data. Health Affairs, 28(2):357–360. Publisher: Health Affairs.

[65] Ransbotham, S., Overby, E. M., and Jernigan, M. C. (2021). Electronic Trace Data and Legal Outcomes: The Effect of Electronic Medical Records on Malpractice Claim Resolution Time. Management Science, 67(7):4341–4361. Publisher: INFORMS.

[66] Rodriguez-Vera, F. J., Marin, Y., Sanchez, A., Borrachero, C., and Pujol, E. (2002). Illegible handwriting in medical records. Journal of the Royal Society of Medicine, 95(11):545–546.

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

[68] Schneier, B. (2008). Schneier on Security. Wiley.

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

[70] 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.

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

[72] 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.

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

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