View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by AIS Electronic Library (AISeL)

IS PATIENT DATA BETTER PROTECTED IN COMPETITIVE HEALTHCARE MARKETS?

Completed Research Paper

Martin S. Gaynor
Heinz College
Carnegie Mellon University
Pittsburgh, PA
mgaynor@cmu.edu

Muhammad Zia Hydari
Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA
zia@cmu.edu

Rahul Telang
Heinz College
Carnegie Mellon University
Pittsburgh, PA
rtelang@andrew.cmu.edu

Abstract

We study the effect of hospital market concentration on the quality of patient data protection practices. We use approximately 200 reported data breaches in US hospitals over the period 2006 - 2011 as a measure of the quality of patient data protection practices. We measure market concentration using the Herfindahl-Hirschman Index (HHI) and estimate our models by exploiting cross-sectional HHI variation. Surprisingly, we find that increased competition is associated with a decline in the quality of patient data protection. Our main result indicates that a 100 point increase in HHI is associated with a 5% decline in the average count of data breach incidents. The results are directionally robust to a number of alternate model specifications. To explain our findings, we posit that hospitals in more competitive markets may be inclined to shift resources to more consumer visible activities from the less consumer visible activity of data protection.

Keywords: IS Security and Privacy, health information systems, competition

Introduction

The venerable Stanford Hospital recently found itself subjected to negative publicity and a $20 million lawsuit due to breach of confidential patient data. A contractor's actions led to the online exposure from September 2010 to August 2011 of 20,000 of Stanford Hospital patient records (Sack 2011). Besides negative publicity and loss of consumer confidence, there may be stiff financial penalties associated with breach of patient data records. In June 2010, five California hospitals were fined $675,000 for breach of confidential patient data. These fines were levied despite the fact that some of the hospitals discovered these breaches themselves during audit and disclosed the breaches[1].

The privacy of health information is a widely accepted notion in the United States (and many Western countries). As of August 2011, legistlation in 46 US states requires hospitals to report data breaches to the individuals affected by such breaches.

We study the effects of firm competition on the quality of patient data protection practices in hospital markets. Information security and privacy have generally been important issues in many societies. However, the widespread adoption of information and communication technologies since the second last decade of the twentieth century has brought information security and privacy to the forefront of societal concerns. With businesses now able to electronically collect, store, and distribute vast amounts of data about consumers, there is a need to institute proper safeguards to protect consumer information.

We consider information security and privacy an aspect of firms' overall quality. We are not aware of any studies that assess the impact of market competition on the quality of information security and privacy. In addition, economic theory is indeterminate on the effect of competition on the likelihood or severity of data breaches at firms. A researcher can make two competing claims based on economic theory, viz. (i) firms with market power may consume their profits and underinvest in activities that thwart data breaches, whereas firms facing tough competition may compete on information security quality (ii) firms with market power may be less conscious about the cost of less visible activies such as data breaches, whereas firms facing tough competition may cut costs on activities that thwart data breaches, as these activities are less visible to consumer. As economic theory cannot provide strong guidance, the impact of competition on information security is an empirical question.

The difficulties with modeling the association between market competition and firm's quality decisions include properly defining markets, measuring competition, and incorporating appropriate controls. We have chosen United States' hospital markets as the context for our study. An obvious advantage of hospital markets is that they have relatively crisp geographical boundaries, which facilitate the creation of measures of market structure and relevant controls. Moreover, both the issue of competition and patient health information privacy within the healthcare markets have been important to consumers, businesses, and governments.

Our study examines the effect of competition in hospital markets on the quality of patient data protection practices. We use approximately 200 reported data breaches in US medical facilities over the period 2006 - 2011 as a measure of quality of patient data protection practices. We use Core Based Statistical Area (CBSA) both to define hospital markets and as the primary unit of our analysis. We use the number of data breach incident reports as a measure of quality and Herfindahl-Hirschman Index (HHI) as a measure of competition within each CBSA. We also control for market size, population, population over 65 years, per capita income, and state-level variation in data breach disclosure laws within each CBSA. We also carry out secondary analysis at hospital level, while controlling for hospital specific characteristics such as hospital size, number of hospital employees, and so on.

Surprisingly, we find that increased competition is associated with a decline in the quality of patient data protection. Our main result indicates that a 100 point increase in HHI is associated with a 5% decline in the average count of data breach incidents. The results are directionally robust to a number of alternate model specifications. To explain our findings, we posit that hospitals in more competitive markets may be

[1] http://web.archive.org/web/20100822093407/http://seattletimes.nwsource.com/html/health/2012114943_fines14.html accessed on September 26, 2011

inclined to shift resources to more consumer visible activities from the less consumer visible activity of data protection.

Related Literature

Economic theory is clear that competition increases quality in regulated markets (with prices above marginal costs). When firms set prices and choose quality, the effect of competition on quality is ambiguous (Gaynor 2006). Although Medicare regulates a segment of the healthcare market, the overall healthcare market is not regulated. In our study, we do not assume that the regulated and non-regulated segments are separable. Although our model is informed by economic theory, our study prioritizes signifiance and fit of statistical model over economic theory, and lets the data "speak"'.

There is a vast and growing body of management literature both on electronic data security in general and data security within healthcare IT in particular. We cite a few recent and relevant articles. (Romanosky et al. 2011) report that data breach disclosure laws led to a decline in identity theft during the 2002 to 2009 study period. Using analytical and numerical modeling, (Romanosky et al. 2010) find that data breach disclosure laws may increase firm costs but may lower social costs. (Miller et al 2011) paradoxically found that encryption of patient data may actually increase publicized data loss. Our study contributes by examining the effects of market competition on data losses in healthcare markets.

As mentioned earlier, theory does not provide a clear-cut answer on how competition would effect IT security and data protection quality, both in general markets as well as healthcare markets. The extant literature does not address the question of competitive impact on the quality of IT security and data protection practices. We empirically examine this question in the context of healthcare markets.

Data Sources and Variable Construction

We sourced hospital data breach information from Privacy Rights Clearing House[2], a non-profit consumer organization with a stated mission of consumer education and consumer advocacy on issues of personal privacy. We sourced data for explanatory variables from American Hospital Association (AHA) yearly surveys, U.S. Department of Health and Human Services Area Resource File (ARF), and Healthcare Information and Management Systems Society(HIMSS) 2009 Analytics Database. Table 1 provides summary statistics for select variables in our dataset.

Table 1: Summary Statistics (2006 - 2011)

Variable	**Mean**	**(Std. Dev.)**	**Min.**	**Max.**
Data Breach Incidents	0.517	(1.385)	0	15
Herfindahl-Hirschman Index (HHI)	4022.472	(2477.683)	307.371	10000
Market Size (Hospital Beds)	2111.536	(3633.667)	99	44120.333
CBSA Population	653217.139	(1154605.876)	55357	11553017.833
CBSA Population (eligible for medicare)	92274.541	(147367.708)	4425	1542206
CBSA Population (> 65 years)	76341.018	(129489.108)	3103	1434894
Per Capita Income	0495.358	(6057.12)	15748.333	64219.333
Disclosure Law Effective Days	1732.969	(719.009)	0.001	2974
N	381			

We used Privacy Rights Clearing House as the source of data for data breaches. Our starting dataset includes all incidents reported from January 1, 2006 up to August 22, 2011 across all industries, but the

[2]https://www.privacyrights.org/about_us.htm

focus of our study is data breaches at hospitals. First, we only retained data breaches at the hospital level and filtered out all other data breach reports including those related to private doctor offices and health insurance companies. Second, we programmatically and manually mapped the hospitals to American Hospital Association's unique hosptial ID for each hospital with a reported data breach. We constructed the following variables, which we use as the dependent variable in a number of our models:

Incidents: The number of data breach incidents that happened at the hospital and CBSA level during the period (2006-2011). The number of incidents at the CBSA is merely an aggregate of all incidents at the hospitals within that CBSA. This is a count variable.

Incident: Whether a data breach incident happened at the hospital and CBSA level during the period (2006 - 2011). This is a binary variable.

Severity: The severity of the data breach coded into three categories, given in Table 2 below in increasing order of severity. *Severity* is an ordinal variable. Table 2 provides a frequency distribution of the levels of severity.

Number of Records: The number of records that were breached during a single incident. This is a count variable.

Table 2: Distribution of the *Severity* of Breaches

Item	Number	Per cent
No breach	281	79
Breach without financial	11	3
Breach with financial	62	18
Total	354	100

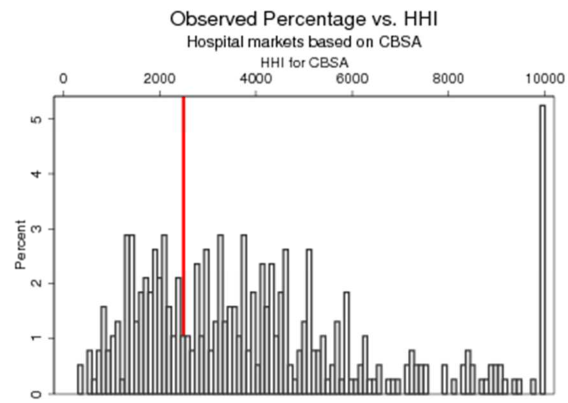

Figure 1: Distribution of HHI

We used American Hospital Association (AHA) yearly survey data set to compute the market size and the Herfindahl-Hirschman Index (HHI) for each CBSA in our study. We note that hospital-beds and patient-days (and similar measures) are highly correlated so HHI based on any of these measures will not effect results. We used the 'number of hospital beds' to define market size as “number of hospital beds" seem to be a commonly used market size definition in extant literature. We construct the following variables:

Market Size: The total number of hospital beds in a CBSA.

HHI: The sum of squared market shares for hospitals in CBSA, with market shares computed using hospital beds.

As of August 2010, US Federal Trade Commission (FTC) threshold for “highly concentrated”' markets is an HHI of over 2500[3]. Figure 1 shows a histogram of mean HHI, with the HHI=2500 points marked by the red vertical line. Figure 1 suggests that most hospital markets are highly concentrated.

We used ARF to compute CBSA-level demographic information such as population, population older than 65, population eligible for medicare, and per capita income. The demographics are reported at the county level in ARF, which we aggregated to the CBSA level.

For hospital level analysis, we further augmented our data with hospital IT variables. We used the Healthcare Information and Management Systems Society 2009 Analytics Database (HIMSS DB) as a

[3] Capps, Dranove: Market Concentration of Hospitals http://www.ahipcoverage.com/wp-content/uploads/2011/06/ACOs-Cory-Capps-Hospital-Market-Consolidation-Final.pdf.

data source on hospital information technology. We found that all hospitals included in the HIMSS DB had Electronic Medical Records (EMR) in the year 2009, so we do not include EMR adoption as an independent variable in our analysis. The following hospital-level IT variables were used in our analysis:

FTE: Full-time equivalent IT employees working at the hospital

FTE Security: Full-time equivalent employees working in IT security

FTE EMR Help: Full-time equivalent employees working to support Electronic Medical Records

The reporting of data breaches is potentially influenced by mandatory data breach disclosure laws, where the legislation is at the geographical states level rather than at the federal level. As the state-level legislation was passed at different times, we constructed a control variable "Law Effective Days", that counts the days elapsed since the mandatory disclosure legislation was passed.

Modeling Framework and Empirical Analysis

We consider patient data protection an aspect of hospital overall quality. Economic theory does not provide clear-cut answers on whether competition will have a positive or negative impact on quality within a non-regulated market (such as hospital markets). Popular perception that "competition is good" notwithstanding, increased competition may lead to decreased quality in certain circumstances. Empirical evidence is also ambiguous (Gaynor 2006).

Most empirical studies on quality and competition within healthcare employ the structure-conduct-performance (SCP) framework, which assumes a causal link between market structure, firm conduct, and industry performance. Market structure is usually measured by Herfindahl-Hirschman Index (HHI), whereas firm conduct is either price (if firms have market power) or some measure of firm-chosen quality. The specification also includes demand-shifters and cost-shifters as controls (Gaynor 2006). The econometric specification usually has the following form:

$$Quality = \beta_0 + \beta_1(DemandShifters) + \beta_2(CostShifters) + \beta_3(HHI) + \varepsilon \quad (1)$$

We employ a similar specification in our study.

We are primarily interested in the association between average count of data breaches and HHI, identified through the observed heterogeneity at the CBSA-level. This in turn helps us understand the association between firm's data security quality choice in the face of market competition. Our observed dependent variable is count of data breaches, which naturally suggests a Poisson regression model (PRM) as our basic modeling framework. Due to potential overdispersion, we also estimate negative binomial regression models (NBRM)[4]. For the mean parameters, the Poisson maximum likelihood estimator (MLE) is fully robust to distributional misspecification. Poisson MLE also maintains some efficiency properties when the distribution is not Poisson (Wooldridge 2002). We use Poisson MLE in software package Stata for parameter estimation.

Market characteristics such as market size, population, average income of the residents, and others may affect the number of breaches as well as HHI (e.g. one would expect more data breaches in a CBSA with a relatively large population). In our models, we control for these observed market characteristics. To further address endogeneity concerns due to omitted variables, we conduct our analysis at the hospital-level with more controls, especially IT controls, at the hospital level (as described in a later section). We are unable to include aggregate measures such as hospital IT adoption at the CBSA-level as it is not possible to construct correct aggregate measures due to missing data at hospital-level.

Our basic modeling framework can be summarized as:

$$DBIR_i \sim Poisson(\lambda_i) \quad (2)$$

[4]NBRM is equivalent to over-dispersed PRM in our setting because of the mean zero assumption on the over-dispersion parameter.

$$\lambda_i = \exp[\beta_0 + \beta_1 HHI_i + \beta_2 (MarketSize)_i + \beta_3 Population_i + \beta_4 Income_i] \quad (3)$$

For the overdispersed case, we assume an error δ_i with the following relation with λ_i. We also assume $E(\delta) = 1$ (for identification) and a gamma distribution with parameter α_i (Long 1997).

$$\lambda_i = \exp[X\beta + \ln\delta_i] \quad (4)$$

$$\delta_i \sim Gamma(\alpha_i, \alpha_i) \quad (5)$$

δ may be viewed as the combined effect of omitted variables (Gourieroux 1984) or a source of randomness (Hausman 1984). A likelihood ratio test for the hypothesis $H_0: \alpha = 0$ provides a test for overdispersion (Long 1997).

Main Results

For our primary results, we use variants of models described in Equations (2), (3), (4), and (5). The dependent variable in all of these models is count of data breach incident reports measured at the CBSA-level. Almost all of the explanatory variables in these models (including the focal predictor *HHI*) are measured at the CBSA level, except that the variable *Law Effective Days* is measured at the geographical state level. We did not include geographical state indicators in these models as we include state-level *Law Effective Days* (linearly dependent on geographical state indicators). The direction of HHI is unaffected even if we include state-level indicators and drop *Law Effective Days*. We estimated the models using Stata with observed information matrix (OIM) variance-covariance estimator, and the results in Table 3 provide a comparison of estimates on various models. The estimates on *HHI* coefficients are directionally similar and statistically significant in all of these models. We find that an increase in HHI is associated with a decrease in average count of data breaches.

Discussion

Table 4 provides detailed results on model NBRM7, which is essentially the same as model NBRM6. The minor difference is that NBRM uses scaled independent variables (HHI, market size, and population) to facilitate discussion. We choose Negative-binomial regression model over Poisson regression model as the likelihood-ratio test of $H_0: \alpha = 0$ reports a chi-square statistic $\chi^2 = 18.22$ with $P[\chi^2 \geq 18.22 | H_0] = 0.000$ providing evidence for overdispersion[5]. Column "factor" in Table 4, reported as 0.949 provides a multiplicative interpretation of the variable effects. Surprisingly, we find that a 100 points increase in *HHI* leads to a 5% decrease in average count of data breaches. A more intuitive example to explain the estimated coefficient is that the change in market from five equally-sized firms (with $HHI = 2000$) to four equally-sized firms (with $HHI = 2500$) is associated with an approximately 25% decrease in the average count of data breaches.

One plausible explanation for the observed decline in data protection quality is that with increased competition, hospitals allocate more resources to customer observable activities and cut costs on less observable activities such as customer data protection. By focusing resources on relatively more observable activities, the hospitals tradeoff between current revenue vs. risk of a data breach.

Economic theory can explain the direction of impact when firms facing competition choose between two or more strategic variables (e.g. a more observable quality attribute and a less observable one). Dorfman-Steiner condition suggests that firms allocate resources depending on the elasticities of strategic variables. Dranove and Satterthwaite provide a similar intuition, where they find that better consumer information about price than quality has an equilibrium with sub-optimal quality provided by the firm. We can draw upon the insights from Dorfman-Steiner condition and Dranove and Satterthwaite when explaining our results. Depending on relative elasticities of various aspects of quality, hospitals may shift resources between activities related to different aspects of quality (Gaynor 2006).

Table 4: Detailed Results with DV=incidents

[5] Stata reports *chibar2(01)=18.22* and $Prob >= chibar2 = 0.000$

	NBRM7			
	b	se	p	factor
incidents				
HHI (by 100)	-5.19e-02***	(9.00e-03)	0.00	0.94940
Size (by 100)	2.20e-03	(6.69e-03)	0.74	1.00220
Population (by 1000)	1.86e-04	(2.18e-04)	0.39	
Population (over 65 years, by 1000))	-4.67e-05	(2.42e-03)	0.98	
Per Capita Income	2.75e-05	(1.51e-05)	0.07	
Log(Law Effective Days)	-3.38e-02	(2.62e-02)	0.20	
vce	oim			
N	380			

A hospital's quality of IT security and data protection is largely invisible to consumers, whereas quality of clinical and hotel aspects of the hospital are relatively more observable. It is reasonable to expect that the ratio between the data protection quality elasticity of demand and clinical quality elasticity of demand will be higher for a monopoly than for firms in competitive markets. This in turn implies that relataively more resources will be spent on data security in a monopoly than in competitive markets. Empirically, (Mukamel et al. 2002) found that intensifying price competition may lead hospitals to allocate more resources to services whose quality customers can more easily evaluate.

Table 3: Main Results with DV = Incidents								
	PRM1	PRM2	PRM3	PRM3a	PRM3b	PRM4	PRM5	NBRM6
	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p
incidents								
HHI	-9.09e-04***	-6.46e-04***	-6.34e-04***	-6.33e-04***	-6.24e-04***	-5.77e-04***	-5.76e-04***	-5.19e-04***
	(7.24e-05)	(7.34e-05)	(7.68e-05)	(7.68e-05)	(7.77e-05)	(7.94e-05)	(7.96e-05)	(9.00e-05)
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Market Size		5.31e-05***	4.00e-05	5.19e-05	1.66e-05	3.18e-05	2.75e-05	2.12e-05
		(7.10e-06)	(2.69e-05)	(4.31e-05)	(4.04e-05)	(2.75e-05)	(2.77e-05)	(5.30e-05)
		0.00	0.14	0.23	0.68	0.25	0.32	0.69
Population			4.94e-08	8.15e-08	-5.38e-08	7.13e-08	8.98e-08	1.83e-07
			(9.77e-08)	(1.33e-07)	(1.65e-07)	(1.00e-07)	(1.01e-07)	(1.75e-07)
			0.61	0.54	0.74	0.48	0.37	0.29
Population (over 65 years))				-6.28e-07				
				(1.77e-06)				
				0.72				
Population (Medicare eligible)					1.47e-06			
					(1.90e-06)			
					0.44			
Per Capita Income						2.46e-05*	2.66e-05*	2.74e-05
						(1.08e-05)	(1.09e-05)	(1.50e-05)
						0.02	0.02	0.07
Log(Law Effective Days)							-3.51e-02	-3.39e-02
							(2.24e-02)	(2.60e-02)
							0.12	0.19
vce	oim	oim	oim	oim	oim	oim	oim	oim
N	380	380	380	380	380	380	380	380

Robustness

As mentioned earlier, mean estimates from Poisson MLE are robust to distributional assumptions. Table 5 provide further evidence of functional form robustness of our results. In model PRM8 and model NBRM9, we include a quadratic control variable by adding $(MarketSize)^2$. Model *PRM10* includes indicator variables for US geographical states where the highest percentage of CBSA population resides, but drops *Law Effective Days* (due to linear dependence). We do not show the estimates for the indicator variables in Table 2 to conserve space. Finally, OLS11 estimates a log-linear model using Ordinary Least Squares, with $Log(Incidents)$ as the dependent variable.

Table 5: DV=Log(Incidents) for OLS, DV=Incidents for Other Models

	PRM8	NBRM9	PRM10	OLS11
	b/se	b/se	b/se	b/se
main				
HHI	-4.56e-04***	-3.89e-04***	-5.03e-04***	-1.52e-04*
	(8.55e-05)	(9.32e-05)	(9.73e-05)	(6.83e-05)
Size	1.65e-04**	1.92e-04*	3.20e-04***	4.87e-04**
	(5.48e-05)	(7.71e-05)	(7.99e-05)	(1.82e-04)
Size Squared	-1.96e-09**	-3.15e-09**	-3.49e-09***	-1.37e-08***
	(6.87e-10)	(1.08e-09)	(9.29e-10)	(2.69e-09)
Population	-8.41e-08	2.43e-08	-3.85e-07*	8.95e-07*
	(1.15e-07)	(1.79e-07)	(1.78e-07)	(4.48e-07)
Per Capita Income	2.60e-05*	2.33e-05	5.90e-06	1.25e-05
	(1.12e-05)	(1.50e-05)	(1.41e-05)	(2.47e-05)
Log(Law Effective Days)	-3.56e-02	-3.48e-02		-3.65e-02
	(2.26e-02)	(2.60e-02)		(3.94e-02)
vce	oim	oim	oim	ols
N	380	380	380	380

Looking at Table 5, we again find that an increase in HHI (i.e. market concentration) is correlated with a decrease in the average count of data breaches. The results are economically and statistically significant in both linear and non-linear models.

Alternate Models at CBSA-level

To further examine the association between competition and data protection practices, we estimate models where the dependent variable is different from the our main models (see Section Main Results). Specifically, we use either *Incident*, or *Severity*, or *Number of Records* as dependent variable in the models that appear in the current section. If our main result holds, we expect to find that these alternative models will also suggest an association between higher competition and lower data security quality.

Association between the Odds of an Incident and HHI

We first look at the likelihood of the occurence of a data breach incident in a CBSA. We expect to find that increasing HHI (i.e. lower competition) is associated with a lower likelihood of data breaches (higher data security quality). We examined the association of the odds of a data breach incident occuring in a CBSA and HHI using a logit model. Table 4 summarizes the results of the estimated model. The estimates suggest that an increase of 100 point in HHI is associated with a decrease of 2.6% in the odds of an incident occuring in the CBSA. We used a robust cluster variance estimator, where the clustering was done on the geographical states. The estimate for HHI are not statistically significant at the 5% level, although the p-value is just 6%. The association is consistent with our main results i.e. increase in competition is correlated with a decline in the quality of data protection practices.

Association between Number of Records Breached and HHI

Next, we look at the association between the number of records breached and HHI. While all breaches are unacceptable, some breaches are worse than other due to larger number of records breached. Although number of records breached are counts and a Poisson regression model may seem natural, we can approximate using a normal distribution due to the high average count. Thus, we use a linear specification with log of the number of records breached as the dependent variable. We used Ordinary Least Squares with robust cluster variance estimator, with CBSA as the cluster variable. Table 6 summarizes the model estimates.

The coefficient on HHI is statistically significant and the magnitude suggests that a 100 points increase in HHI is associated with an approximately 25% decrease in the number of records breached (ceteris paribus). A clarification about sample size is in order - we deleted[6] 24 observations where the number of records breached was missing in the original data set. Therefore, the sample size is $N = 354$ rather than $N = 380$ as in the earlier models.

Association between Severity and HHI

Finally we look at the association between severity of breach and HHI, where severity was defined earlier (see subsection **Error! Reference source not found.**). The dependent variable for this model is the ordinal variable $Severity$. We estimate an ordered logistic model to measure the association between the severity of breaches and HHI. The robust cluster variance estimator uses CBSA for clustering. Table 7 summarizes the estimated model. The coefficient on HHI imply a multiplicative factor of approximately 0.97. The interpretation would be that a 100-point increase in HHI is associated with a 3% decrease in a $(Severity = 3)$ incident vs. $(Severity < 3)$ incident. Again, we find that an increase in competition is associated with a decrease in the quality of data protection practices.

[6]We deleted observations with missing data rather than using imputation methods.

Table 5: Incident Logit (CBSA)			
	LOGIT		
	b	p	factor
incident			
HHI (by 100)	-2.62e-02	0.06	0.97415
Size (by 100)	5.01e-03	0.77	1.00502
Population (by 1000)	1.63e-03**	0.00	
Per Capita Income	-2.42e-02	0.34	
vce	cluster		
N	380		

Table 6: Number of Records Breached (CBSA)		
	OLS	
	b	p
HHI (by 100)	-2.52e-01***	0.00
Size (by 100)	-1.14e-01	0.64
Population (by 1000)	1.42e-02	0.05
Per Capita Income	4.72e-01	0.12
Law Effective Days	1.02e-03	0.64
vce	cluster	
N	354	

Table 7: Severity (CBSA)				
	OLOGIT			
	b	se	p	factor
severity				
HHI (by 100)	-2.98e-02*	(1.34e-02)	0.03	0.97065
Size (by 100)	1.58e-02	(1.97e-02)	0.42	1.01593
Population (by 1000)	2.25e-04	(4.14e-04)	0.59	
Per Capita Income	3.15e-02	(2.31e-02)	0.17	
Law Effective Days	1.42e-04	(2.34e-04)	0.54	
cut1				
Constant	2.09e+00*	(9.00e-01)	0.02	
cut2				
Constant	2.36e+00**	(9.00e-01)	0.01	
vce	cluster			
N	354			

Alternate Models at Hospital-level

As our final set of models, we investigate the association between HHI and the quality of data protection by using hospitals as the unit of our analysis. By analyzing at the hospital-level, we are better able to control for hospital specific heterogeneity. As mentioned earlier, we have two sources for hospital level variables viz. AHA annual survey and HIMSS 2009 Analytics Database. We merge AHA and HIMSS data on Medicare number, which results in the loss of hundreds of observations. Finally, there are missing entries for Information Systems department full-time employee (FTE) counts, which results in further loss of observations in the models that follow.

If more competition is indeed associated with lower IT security quality, then we would find the odds of a breach to be higher in more competitive markets. To investitage this, we estimate the odds of a breach at a hospital given the HHI for the CBSA (and other control variables) using logit models. Table 8 summarizes the estimates for a number of logit models. We used a robust cluster variance estimator, where the clustering was done on the CBSA. In addition for the control variables shown in Table 8, we also control for hospital ownership type, hospital system membership, and hospital JCAHO[7] accreditation status.

Except for model L1, the coefficient estimates are not statistically significant and cannot be used for general inference. As descriptive statistic for the given sample, all models suggest a negative association between HHI and the odds of a breach occuring at a hospital.

Table 8: Incident Logit (Hospital)

	L1	L2	L3	L4	L5	L6
	b/se	b/se	b/se	b/se	b/se	b/se
incident						
HHI (by 100)	-1.82e-02	-7.37e-03	-1.27e-02	-1.15e-02	-1.08e-02	-1.64e-02
	(6.83e-03)	(7.17e-03)	(7.62e-03)	(7.60e-03)	(1.10e-02)	(1.11e-02)
CBSA Size (by 100)		-2.91e-03	-8.07e-04	-4.55e-03	-1.12e-02	-1.31e-02
		(4.83e-03)	(4.70e-03)	(4.62e-03)	(8.81e-03)	(1.02e-02)
Population (by 1000)		1.95e-04	2.74e-04	2.03e-04	3.17e-05	-1.91e-04
		(1.14e-04)	(1.10e-04)	(9.76e-05)	(2.17e-04)	(2.41e-04)
Population (> 65, by 1000)		-1.21e-03	-2.43e-03	-6.18e-04	3.30e-03	5.49e-03
		(1.19e-03)	(1.55e-03)	(1.46e-03)	(2.97e-03)	(3.21e-03)
Income (by 1000)		2.25e-02	1.10e-02	5.31e-03	6.68e-03	-2.85e-02
		(1.23e-02)	(1.25e-02)	(1.21e-02)	(1.76e-02)	(3.17e-02)
System Size (in CBSA)		5.74e-04	1.31e-04	6.79e-05	2.15e-04	4.81e-04
		(1.34e-04)	(1.57e-04)	(1.74e-04)	(1.91e-04)	(1.88e-04)
Hospital Size (by 100)			2.99e-01	2.43e-01	2.46e-01	2.63e-01
			(3.76e-02)	(4.87e-02)	(7.13e-02)	(1.24e-01)

[7] ``An independent, not-for-profit organization, The Joint Commission accredits and certifies more than 19,000 health care organizations and programs in the United States. Joint Commission accreditation and certification is recognized nationwide as a symbol of quality that reflects an organization's commitment to meeting certain performance standards" (Source: http://www.jointcommission.org/about_us/about_the_joint_commission_main.aspx, accessed on Jan 03, 2012).

Number of FTE				1.55e-04	3.72e-05	-6.86e-05
				(5.73e-05)	(6.68e-05)	(1.89e-04)
FTE in IS					3.25e-03	8.49e-03
					(1.36e-03)	(3.62e-03)
FTE in IS Security						8.74e-02
						(7.64e-02)
FTE in EMR Support						-2.46e-02
						(9.91e-03)
Intercept	-2.79e+00	-4.08e+00	-5.64e+00	-4.99e+00	-3.76e+00	-2.67e+00
	(1.61e-01)	(4.70e-01)	(7.48e-01)	(8.16e-01)	(7.45e-01)	(1.14e+00)
	0.00	0.00	0.00	0.00	0.00	0.02
VCE	cluster	cluster	cluster	cluster	cluster	cluster
N	4040	4040	4040	2646	995	562

Evidence on Higher Observable Quality in Competitive Market

To expain the observed association of better data security quality with less competitive markets, we have profferred that hospitals in competitive markets may shift resources towards activities whose quality is more easily observable to the consumers. Objectively measuring overall hospital quality is not easy and healthcare researchers have usually focused on the outcomes related to very specific ailments such as pneumonia or acute myocardial infarction. Instead of focusing on the outcome of specific ailments, we instead investigate the relation between HHI and highly-visible and costly quality signals such as existence of a residency program, medical school affiliation, and membership in the Council of Teaching Hospitals (COTH). Since the observed output variables (medical school affiliation and so on) are binary in nature, it is natural to model the odds of participation (success) through a logit specification.

Table 9 summarizes the estimates, where Models Q1, Q2, and Q3 have existence of residency program, medical school affiliation, and COTH membership as output variables. We include the usual controls but focus on the sign and statistical significance of the coefficient of HHI in these models. For all of these models, we find that more competitive markets are associated with higher odds of success on these binary variables of highly visible quality signals.

While these models do not offer conclusive proof of our resource shifting hypothesis, the associations suggested by models Q1, Q2, and Q3 along with the theoretical models from extant literature[8] add to the plausbility of our argument.

Table 9: Observed Hospital Quality versus HHI

	Q1	Q2	Q3
	b/se	b/se	b/se
main			

[8]Dorfman-Steiner condition and Dranove-Satterthwaite model, see (Gaynor 2006).

HHI (by 100)	-2.00e-02	-1.48e-02	-2.85e-02
	(3.87e-03)	(3.31e-03)	(7.27e-03)
CBSA Size (by 100)	9.58e-03	9.20e-03	6.52e-03
	(2.71e-03)	(2.53e-03)	(3.91e-03)
Population (by 1000)	-1.59e-04	-2.41e-04	-1.23e-04
	(8.21e-05)	(7.66e-05)	(1.22e-04)
Population (> 65, by 1000)	-1.81e-03	-8.14e-04	-9.39e-04
	(9.97e-04)	(9.22e-04)	(1.52e-03)
Income (by 1000)	3.39e-02	3.35e-02	3.39e-02
	(7.30e-03)	(6.86e-03)	(1.08e-02)
Hospital Size (by 100)	6.07e-01	5.84e-01	6.21e-01
	(2.63e-02)	(2.58e-02)	(3.07e-02)
Owner Type	-3.98e-01	-3.63e-01	-4.05e-01
	(4.44e-02)	(4.12e-02)	(6.81e-02)
Intercept	-1.99e+00	-1.76e+00	-3.68e+00
	(3.18e-01)	(2.95e-01)	(4.98e-01)
N	4040	4040	4040

Conclusion

We find a robust association between increase in competition and decrease in the quality of patient data protection practices. We find the association to hold at:

Unit of Analysis: CBSA-level analyis or hospital-level analysis

Outcome Measure: Count of incidents, odd of an incident, severity of breach, number of records breached

Functional Forms: Model specification was driven by the dependent variable but we used different functional forms in our specification

Our main result indicates that a 100-point increase in HHI is associated with approximately 5% decline in the average count of data breach incidents at the CBSA level. We find statistically and economically significant support of this finding through a number of other models that we report in this article.

As explanation for our finding, we posit that hospitals in competitive markets may shift resources to more visible activities (such as medical and hotel services) away from less visible activities such as data security. In doing so, the hospitals may increase the risk of data breaches. We find support for this resource shifting explanation both from economic theory (Dorfman-Steiner condition) and from empirical research (Mukamel 2002). We also find some support for our resource shifting hypothesis from the data as we observe hospital in more competitive markets (ceteris paribus) to have higher likelihood of costly high quality signals such as residency programs, medical school affiliation, and membership in Council of Teaching Hospitals (COTH).

Our finding may have interesting policy implications. The extant policy has been to let hospitals decide on the level of data security investments and only penalize when a data breach is reported. Although not without its own complications and unintended effects, an alternate policy route would be to require certification to a minimum level of compliance to data protection practice.

Our finding may also have indirect implications for general managers and IT managers. While firms in competitive markets may be maximizing profits in expectation, they may be miscalculating the risks of a future breach and thus underinvesting in IT and data security. This may open firms to future losses that have not been correctly anticipated.

References

Dranove, D. and Shanley, M. and Simon, C. Is hospital competition wasteful?. *The Rand journal of economics*, :247--262, 1992.

Gaynor, M. *What do we know about competition and quality in health care markets?*. National Bureau of Economic Research Cambridge, Mass., USA, 2006.

Miller, A.R. and Tucker, C.E.. Encryption and the loss of patient data. *Journal of Policy Analysis and Management*, 2011.

Mukamel, D. and Zwanziger, J. and Bamezai, A. Hospital competition, resource allocation and quality of care. *BMC Health Services Research*, 2(1):10, 2002.

Romanosky, S. and Sharp, R. and Acquisti, A. Data Breaches and Identity Theft: When is Mandatory Disclosure Optimal?. *Ninth Workshop on the Economics of Information Security (WEIS 2010), Harvard University, Cambridge, MA*, 2010.

Romanosky, S. and Telang, R. and Acquisti, A. Do data breach disclosure laws reduce identity theft?. *Journal of Policy Analysis and Management*, 30(2):256--286, 2011.