Research Paper (WEIS 2012)

Is Patient Data Better Protected in Competitive Healthcare Markets?

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

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

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

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 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.

Key words: healthcare IT, patient data privacy, data breaches, healthcare competition, information security, IT investments

History:

1. Introduction

1.1. The Broad Issue

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 in the 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 are not aware of any studies that assess the impact of market competition on information security and privacy. While existing work on quality and competition provides some guidance, economic theory is indeterminate on the effect of competition on the likelihood or extent of data breaches at firms. It is plausible that firms with market power may consume their profits and underinvest in activities that thwart data breaches. On the other hand, it is also plausible that firms facing tough competition may cut costs on activities that thwart data breaches, as these activities are less visible to consumer. As a consequence, the impact of competition on information security is an empirical question.

We consider information security and privacy an aspect of firms' overall quality. 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.

1.2. Context

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 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 problem of patient data security is not limited by geography or the nature of

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

medical system (multiple private-public providers vs. single national provider) as evident from the recent breach in United Kingdom (UK). The East Surrey Hospital, run by the National Health Service (NHS) trust, recently lost a memory stick containing unencrypted confidential records for 800 patients. This led the United Kingdom Information Commissioner's Office to warn the hospital that it will face formal regulatory action if any data breach were to happen again(BBC News 2011).

The privacy of health information is a widely accepted notion in the United States (and many Western countries). The Privacy Rule of the federal Health Insurance Portability and Accountability Act (HIPAA) explicitly protects health information such as medical records, billing information, and other medical information about individuals. HIPAA Security Rule explicitly protects individuals' *electronic* health information[2]. Data breaches occur when personally indentificable information (PII)[3] is disclosed - either inadvertently or through a malicious act. As of August 2011, legistlation in 46 US states requires hospitals to report data breaches to the individuals affected by such breaches.

1.3. Overview of the Study and Data Set

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) to define hospital markets and as the primary unit of our analysis. We use the number of data breaches incident reports as a measure of quality and HHI as a measure of competition within each CBSA. We also control for market size, population, population over

[2] `http://www.hhs.gov/ocr/privacy/hipaa/understanding/consumers/index.html` accessed on August 26, 2011. Also `http://www.hhs.gov/ocr/privacy/hipaa/administrative/securityrule/index.html` which describes the Security Rule as "The HIPAA Security Rule establishes national standards to protect individuals electronic personal health information that is created, received, used, or maintained by a covered entity. The Security Rule requires appropriate administrative, physical and technical safeguards to ensure the confidentiality, integrity, and security of electronic protected health information".

[3] We define PII as information about individuals that is generally considered private and has the potential of misuse if disclosed. Office of Budget and Management (OMB) defines PII as any information about an individual maintained by an Agency, including but not limited to, education, financial transactions, medical history, and criminal or employment history and information which can be used to distinguish or trace an individuals identity, such as their name, social security numbers, date and place of birth, mothers maiden name, biometric records, etc., including any other personal information which is linked or linkable to an individual. `http://www.ornl.gov/doe/doe_oro_dmg/TMR/TMRs/DOE_CIO_TMR-22_-_SUI_101607_Final.pdf` accessed on August 26, 2011.

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.

1.4. Summary of Results

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, where we model the odds of a data breach incident, the severity of incident, and the number of breaches. 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.

2. Related Literature

Our work is related to the following streams of research:

2.1. Economics Literature on Competition and Quality in Healthcare Markets

Gaynor (2006) surveys competition and quality specifically in the health care markets. 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" (Chintagunta et al. 2006).

Finally, an issue germane to our article is how hospitals allocate resources between various activities necessary to provide hospital product and services. Mukamel et al. (2002) finds that in the face of price competition, hospitals may allocate resources to activities that are more easily evaluated by consumers.

2.2. Management Literature on Electronic Data Security within Healthcare

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 lead 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 and Tucker (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.

2.3. Contribution of Our Study

As mentioned earlier, theory does not provide a clear-cut answer on how competition impacts 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.

3. Data Sources and Variable Construction

3.1. Data Sources and Data Summary Statistics

We sourced hospital data breach information from Privacy Rights Clearing House[4], 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.

3.2. Data Breach Incidents at Hospitals

For data breaches, we used Privacy Rights Clearing House as the source of data. Our starting dataset includes all incidents reported from January 1, 2006 up to August 22, 2011 across all industries, but the focus of our study is data breaches at hospitals. First, we only retained data

[4] `https://www.privacyrights.org/about_us.htm`

Table 1 Summary Statistics at CBSA-level

Variable	Mean	(Std. Dev.)	Min.	Max.
Data Breach Incidents	0.517	(1.385)	0	15
Herfindahl-Hirschman Index	4022.472	(2477.683)	307.371	10000
Market Size (Hospital Beds)	2111.536	(3633.667)	99	44120.333
Population	653217.139	(1154605.876)	55357	11553017.833
Population (eligible for medicare)	92274.541	(144482.482)	4425	1398709
Population ($>$ 65 years)	76341.018	(126618.528)	3103	1302537
Per Capita Income	30495.358	(6057.12)	15748.333	64219.333
Disclosure Law Effective Days	1732.969	(719.009)	0.001	2974
N		381		

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. Before we describe our variables, it may be worthwhile to briefly discuss as to how data breaches happen and how data breaches get reported.

3.2.1. Data Breach The causes of data breaches can be broadly divided into three categories viz. malicious attacks, unauthorized disclosures, and lost or misplaced protected health information.

Malicious attacks take place when individuals access or steal data with malicious intent. These malicious accesses may be through remote hacking into computer systems, physical stealing of computers and storage devices, or insider access of insufficiently protected computer systems.

Unauthorized disclosures can happen when private patient data is made available to individuals who are not authorized to access this private data. Usually there is no malicious intent. Examples of unauthorized disclosures includes:

- Wrong person, lab, physicial (electronic and physical)
- Family members (unauthorized)
- Publicly accessible computer records
- Insider access due to insufficient access controls

Lost or misplaced protected health information incidents can happen when laptops or portable storage devices such as thumb drives are lost or misplaced.

Table 2 provides a frequency distribution of the various types of data breach incidents in our data set.

Table 2 Frequency Distribution of Type of Data Breach

	Freq.	Percent	Cum.
Hacking or malware	9	4.46	4.46
Insider	46	22.77	27.23
Physical loss	28	13.86	41.09
Portable device	78	38.61	79.70
Stationary device	19	9.41	89.11
Unintended disclosure	20	9.90	99.01
Unknown or other	2	0.99	100.00
Total	202	100.00	

3.2.2. Data Breach Reports We only know about data breach incidents that are reported and there is no way to exactly know how many data breach incidents actually happened - some breaches may never get reported and even some may never get discovered. However, there are a few main reasons why hospital data breaches are reported:

1. Patients resides in a geographical state with mandatory data breach disclosure laws. Hospitals are forced to disclose data breach to avoid breaking the law although the hospital may be dismissive of the impact of the data breach.

2. Disgruntled (ex-) employees may report a data breaches to settle scores with the hospital.

3. Consumers, press, or privacy organization may report a data breach discovered by chance (e.g. public posting of private patient information)

4. Consumers and law enforcement may report data breach discovered as the cause of an indentity theft

3.2.3. Data Breach Variables 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 is an ordinal variable coded into three categories as enumerated in Table 3, which also provides a frequency distribution of the levels of severity. Please note that *Severity* is a CBSA-level measure so a frequency of 281 for "No Breach" means that a total of 281 CBSA had no data breaches[5].

Table 3 Frequency Distribution of Severity of Data Breach in a CBSA

		Freq.	Percent	Cum.
1	No breach	281	79.38	79.38
2	Breach without disclosure of financial data	11	3.11	82.49
3	Breach with disclosure of financial data	62	17.51	100.00
Total		354	100.00	

Number of Records The number of records that were breached within a CBSA. This is a count variable. Table 4 provide summary statistics on total records breached and Table 5 provide summary statistics on the number of records breached, given a breach occured.

Table 4 Summary Statistics for Records Breached and Severity

Variable	Mean	(Std. Dev.)	Min.	Max.
Total Records Breached	28748.741	(205504.05)	0	2204800
Severity	1.38	(0.766)	1	3
N		355		

Table 5 Summary Statistics for Records Breached, Given Breach

Variable	Mean	(Std. Dev.)	Min.	Max.
Total Records Breached	139805.521	(438061.927)	13	2204800
Log(Records Breached)	8.756	(2.645)	2.565	14.606
N		73		

We note that the standard deviation is much larger than the mean number of records breached so it may be helpful to use a logarithm of number of records (given breach) while estimating a model.

[5]The total CBSA reported here is lower because a few CBSA were dropped due to missing data.

3.3. Competition and Market Size

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. Hospital market size can be calculated using the total number of hospital beds, in-patient days, or similar measures (e.g. AHA reports adjusted patients days by adjusting in-patient days with the ratio of out-patient revenue and in-patient revenue). Table 6 shows that total beds, in-patient days, and adjusted patient days are significantly and positively correlated (so *HHI* based on either of these measure do not differ much). We thus 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.

Table 6 Correlation Between Beds, Inpatient Days, and Adjusted Patient Days (with significance)

Variables	Total Beds	Total Inpatient Days	Adjusted Patient Days
Total Beds	1.000		
Total Inpatient Days	0.974	1.000	
	(0.000)		
Adjusted Patient Days	0.941	0.946	1.000
	(0.000)	(0.000)	

3.3.1. Remarks About HHI Calculation It is worthwhile to make a few remarks about the calculation of HHI and its distribution:

1. We counted all hospitals belonging to one hospital system within a market as a single entity. As a hypothetical example, if there are two hospitals in a CBSA owned by the same hospital system, the market is considered a monopoly with an HHI of 10,000.

2. Since AHA provides yearly data, we first calculated HHI for each CBSA-year and then averaged over the years to obtain HHI for the CBSA.

3. As of August 2010, US Federal Trade Commission (FTC) threshold for “highly concentrated”

markets is an HHI of over 2500[6]. 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.

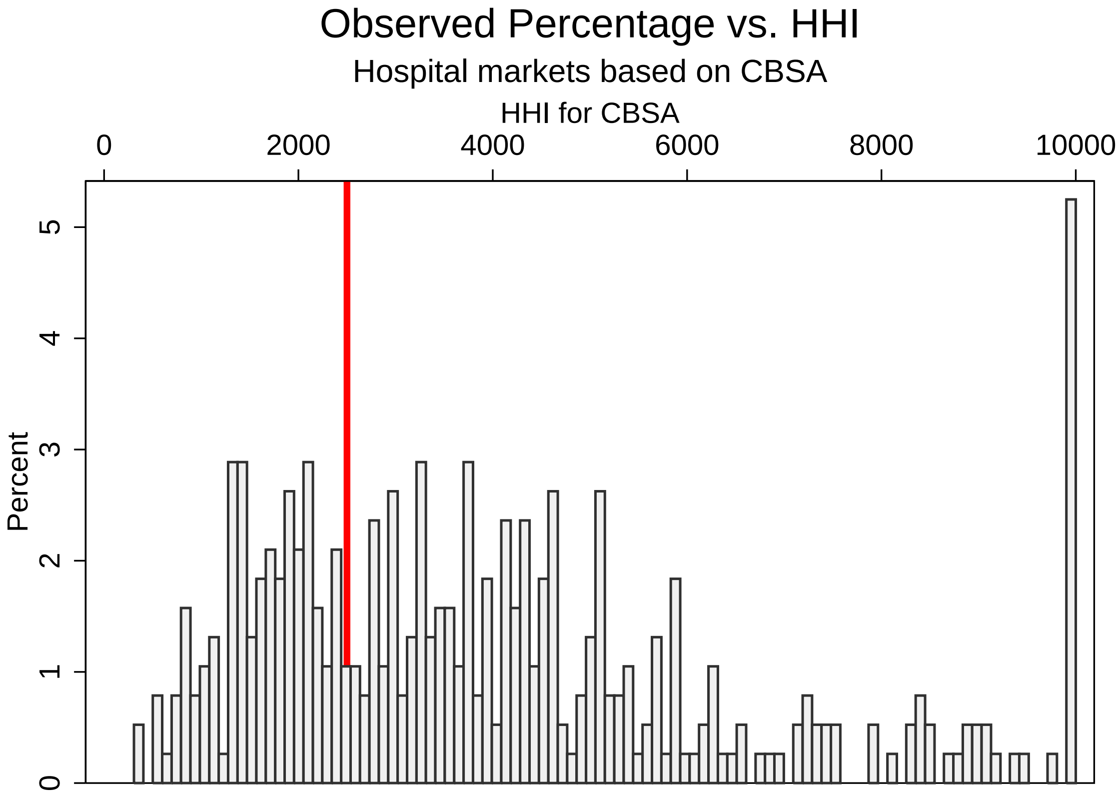

Figure 1

3.4. CBSA Demographics

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.

Population The total population residing in a CBSA, calculated as a sum of the population in underlying counties.

Population Over 65 Total population over 65 residing in a CBSA

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

Per Capita Income The weighted average of per capital income at the CBSA level

3.5. Hospital Data

General attributes of a hospital such as ownership structure and system membership may affect the competitive intensity as well as the quality of data security. We sourced data from AHA yearly survey to construct such general hospital-level variables.

Owner Type Whether the hospital is government, non-profit, or for-profit.

System Member Whether the hospital is part of a larger system

Table 8a and 9a provide descriptive statistics on hospital ownership and system membership respectively. We note that most hospitals are not-for-profit and most hospitals belong to a system. Finally, Table 7a provides a tabulation of incidents at the hospitals.

Since a hospital's Information Technology choices will potentially affect data breaches, 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 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 Desk Full-time equivalent employees working to support Electronic Medical Records

3.5.1. Effect on Sample Size of Merging Hospital Data from AHA and HIMSS The AHA survey covers more hospitals than HIMSS, so it is natural that we would lose observations if we attempt to merge the two data sets. We merge AHA and HIMSS data on Medicare number, which results in the loss of hundreds of observations. Finally, there are missing entries in HIMSS data for Information Systems department full-time employee (FTE) counts, which results in further

Table 7 Frequency Distribution of Incidents at Hospital-Level

(a) AHA Data Only

		Freq.	Percent	Cum.
0	No Incident	3873	95.87	95.87
1	One or More Incidents	167	4.13	100.00
Total		4040	100.00	

(b) After AHA & HIMSS data merge

		Freq.	Percent	Cum.
0	No Incident	2508	94.29	94.29
1	One or More Incidents	152	5.71	100.00
Total		2660	100.00	

Table 8 Frequency Distribution of Hospital Ownership

(a) AHA Data Only

		Freq.	Percent	Cum.
1	Government, non-federal	597	14.78	14.78
2	Government, federal	162	4.01	18.79
3	Not-for-profit	2029	50.22	69.01
4	For-profit	1252	30.99	100.00
Total		4040	100.00	

(b) After AHA & HIMSS data merge

		Freq.	Percent	Cum.
1	Government, non-federal	349	13.12	13.12
3	Not-for-profit	1705	64.10	77.22
4	For-profit	606	22.78	100.00
Total		2660	100.00	

Table 9 Frequency Distribution of Hospital System Membership

(a) AHA Data Only

		Freq.	Percent	Cum.
0	Not in System	1424	35.25	35.25
1	In a System	2616	64.75	100.00
Total		4040	100.00	

(b) After AHA & HIMSS data merge

		Freq.	Percent	Cum.
0	Not in System	902	33.91	33.91
1	In a System	1758	66.09	100.00
Total		2660	100.00	

loss of observations in the models that follow. Tables 7b, 8b, and 9b tabulate hospital-level variables after the merge with HIMSS data.

3.6. Other Variables

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.

4. Modeling Framework and Empirical Analysis

4.1. Motivating the Model

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 about the impact of competition on quality as summarized in Subsection A (in the Appendix).

The common framework used in empirical work in this area is the structure-conduct-performance (SCP) framework but with a focus on market structure and firm conduct. The usual setup for the SCP framework is as follows:

- Herfindahl-Hirschman Index (HHI) is used as a measure of competition
- Firm's choice of price or quality-level is used as a measure of firm conduct
- Demand-shifters and cost-shifters are included as controls (see Gaynor 2006, pp. 11, 14-15).

The econometric specification usually has the following form:

$$Quality = \beta_0 + \beta_1(Demand\ Shifters) + \beta_2(Cost\ Shifters) + \beta_3(HHI) + \epsilon \tag{1}$$

We employ a similar specification in our study. A common concern is the endogeneity of HHI and some authors have used instruments to overcome such concerns. For example, a number of authors have used patient choice models determined by distance from hospital to overcome endogeneity concerns. We do not have access to patient-level data and hence it was not feasible for us to employ patient-choice models in this work.

4.2. Model Estimators

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)[7]. 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 (see Wooldridge 2002, chap. 19). 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:

$$DBIR_i \sim Poisson(\lambda_i) \tag{2}$$

$$\lambda_i = \exp[\beta_0 + \beta_1 HHI_i + \beta_2 (Market\ Size)_i + \beta_3 Population_i + \beta_4 Income_i] \tag{3}$$

[7]NBRM is equivalent to overdispersed PRM in our setting because of the mean zero assumption on the overdispersion parameter.

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 (see Long 1997, pp. 231-232).

$$\lambda_i = \exp[X\beta + \ln\delta] \tag{4}$$

$$\delta_i \sim Gamma(\alpha_i, \alpha_i) \tag{5}$$

δ may be viewed as the combined effect of omitted variables (Gourieroux et al. 1984) or a source of randomness (Hausman et al. 1984). A likelihood ratio test for the hypothesis $H_0 : \alpha = 0$ provides a test for overdispersion(see Long 1997, p. 237)..

4.3. Main Results

For our primary results, we use variants of models described in Equations (2), (3), (4), and (5) (see page 14). The models are labeled[8] PRM1, PRM2, PRM3a, PRM3b, PRM4, PRM5, NBRM6 in Table 10. 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* as discussed in Subsection 4.5 (on page 18).

We estimated the models using Stata with observed information matrix (OIM) variance-covariance estimator[9], and the results in Table 10 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 correlated with a decrease in average count of data breaches.

[8]The prefix PRM stands for "Poisson Regression Model", whereas the prefix NBRM stands for "Negative-binomial Regression Model".

[9]Standard errors and coefficient covariance matrix are estimated using observed information matrix, which is the inverse of the negative Hessian matrix (Gould et al. 2006, p. 247)

Table 10 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	5.18e-05	1.65e-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	8.13e-08	-5.41e-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))				-6.25e-07				
				(1.77e-06)				
				0.72				
Population (Medicare eligible)					1.48e-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

4.4. Discussion

Table 11 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 $H_0 : \alpha = 0$ (see Section 4.2) reports a chi-square statistic $\chi^2 = 18.22$ with $P[\chi^2 \geq 18.22|H_0] = 0.000$ providing evidence for overdispersion[10].

Table 11 Detailed Results of NBRM

	NBRM7			
	b	se	p	factor
incidents				
HHI (by 100)	-5.19e-02***	(9.00e-03)	0.00	0.94940
Size (by 100)	2.19e-03	(6.69e-03)	0.74	1.00219
Population (by 1000)	1.85e-04	(2.18e-04)	0.40	
Population (over 65 years, by 1000))	-4.06e-05	(2.42e-03)	0.99	
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			

Column "factor" in Table 11 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.

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

4.4.1. Support for Resource Shifting Explanation from Earlier Work Economic theory can explain the direction of impact when firms facing competition choose between two or more strategic variables (e.g. price and quality). 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 if consumers are better informed about one strategic variable over another - price and quality in their case - then the firm may provide sub-optimal quality in equilibrium (see Gaynor 2006, pp. 6-8 for details).

As providing relief from illness is the primary *product* of the hospital, most of the attention of the academic community has been on quality indicators related to health outcomes. However, activities related to various aspects of the overall quality (healthcare, hotel services, patient data protection, and so on) compete for allocation from the finite hospital budget. Depending on relative elasticities of price and various aspects of quality, hospitals may shift resources between activities related to different aspects of quality.

The insights from Dorfman-Steiner condition and Dranove-Satterthwaite (as mentioned earlier) support these explanations. 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.

4.5. Robustness

As mentioned earlier, mean estimates from Poisson MLE are robust to distributional assumptions. Table 12 provide further evidence of functional form robustness of our results. In model PRM8 and model NBRM9, we include a quadratic control variable by adding $(\textit{Market Size})^2$. 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 12 to conserve space. Finally, OLS11 estimates a log-linear model using Ordinary Least Squares, with $Log(Incidents)$ as the dependent variable.

Table 12 DV=Log(Incidents) for OLS, DV=Incidents for Other Models

	PRM8	NBRM9	PRM10	OLS11
	b/se/p	b/se/p	b/se/p	b/se/p
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)
	0.00	0.00	0.00	0.03
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)
	0.00	0.01	0.00	0.01
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)
	0.00	0.00	0.00	0.00
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)
	0.46	0.89	0.03	0.05
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)
	0.02	0.12	0.68	0.61
Log(Law Effective Days)	-3.56e-02	-3.48e-02		-3.65e-02
	(2.26e-02)	(2.60e-02)		(3.94e-02)
	0.11	0.18		0.36
vce	oim	oim	oim	ols
N	380	380	380	380

Looking at Table 12, 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. Table 13 summarizes the evidence from the non-linear models with unscaled independent variables. The first row reports the factor by which average count of data breaches should be multiplied for each unit increase in HHI. The second row reports the p-value for the estimates in the first row. To reiterate, all results in Table 13 are statistically significant and suggest that an increase in HHI is correlated with a decrease in average count of data breaches.

Table 13 Factor Change in Expected Count of Incidents

	PRM1 factor/p	PRM2 factor/p	PRM3 factor/p	PRM4 factor/p	PRM5 factor/p	NBRM6 factor/p	PRM8 factor/p	NBRM9 factor/p	PRM10 factor/p
incidents									
HHI	0.9991	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9996	0.9995
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

5. 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.

5.1. Association between Odds of an Incident and HHI using Simple Logit Regression Model

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 14 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 14, we also control for hospital ownership type, hospital system membership, and hospital JCAHO[11] 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.

5.2. Association between Odds of an Incident and HHI using Multi-level Model

Our data set has an inherent multilevel or hierarchical structure - hospitals are embedded within CBSA; CBSA are embedded within geographical states. Our outcome variable - incident of data

[11] "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 organizations commitment to meeting certain performance standards" (Source: `http://www.jointcommission.org/about_us/about_the_joint_commission_main.aspx`, accessed on Jan 03, 2012).

Table 14 Incident Logit (Hospital)

	L1	L2	L3	L4	L5	L6
	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p	b/se/p
incident						
HHI (by 100)	-1.82e-02	-7.36e-03	-1.27e-02	-1.15e-02	-1.05e-02	-1.64e-02
	(6.83e-03)	(7.17e-03)	(7.62e-03)	(7.60e-03)	(1.11e-02)	(1.11e-02)
	0.01	0.30	0.10	0.13	0.34	0.14
CBSA Size (by 100)		-2.83e-03	-6.81e-04	-4.45e-03	-1.02e-02	-1.27e-02
		(4.83e-03)	(4.68e-03)	(4.60e-03)	(8.74e-03)	(1.01e-02)
		0.56	0.88	0.33	0.24	0.21
Population (by 1000)		1.97e-04	2.77e-04	2.06e-04	5.60e-05	-1.83e-04
		(1.14e-04)	(1.10e-04)	(9.76e-05)	(2.16e-04)	(2.41e-04)
		0.08	0.01	0.04	0.80	0.45
Population (>65, by 1000)		-1.26e-03	-2.50e-03	-6.72e-04	2.80e-03	5.30e-03
		(1.18e-03)	(1.54e-03)	(1.44e-03)	(2.91e-03)	(3.18e-03)
		0.29	0.11	0.64	0.34	0.10
Income (by 1000)		2.25e-02	1.11e-02	5.34e-03	7.34e-03	-2.83e-02
		(1.23e-02)	(1.25e-02)	(1.21e-02)	(1.75e-02)	(3.17e-02)
		0.07	0.38	0.66	0.68	0.37
System Size (in CBSA)		5.74e-04	1.31e-04	6.81e-05	2.13e-04	4.80e-04
		(1.34e-04)	(1.57e-04)	(1.74e-04)	(1.93e-04)	(1.88e-04)
		0.00	0.40	0.70	0.27	0.01
Hospital Size (by 100)			2.99e-01	2.43e-01	2.46e-01	2.62e-01
			(3.76e-02)	(4.87e-02)	(7.12e-02)	(1.23e-01)
			0.00	0.00	0.00	0.03
Number of FTE				1.55e-04	3.78e-05	-6.80e-05
				(5.73e-05)	(6.67e-05)	(1.87e-04)
				0.01	0.57	0.72
FTE in IS					3.20e-03	8.49e-03
					(1.37e-03)	(3.62e-03)
					0.02	0.02
FTE in IS Security						8.63e-02
						(7.61e-02)
						0.26
FTE in EMR Support						-2.45e-02
						(9.89e-03)
						0.01
Intercept	-2.79e+00	-4.08e+00	-5.64e+00	-4.99e+00	-3.78e+00	-2.67e+00
	(1.61e-01)	(4.70e-01)	(7.48e-01)	(8.16e-01)	(7.46e-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

breach - is measured at the hospital-level. Our focal predictor - HHI - is measured at the CBSA-level, where as control variables are measured at hospital-level, CBSA-level, and geogrpahical state level. Multilevel modeling allows us to take into account the hierarchy in our data.

The non-hierarchical hospital-level model estimates (as in subsection 5.1) suffer from multiple problems. First, the estimate for the coefficient of HHI are not statistically significant when the estimated model includes any control variables - only model L1 in Table 14 has statistically significant coefficient. Second, we cannot include CBSA-level indicators as CBSA-indicators and any of the CBSA-level variables will not be linearly independent. Thus, model estimation requires a strong assumption that the intercept term in these models do not vary at the CBSA-level and the included variables capture all variation attributable to CBSAs. A multi-level model addresses the later issue, by allowing us to model the CBSA and geographical-state-level heterogeneity as described below.

Let us assume that subscripts i, j, k represent the hospital, CBSA, and geographical states in the following equations. We can then write a three-level model as:

$$\Pr(\mathit{Incident}_i = 1) = \operatorname{logit}^{-1}[\alpha_{j[i]} + \beta_1 \mathit{HHI} + \boldsymbol{\beta_2}\,\textit{(Hospital-level Preditors)}] \tag{6}$$

$$\alpha_j = \gamma_{k[j]} + \boldsymbol{\beta_3}\,\textit{(CBSA-level Predictors)} + \zeta_j \tag{7}$$

$$\gamma_k = \mu + \eta_k \tag{8}$$

The error terms ζ_j and η_k are assumed to be standard normal random variables. Equation 6 is the main model that estimates the association between *HHI* and the quality of data protection practice. In addition to the focal predictor *HHI*, we also include hospital-level predictors and an intercept term that varies over CBSA. Equation 7 then models the CBSA-level intercept as a normal random variable with mean determined by the CBSA-level predictors and an intercept that varies over geographical states. Thus, we are better able to control for the differences across CBSA and geographical states beyond those captured by the included predictors.

For estimation of these models, we used `lme4`, a multi-level modeling package for the statistical programming language `R`. We only present the estimates on the first-level i.e. hospital-level and suppress the estimates at the CBSA-level and geographical-state level as the latter are not necessary for interpretation. Table 15 provides a summary of estimates for the variables included in the hospital-level model. Since we do not include estimates for the CBSA-level and state-level models, we include the following equations to clarify the functional form at those levels:

$$\alpha_j = \gamma_{k[j]} + \beta_3(CBSA\ Size) + \beta_3(Average\ Income) + \zeta_j \tag{9}$$

$$\gamma_k = \mu + \eta_k \tag{10}$$

The coefficient estimate for *HHI* is statistically significant and implies that a 100-point increase in HHI is associated with a 1.5% decrease in the odds of a data breach incident. The association suggested here is again consistent with our main result.

Table 15 **Three-level model with varing intercept (second-level predictors)**

HHI (by 100)	-0.015^{*}
	(0.006)
Hospital Size (by 100)	0.309^{***}
	(0.028)
Hospital in System	0.270
	(0.194)
JCAHO Accreditation	1.669^{**}
	(0.604)
Owned by Fed Gov	−0.600
	(0.481)
Not-for-Profit	0.035
	(0.231)
For-profit	-1.632^{***}
	(0.412)
(Intercept)	-5.320^{***}
	(0.633)
N	4033

6. 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

Dependent Variables Count of incidents, or odd of an incident

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) from less visible activities such as data security. In doing say, the hospitals increase the risk of data breaches. We find support for this explanation both from economic theory (Dorfman-Steiner condition and Dranove et al. (1992)) and from empirical research (Mukamel et al. 2002).

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 than have not been correctly anticipated.

Appendix

A. Empirical Studies on Competition and Healthcare Quality: Hospitals Set Prices and Quality

A.1. Competition Increases Quality

Joskow (1980) Excess beds vs. HHI

Robinson and Luft (1985) Length of stay vs. HHI

Dranove, Shanley, and Simon (1992) High-technology services vs. HHI

Gowrisankaran and Town (2003) Mortality vs. HHI

Sohn and Rathouz (2003) Mortality vs. Competition coefficient

Sari (2002) Quality indicators vs. HHI

Abraham, Gaynor, and Vogt (2005) Quality consumed vs. Number of hospitals

A.2. Competition Decreases Quality

Mukamel, Zwanziger, and Bamezai (2002) Mortality vs. HHI

Volpp, Williams, Waldfogel, Silber, Schwartz, and Pauly (2003) Mortality vs. Number of competitors

Encinosa, Bernard, Steiner, and Chen (2005) Patient safety vs. Hospital margin

Propper, Burgess, and Green (2004) Mortality vs. Number of competitors

A.3. Competition Decreases Quality

Ho and Hamilton (2000) Mortality vs. Merger

Capps (2005) Patient safety indicators vs. Merger

B. Association between Incident Count and HHI without Hacks or Malware

Table 16 is included for comparison with Table 11. As is clear, the direction of the result is unchanged and the magnitude changes only very slightly.

Table 16 Detailed Results of NBRM (without Hacks or Malware)

	NBRM7			
	b	se	p	factor
incidents				
HHI (by 100)	-5.12e-02***	(9.30e-03)	0.00	0.95012
Size (by 100)	1.51e-03	(6.41e-03)	0.81	1.00151
Population (by 1000)	1.58e-04	(2.27e-04)	0.49	
Population (over 65 years, by 1000))	5.68e-04	(2.43e-03)	0.82	
Per Capita Income	2.76e-05	(1.55e-05)	0.07	
Log(Law Effective Days)	-4.05e-02	(2.63e-02)	0.12	
vce	oim			
N	381			

C. Theoretical Model of Quality Choice

In this section, we show that a competitive firm may chose lower-quality level for data security versus a monopoly firm under some reasonable assumption. The model will be inserted here in future.

References

Abraham, J.M., M.S. Gaynor, W.B. Vogt. 2005. Entry and competition in local hospital markets. Tech. rep., National Bureau of Economic Research.

BBC News. 2011. Patient data lost from hospital. *BBC* URL `http://www.bbc.co.uk/news/uk-england-surrey-15176343`.

Capps, C.S. 2005. *The quality effects of hospital mergers*, vol. 5. US Dept. of Justice, Antitrust Division.

Chintagunta, Pradeep, Tlin Erdem, Peter E. Rossi, Michel Wedel. 2006. Structural modeling in marketing: Review and assessment. *Marketing Science* **25**(6) 604–616. doi:10.1287/mksc.1050.0161. URL `http://mktsci.journal.informs.org/content/25/6/604.abstract`.

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

Encinosa, W.E., D.M. Bernard, C.A. Steiner, C.C. Chen. 2005. Use and costs of bariatric surgery and prescription weight-loss medications. *Health Affairs* **24**(4) 10391046.

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

Gould, W., J. Pitblado, W. Sribney. 2006. *Maximum likelihood estimation with Stata*. Statacorp Lp.

Gourieroux, C., A. Monfort, A. Trognon. 1984. Pseudo maximum likelihood methods: applications to poisson models. *Econometrica: Journal of the Econometric Society* 701–720.

Gowrisankaran, G., R. J Town. 2003. Competition, payers, and hospital quality. *Health Services Research* **38**(6p1) 1403–1422.

Hausman, J.A., B.H. Hall, Z. Griliches. 1984. *Econometric models for count data with an application to the patents-R&D relationship*. National Bureau of Economic Research Cambridge, Mass., USA.

Ho, V., B.H. Hamilton. 2000. Hospital mergers and acquisitions: does market consolidation harm patients? *Journal of Health Economics* **19**(5) 767791.

Joskow, P.L. 1980. The effects of competition and regulation on hospital bed supply and the reservation quality of the hospital. *The Bell Journal of Economics* 421447.

Long, J.S. 1997. *Regression models for categorical and limited dependent variables*, vol. 7. Sage Publications, Inc.

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

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

Propper, C., S. Burgess, K. Green. 2004. Does competition between hospitals improve the quality of care?:: Hospital death rates and the NHS internal market. *Journal of Public Economics* **88**(7-8) 12471272.

Robinson, J.C., H.S. Luft. 1985. The impact of hospital market structure on patient volume, average length of stay, and the cost of care. *Journal of Health Economics* **4**(4) 333356.

Romanosky, S., R. Sharp, A. Acquisti. 2010. Data breaches and identity theft: When is mandatory disclosure optimal? *Ninth Workshop on the Economics of Information Security (WEIS 2010), Harvard University, Cambridge, MA*.

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

Sack, Kevin. 2011. Stanford hospital patient data breach is detailed. *The New York Times* URL `http://www.nytimes.com/2011/10/06/us/stanford-hospital-patient-data-breach-is-detailed.html`.

Sari, N. 2002. Do competition and managed care improve quality? *Health economics* **11**(7) 571584.

Sohn, M.W., P.J. Rathouz. 2003. Competition among hospitals and quality of care: Hospital-level analysis. *unpublished paper, University of Chicago* .

Volpp, K.G.M., S.V. Williams, J. Waldfogel, J.H. Silber, J.S. Schwartz, M.V. Pauly. 2003. Market reform in new jersey and the effect on mortality from acute myocardial infarction. *Health Services Research* **38**(2) 515533.

Wooldridge, J.M. 2002. *Econometric analysis of cross section and panel data*. The MIT press.