JOURNAL OF
EMPIRICAL LEGAL STUDIES

Journal of Empirical Legal Studies
Volume 11, Issue 1, 74–104, March 2014

Empirical Analysis of Data Breach Litigation

*Sasha Romanosky, David Hoffman, and Alessandro Acquisti**

In recent years, many lawsuits have been filed by individuals seeking legal redress for harms caused by the loss or theft of their personal information. However, very little is known about the drivers, mechanics, and outcomes of those lawsuits, making it difficult to assess the effectiveness of litigation at balancing organizations' usage of personal data with individual privacy rights. Using a unique and manually collected database, we analyze court dockets for more than 230 federal data breach lawsuits from 2000 to 2010. We investigate two questions: Which data breaches are being litigated? and Which data breach lawsuits are settling? Our results suggest that the odds of a firm being sued are 3.5 times greater when individuals suffer financial harm, but 6 times lower when the firm provides free credit monitoring. Moreover, defendants settle 30 percent more often when plaintiffs allege financial loss, or when faced with a certified class action suit. By providing the first comprehensive empirical analysis of data breach litigation, our findings offer insight into the debate over privacy litigation versus privacy regulation.

I. Introduction

The surge in popularity of social media, e-commerce, and mobile services is proof of the benefits consumers are enjoying from information and communication technologies. However, these same technologies can create harm when personal consumer information is lost or stolen, causing emotional distress or monetary damage from fraud and identity theft.[1] Since 2005, an estimated 543 million records have been lost from over 2,800 data

*Address correspondence to Sasha Romanosky, New York University School of Law, 406 Wilf Hall, 139 MacDougal St., New York, NY 10012; email: sromanos@cmu.edu. Romanosky is a Microsoft Research Fellow at the Information Law Institute, New York University School of Law; Hoffman is the James E. Beasley Professor of Law at Temple University Beasley School of Law; Acquisti is an Associate Professor of Information Technology and Public Policy at the Heinz College of Carnegie Mellon University.

This research was supported by CyLab at Carnegie Mellon under Grants DAAD19-02-1-0389 and W911NF-09-1-0273 from the Army Research Office, by Temple Law Schools Conwell Corps Program, and by the Information Law Institute at New York University School of Law. We thank Antima Chakraborty, Carol Anne Donohoe, Ian Everhart, Caitlin Jones, Kevin Leary, and Jake Oresick for their research assistance. We also thank Paul Bond, Aaron Burnstein, Jim Graves, Fainna Kagan, Amelia Haviland, Mark Melodia, Kristen Matthews, Peter Oh, Barrie Nault, David Navetta, Mohammad Rahman, Theresa Romanosky, Boris Segalis, Brendon Tavelli, seven anonymous attorneys, and the anonymous reviewers and editors of *JELS* for their valuable insights and suggestions.

[1]See Solove (2010) for a description of the potential harms associated with breaches of personal information.

breaches,[2] and identity theft caused $13.3 billion in consumer financial loss in 2010 (BJS 2011). In response, federal legislators have introduced numerous bills that define appropriate business practices regarding the collection and protection of consumer information,[3] and federal regulators have drafted privacy frameworks for consumer data protection (Department of Commerce 2010; FTC 2010). A significant concern for policymakers, therefore, is balancing ex ante regulation with ex post litigation to protect both consumer and commercial interests. For instance, the Department of Commerce inquired: "should baseline commercial data privacy legislation include a private right of action?" (Department of Commerce 2010:30). At issue is the degree to which the current liability regime sufficiently addresses modern privacy harms, or whether a new, more effective federal liability standard is required.

On one hand, a weak litigation regime would be ineffective at deterring a firm's harmful or negligent behavior. Lawsuits that are inappropriately disposed of eliminate a plaintiff's ability to obtain appropriate relief for legitimate harms. For example, a case was successfully brought against Rite Aide for carelessly tossing pharmacy labels and employment applications in a public trash dumpster.[4] In the settlement, Ride Aide agreed to "a comprehensive information security program that is reasonably designed to protect the security, confidentiality, and integrity of personal information collected from or about consumers."[5] Without legal action, such careless practices may have never been corrected.

On the other hand, a heavy-handed litigation regime could impose excessive legal fees and damage awards and—according to some—stifle innovation. For instance, Netflix, an online movie rental site, offered a $1 million prize to anyone who could sufficiently improve its movie recommendation algorithm. To facilitate the contest, Netflix published (what was believed to be) anonymized rental information for a sample of its users. Due to lawsuits stemming from the reidentification of these data, Netflix cancelled a subsequent contest. While the total social value of such innovation may be limited, the Netflix case provides one example of how litigation can impact firms' product development.

Our research attempts to offer novel insight into this debate by providing the first comprehensive empirical analysis of data breach litigation, and investigates the drivers, mechanisms, and outcomes of data breach litigation.

Determining whether current U.S. privacy laws are too weak or too strong is not easy. It is difficult (and perhaps impossible) to assess the aggregate costs and benefits for both consumers and firms of different privacy regimes in purely monetary terms (Romanosky &

---

[2]See Privacy Rights Clearinghouse, http://www.privacyrights.org/data-breach, last accessed Jan 22, 2012.

[3]For example, the Cyber Security and American Cyber Competitiveness Act of 2011 (S.21), the Data Security and Breach Notification Act of 2011 (S.1207), the Commercial Privacy Bill of Rights Act 2011 (S.799), the Personal Data Privacy and Security Act of 2011 (S.1151), the Data Breach Notification Act (S.1408), the Personal Data Protection and Breach Accountability Act of 2011 (S.1535), the Secure and Fortify Electronic Data Act of 2011 (H.R.2577), and the Cybersecurity Enhancement Act of 2011 (H.R. 2096).

[4]See In re Rite Aid Corp., FTC File No. 072-3121 (July 27, 2010).

[5]Id.

Acquisti 2009). However, even just understanding the landscape of privacy litigation is a problem. Little is known about the trends in data breach litigation—which breaches are litigated and which are not, and with what outcomes. While there exists some legal scholarship regarding data breach litigation (Citron 2007, 2011; Rice 2007; Serwin 2009), it typically examines a narrow subset of lawsuits, focusing on high-profile cases or those with published opinions. Unfortunately, given that as few as 15 percent of all federal lawsuits produce reported opinions (Hoffman et al. 2007), any conclusions reached from examining particular, high-profile cases are likely unrepresentative of the full population of data breach lawsuits. Consequently, it remains still unclear what characteristics these lawsuits actually possess, and how "successful" they have been.

To our knowledge, no empirical research involving data breach lawsuits has been conducted. The purpose of this article is to address this research and policy gap by investigating empirically a representative collection of federal data breach lawsuits and their outcomes. We overcome common sample selection issues by searching Westlaw and acquiring data directly from court dockets (PACER), in combination with other publicly available data sources.[6]

In addition to presenting rich descriptive information about these lawsuits, we explore two sets of questions. First, what kinds of data breaches are being litigated in federal court, and why? Second, what kinds of data breach lawsuits are settling, and why? Our first question examines federal lawsuits resulting from reported data breaches, while the second question includes all known federal lawsuits related to the unauthorized disclosure of personal information.

Our analysis reveals that federal data breach lawsuits typically exhibit a number of significant characteristics. First, plaintiffs seek relief for one or more of: actual loss from identity theft (e.g., financial or medical fraud), emotional distress, cost of preventing future losses (e.g., credit monitoring and identity theft insurance), and the increased risk of future harm. Second, the lawsuits are usually private class actions, though some are brought by public entities such as the Federal Trade Commission or state attorneys general. Third, defendants are typically large firms such as banks, medical/insurance entities, retailers, or other private businesses. Fourth, complaints allege a staggering range of both common-law (tort, breach of contract) and statutory causes of action. And fifth, the vast majority of cases either settle, or are dismissed, either as a matter of law, or because the plaintiff was unable to demonstrate actual harm.

In addition, we find that that the odds of a firm being sued are 3.5 times greater when individuals suffered financial harm, but over 6 times lower when the firm provides free credit monitoring to those affected by the breach. Moreover, the odds of a firm being sued as a result of improperly disposing of data are 3 times greater relative to breaches caused by lost/stolen data, and 6 times greater when the data breach involved the loss of financial information. Our analysis suggests that defendants settle 30 percent more often when plaintiffs allege financial loss from a data breach, or when faced with a certified class action suit. The odds of a settlement are found to be 10 times greater when the breach is caused

---

[6]We discuss the consequences of limiting our search to federal lawsuits in Section VI.

by a cyber attack, relative to lost or stolen hardware, and the compromise of medical data increases the probability of settlement by 31 percent.

By providing a comprehensive empirical analysis of data breach litigation, these findings offer insight into the debate over privacy litigation versus privacy regulation. Specifically, we believe that answering these questions will help inform firms, consumers, and policymakers regarding the risks associated with the collection and use of personal information, and the characteristics and outcomes of federal data breach litigation.

The next section provides background literature related to data breaches, docket analysis, and litigation. We then examine which breaches are litigated and, conditional on suit, which cases settle. Discussions of limitations and final conclusions complete the article.

II. Related Work

In recent years, economists have researched empirical and theoretical aspects of data breaches, such as the effect of breaches on a firm's stock market price (Campbell et al. 2003; Cavusoglu et al. 2004; Acquisti et al. 2006; Kannan et al. 2007; Gordon et al. 2011), the effect of data breach disclosure laws on identity theft (Romanosky et al. 2011), and the conditions under which disclosure laws may reduce the social costs of these breaches (Romanosky et al. 2010). That work shows that while disclosure of a breach does appear to reduce identity theft, the evidence of the impact on stock market price is not conclusive. In addition, a growing body of legal scholarship relates to data breaches. For example, policy and legal scholars have discussed the outcomes of data breach litigation (Citron 2007; Hutchins 2008; Lesemann 2012; Solove 2008); they have summarized the legal theories that plaintiffs allege when trying to recover damages from data breaches (Citron 2007, 2011; Rice 2007; Serwin 2009); and they have examined alternative policy mechanisms that can be used to reduce the harm from data breaches (Romanosky & Acquisti 2009).

In addition, there are efforts to construct repositories for domain-specific lawsuits, making them available for public analysis and research. Such efforts include the Securities Class Action Clearinghouse,[7] Intellectual Property Litigation Clearinghouse,[8] Civil Rights Litigation Clearinghouse,[9] and the Equal Employment Opportunity Commission (Kim et al. 2009).

An emerging body of (nonprivacy) legal scholarship analyzes court dockets. This form of empirical research (docketology) makes very practical use of the publicly available—and generally very detailed—collection of pleadings, motions, rulings, and administrative record keeping that compose a legal dispute. For example, Kim et al. (2009) use docket analysis to compare judicial decisions between district and appellate judges. Hoffman et al. (2007) use dockets to examine the incentives for judges to justify their legal

---

[7]See http://securities.stanford.edu/, last accessed Jan. 22, 2013.

[8]See http://www.law.stanford.edu/program/centers/iplc/, last accessed Jan. 22, 2013.

[9]See http://www.clearinghouse.net/, last accessed Jan. 22, 2013.

decisions (i.e., orders vs. opinions) and to publish these decisions in court reporters. Boyd and Hoffman (forthcoming) use dockets to examine federal veil-piercing litigation and examine the characteristics that lead to a plaintiff's greater "success" rate. We draw insights from this body of work.

We also draw insights from a small body of related work on litigation rates. For example, Miller and Sarat (1980) surveyed 1,000 households in order to discover grievance- and court-filing rates across a number of categories such as property, discrimination, landlord, and the like. Similarly, Hensler et al. (1991) surveyed 26,000 households in order to understand, inter alia, litigation rates for personal injuries such as motor vehicle and product-associated accidents.

Intuitively, economic analysis of litigation suggests that individuals are more likely to file suit when their expected rewards exceed their expected costs (Cooter & Ulen 2008:414–84; Cooter & Rubinfeld 1989). This hypothesis has been supported by some empirical work (Clermont & Eisenberg 2002), especially in the area of financial patent litigation (Lerner 2010). For instance, Dunbar and Sabry (2007) examine plaintiff demographics, injury severity, and economic factors in the propensity for victims of work, car, or product-related injuries to sue. They find that severity of injury is significantly correlated with litigation. Viscusi (1986:326) provides evidence that case outcomes are correlated with the defendant's alleged violations of government regulations.

Priest and Klein (1984) propose a theoretical model of the plaintiff win rate, which holds under general conditions and is robust to multiple types of liability regimes, judicial biases, and distribution of disputes (for empirical validation, see Wittman 1988). Alternatively, Shavell (1996) presents a brief but competing model in which he argues, under other conditions, that any frequency of plaintiff victory is possible. A finding from this literature that holds particular relevance is that statistical models studying outcomes often suffer from omitted variable and sample selection biases when the collection of suits reaching judicial ruling (or settlement) is not representative of the larger set of cases that begin a dispute (Clermont & Eisenberg 1998, 2002; Boyd & Hoffman forthcoming).

As a whole, these streams of literature help inform this article in a number of ways. First, the economic analysis of litigation provides the foundational theories upon which we develop our hypotheses. In addition, we leverage the existing research on docketology to help inform our data collection, and we leverage the research on settlement and dispute resolution to overcome chronic forms of bias. However, while economic and legal scholarship has examined various aspects of data breaches, their harms, and the legal theories brought by plaintiffs, to our knowledge this is the first article to empirically examine privacy litigation generally, and data breach litigation specifically.

III. Data

This research is based on a number of data sets we collected and combined. We first obtained a list of publicly reported data breaches. We then used Westlaw (an online legal research service) to identify federal data breach lawsuits. Finally, we used PACER (Public Access to Court Electronic Records) to obtain docket filings. For the purpose of this

research, a data breach is defined broadly as the unauthorized disclosure of personal information by an organization.

A. Data Collection

To address our first research question ("Which breaches are being litigated?"), we first gathered a list of reported U.S. data breaches from the Open Security Foundation (Datalossdb), a nonprofit organization devoted to collecting and recording data breaches and IT vulnerabilities, and that is one of the most comprehensive collections of reported data breaches.[10] This data set contains the name and industry of the breached entity, the number of records compromised, the date of breach, the cause of breach,[11] and the types of information lost. Then, we used Westlaw to identify which of these reported breaches resulted in federal litigation.

To address our second research question ("Which data breach lawsuits settle?"), we used Westlaw to perform a systematic search for all federal lawsuits in which plaintiffs alleged an unauthorized disclosure of their personal information.[12] (The lawsuit observations previously used are, of course, a subset of the results from this search.) Specifically, we searched Westlaw's Pleadings database using the following search strings: "personally identifiable information," "personal information," and either "data breach," "security breach," or "privacy breach." These search terms balance specificity without biasing search results to specific causes or types of data breach lawsuits. We then manually examined the results and extracted those cases relating to unauthorized disclosure of personal information.[13] We believe this is an appropriate combination of methods for identifying all lawsuits either filed in, or removed to, federal court and therefore represents the most complete collection of federal data breach lawsuits. We address issues related to collecting state actions later in this article.[14]

---

[10]These data are used per the OSF license agreement, which states: "permission is granted to use this database in non-profit works and research."

[11]The causes of the breach, as coded by Datalossdb, included 51 unique types. However, many categories are variations that can easily be reduced to the following three: "loss or theft" (i.e., accidental loss or theft of computing hardware that happened to contain personal information), "disclosure" (i.e., personal data carelessly made publicly available), and "hack" (i.e., the deliberate theft of personal information through cyber attack).

[12]Certainly, the ideal data set would include all state and federal suits. However, we focus on federal suits only in this article.

[13]For consistency in analysis, we omitted cases relating to, for example, a breached entity suing an individual alleged to have stolen data (21 instances), individuals suing entities for unauthorized collection or use of personal information (114), or known state actions (79).

[14]Legal scholars increasingly use the Pleadings database in a way that suggests that they believe it to be fairly comprehensive (Heise & Sisk 2012; Hoffman 2011; Mammem 2009; Boyd & Hoffman forthcoming). However, there is no easy way to test this assertion rigorously. One possible strategy would be to compare a sample of all federal cases in a given year using AO codes with a sample of cases obtained from Westlaw's Pleadings database. Our preliminary investigations in this vein suggested that some fields of relevance to this project—fraud, consumer credit, and FOIA cases—were comprehensively covered by Westlaw's docketing databases in the time period at issue. Westlaw appeared

We used PACER to retrieve the court docket for each case. From the docket itself we coded the following information: presiding judge, date filed, date terminated, forum, the law firms involved in the suit, and number of docket filings. We then purchased the complaint (or amended complaint where appropriate) and coded information relating to the breach such as the date of breach, size and cause of the breach, types of information compromised, and all causes of action. We also identified whether any dispositive motions were filed, and coded the disposition of the case. Settlement information (such as actual confirmation of a settlement, and amounts of any damage awards) was obtained either from the docket filings, or from directly contacting the litigating attorneys.[15]

B. Data-Generating Process

Data breach and lawsuit data are generated from the processes shown in Figure 1.

1. Stage 1: Reported and Unreported Breaches

As mentioned, for the purpose of this research, a "data breach" is defined as the unauthorized disclosure of personal information. From this population of events only a subset will become public knowledge and "reported" by the Datalossdb clearinghouse. Specifically, the only breaches that are included in this clearinghouse are those relating to Social Security numbers, financial/banking information, credit card numbers, or medical information, and where the number of records compromised exceeds 10.[16] This collection of reported breaches originates from a community of dedicated security professionals who obtain data breach information from news sources across the country, from Freedom of Information Act (FOIA) requests to state agencies, and from many individual contributing members across the country. This group has been systematically collecting data breach information since at least 2005.[17]

Awareness of breaches also stems from U.S. state laws requiring companies to notify individuals when their personal information is lost or stolen. California first adopted this type of law in 2003, with other states following (by the end of 2011, at least 46 states had

---

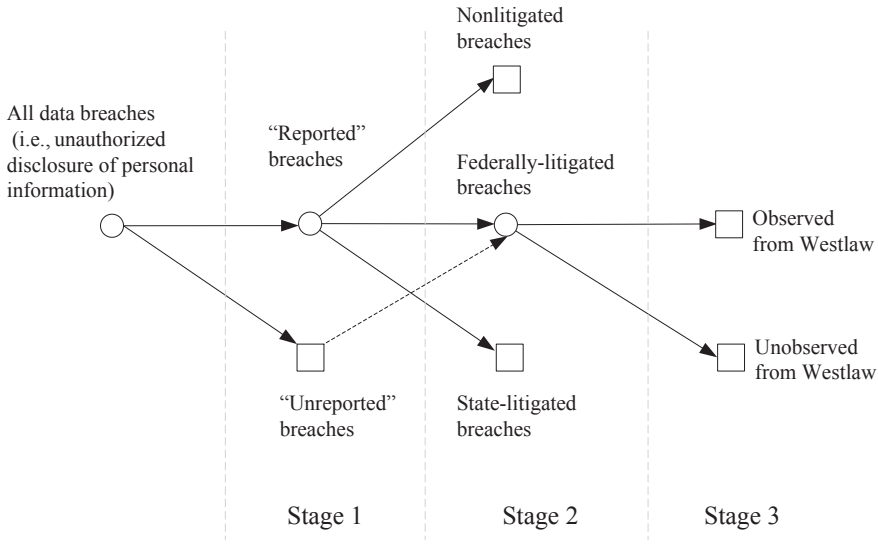
to collect fewer cases in which individuals sued the government. This would be an unsurprising finding given the purposes of Westlaw's data collection and monetization efforts. However, given that AO coding is under- and overinclusive (Boyd et al. forthcoming) it is not obvious that these results are dispositive. To be sure, if the way that Westlaw collects federal dockets is systematically biased against particular kinds of legal theories—which we have no evidence to be the case—then our results would be correspondingly skewed.

[15]Class action settlements were sometimes publicly available, and in some cases we were able to obtain settlement details for individual actions. Many times, however, only confirmation of settlement was available, with all other details being privileged.

[16]Note that the sample of "unreported breaches" (the dotted line from Stage 1 to Stage 2 in Figure 1) also contains observations that would be nonlitigated, federally litigated, or state litigated. However, when addressing our first research question ("Which data breaches are litigated?"), we do not include these observations.

[17]See http:// datalossdb.org/about, last accessed Jan. 22, 2013.

Figure 1: Process by which data breach observations and, subsequently, data breach lawsuits are generated.



NOTE: First, from the population of all breaches, some will be publicly reported while others will not. Clearly, we only observe reported breaches and therefore limit our inferences to this sample. Next, we use Westlaw to identify which reported breaches are litigated in federal court. Estimates of the effects of federally-litigated breaches pool state-litigated and nonlitigated breaches. Finally, we use Westlaw to acquire a representative sample of all federally-litigated data breaches.

adopted similar laws).[18] Two characteristics of state disclosure laws can affect the proportion of all breaches that are reported. First, there is heterogeneity among the state laws regarding the threshold of disclosure; about half the laws require notification only if there is a reasonable risk of malicious use of the data (high threshold), as opposed to simple loss of the data (low threshold). Second, it is the residence of the individual that drives disclosure, not the location of the breach. That is, disclosure to an individual is required only if the state in which the individual is a citizen has adopted a disclosure law. These properties suggest that breaches are less likely to be systematically reported if they affect citizens of states with higher thresholds for notification, or affect citizens of a state without a disclosure law.

There are, however, a number of mitigating factors that should reduce this systematic nonreporting. First, conversations with defense attorneys suggest that because it is quite costly for firms to separate disclosure requirements among differing states' citizens, it is easier for firms to simply notify all individuals. Indeed, firms also choose to notify all individuals independently of any particular state law as a means of managing public relations or due to pressure from states' attorneys general. For example, Choicepoint

[18]See http://www.ncsl.org/issues-research/telecommunications-information-technology/security-breach-legislation-2011.aspx, last accessed Jan. 22, 2013.

notified consumers from all affected states of its breach in 2005 even though only California had a disclosure law at that time (Ryan nd). Similarly, firms, once confronted with legal requirements from disparate states, may simply choose to follow the strictest law—which would require notification regardless of any threshold.

In addition, one may be concerned that organizations weigh the costs and benefits of disclosure and rationally choose not to notify consumers. However, conversations with privacy attorneys suggest that firms find this practice too risky and therefore obey the law. Together, these effects should minimize systematic nonreporting of data breaches.

2. Stage 2: Nonlitigated, State-Litigated, and Federally-Litigated Data Breaches

Stage 2 describes three separate outcomes from the sample of reported breaches: nonlitigated, federally litigated, or state litigated.[19] Because our key research questions relate to federal policy solutions to resolving the externalities caused by data breaches, our empirical focus compares *federally-litigated breaches* with *non-federally-litigated breaches* (i.e., both state-litigated and nonlitigated breaches). It is important to note that by pooling state-litigated and nonlitigated breaches we are still able to obtain unbiased estimates of *federal* lawsuits resulting from *reported* data breaches.[20] We discuss data limitations from unobserved state lawsuits in Section VI.

3. Stage 3: Federal Lawsuits Observed from Westlaw

For Stage 3, we obtained a sample of federal lawsuits through Westlaw using a systematic search strategy designed to identify the largest collection of data breach lawsuits practical, and then manually edited the list of suits matching our research question. Investigations by researchers have concluded that the Westlaw Pleadings database (used in this analysis) "covers or nearly covers the universe of federal claims [as it related to veil-piercing lawsuits]" and that it "was designed to collect all federal complaints since 2000 that lawyers litigating commercial cases would have a plausible interest in learning about. Thus, Pleadings may exclude civil rights cases, or habeas petitions, or family disputes, but attempts to collect every tort, contract, or federal statutory claim brought against corporate defendants" (Boyd & Hoffman forthcoming). Therefore, we do not believe that the use of Westlaw would pose any significant selection bias for our analysis.

It is relevant to also mention that the sample of unreported breaches may result in no federal or state litigation, although—for clarity—only the path to federally-litigated

---

[19]Arbitration is one further category of outcome that may exist. In these cases, plaintiffs, as a result of enjoying a firm's good or service, are contractually bound to resolve any legal dispute through arbitration, rather than civil court. However, we are unaware of any arbitrations in which privacy rights have been adjudicated.

[20]Alternatively, had we complete data on all three outcomes, one might choose to estimate a multinomial logit model in order to separately estimate marginal effects on federal-litigated versus state-litigated breaches. Or, one might pool state and federal suits together in order to draw inferences about all litigated breaches. However, because our topic of interest is primarily federal policy matters, we pool all non-federally-litigated outcomes (i.e., state-litigated and nonlitigated breaches).

breaches is drawn in Figure 1 (these data are included for the purpose of our second research question: "Which data breach lawsuits settle?").

IV. Which Data Breaches Are Litigated in Federal Court?

A. Hypotheses

Cooter and Rubinfeld (1989) examine prior theoretical models of litigation to create a unified framework for legal disputes. They present an analytical foundation describing the tensions faced by injurer and victim (defendant and plaintiff) at each stage of a dispute. First, when deciding whether or not to prevent an accident, an injurer balances the (marginal) cost of care with the (marginal) cost of an accident. Then, when deciding whether or not to sue, a plaintiff compares the cost of litigation with the expected benefit from an award. Finally, when deciding whether to settle or proceed to trial, both plaintiff and defendant balance their expected costs of litigation with the outcome from trial. This section is concerned with the second stage (the alleged victim's decision to file suit), which is increasing in both the probability of success and magnitude of award (his or her expected gain). Below, we adapt these conditions to data breach litigation to construct appropriate hypotheses.

First, we consider the magnitude of a potential award. Given that most data breach lawsuits are filed as class actions (76 percent in our data set), the magnitude of a plaintiff's award becomes a function of the size of the class, which is proportional to the number of records compromised in the data breach. If it is true that class action lawsuits are, in general, driven by class action plaintiffs' attorneys, it follows that the larger the data breach, the greater the potential fee award to the attorney, and the greater the incentive to bring and litigate the suit.[21] Therefore, *the probability of a lawsuit is positively correlated with the number of records lost* (H1a).

Next, the probability of a favorable outcome is multifaceted. Among other things, it is a function of whether an alleged harm can be attributed directly to the breach, the cause of the breach, and the types of information lost.

Plaintiffs in many data breach lawsuits seek relief for harms such as actual financial loss from identity theft, emotional distress, costs of credit monitoring, and anticipated future losses. However, a critical factor affecting the success of a lawsuit is the presence of a cognizable harm for which the law could provide a remedy. In the context of data breach litigation, this is manifested by whether the plaintiff can allege (though would not yet have to prove) financial harm. Moreover, plaintiff harm (loss) is also a function of whether the breached firm provided any initial compensation immediately following the breach and before litigation. This redress is commonly offered in the form of credit monitoring or identity theft insurance. Full compensation for any loss will decrease plaintiffs' legal

---

[21]It is not the purpose of this research to address the motivations of attorneys, but merely to understand and apply relevant behavior in forming reasonable hypotheses. Conversations with class action plaintiffs attorneys confirm that while it is true that attorneys do seek plaintiffs, plaintiffs also seek attorneys for class action litigation.

remedies. Therefore, *the probability of a lawsuit is positively correlated with the presence of actual harm, and negatively correlated with credit monitoring* (H1b).

The legal merits matter. In the context of data breaches, a plaintiff's case is strengthened by his or her ability to prove that the defendant had a legal duty to protect the plaintiff's personal information, and somehow failed in that duty. This could occur in two different ways.

The first manner relates to the cause of the breach, which typically occurs in one of three ways: improper disclosure or disposal of personal information (e.g., tossing tax records in a dumpster); a computer hack (e.g., computer-based theft of information); or loss or theft of hardware (e.g., computer hardware that happens to contain personal information). Of these methods, we consider that the first cause (the careless handling of personal information) may provide the strongest legal argument because it involves negligent behavior on the part of the data custodian, as opposed to misfortune or petty theft. Therefore, *lawsuits are more likely to occur from breaches caused by improper disclosure of information, relative to the computer hack, or loss of hardware* (H1c).

The second manner relates to the types of information compromised. It is reasonable to consider that the greater the legal duty to protect certain information (typically enforced through statute), the greater the probability of a favorable outcome. For instance, organizations using medical and financial data are governed by a regulatory environment requiring the enhanced protection of such data. The Health Information Portability and Accounting Act (HIPAA) requires patient consent before the disclosure of medical information between health agencies. The Gramm-Leach-Bliley Act (GLBA) and Fair Credit Reporting Act (FCRA) require greater security controls protecting an individual's credit data. In addition, many state and federal laws require the proper disposal of Social Security numbers (Dickey et al. 2011) and the storage and transmission of credit card data is also protected through contractual agreements by the credit card companies under the Payment Card Industry Data Security Standard (PCI-DSS). Therefore, *the probability of a lawsuit is positively correlated with the compromise of personal information requiring a heightened level of protection, such as Social Security numbers and financial, credit card, and medical data* (H1d).[22]

B. Descriptive Statistics

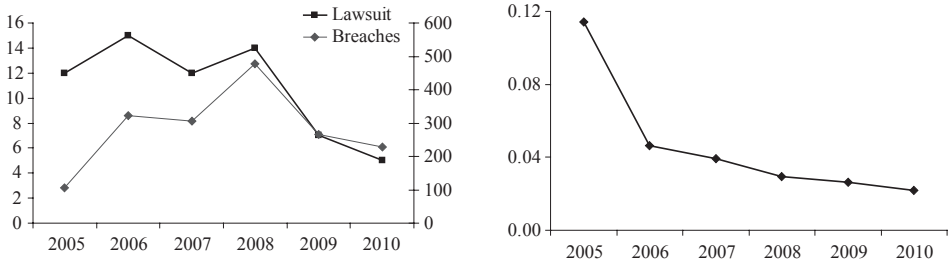
The entire Datalossdb clearinghouse consists of almost 3,000 data breaches. However, since the primary research question of this section focuses on estimating the probability of a federal lawsuit conditional on covariates, we prune the data set in a number of ways. Given that systematic collection of data breaches did not begin until 2005 (after the first breach disclosure law was adopted), we limit the duration of our analysis from 2005 to 2010. Observations with missing or ambiguous data are also omitted,[23] though the descriptive analyses presented below are robust to their inclusion. The resulting data set consists of

---

[22]Note that we employ the general categories used in the Dataloss clearinghouse and that these categories are not mutually exclusive: a data breach can compromise one or more types of data.

[23]For example, the number of records compromised in some breaches is not known.

Figure 2: Absolute and relative levels of data breaches and data breach lawsuits.



NOTE: The left panel separately compares the levels of total data breaches with litigated breaches, while the right panel presents the same data as a ratio (i.e., the portion of reported breaches that are litigated). In the left panel, the left axis reflects the number of lawsuits (0–16), while the right axis reflects the number of breaches (0–600).

SOURCES: Datalossdb clearinghouse (reported breaches); Westlaw (litigated breaches).

1,772 U.S. data breach observations, of which only 65 (3.7 percent) were litigated in federal court.[24]

Figure 2 compares the number of reported data breaches with the number of federally-litigated breaches during the period 2005 to 2010. In the left panel, lawsuits are scaled according to the left axis (0–16), while reported breaches are scaled according to the right axis (0–600). The right panel shows the ratio of filed lawsuits to the number of breaches reported in that year (i.e., the portion of federally-litigated breaches over time). The right panel shows that in 2005, the proportion of federal lawsuits was about 10 percent. However, since 2005, the proportion of federal lawsuits appears to be declining slightly, reaching around 3 percent in 2010. This 3 percent litigation rate appears to be on order with other estimates of personal injury or loss. Miller and Sarat (1980:Table 1) find a litigation rate of 3 percent in "consumer-related" court filings, and Hensler et al. (1991:Table 4) find 7 percent and 3 percent litigation rates in work and non-work and non-motor-vehicle accidents, respectively. Notably, however, the 3 percent observed filing rate in our data treats each breach as a single harm: if we focused on individuals potentially affected by the loss of their PII, the "real" filing rate would be far lower.

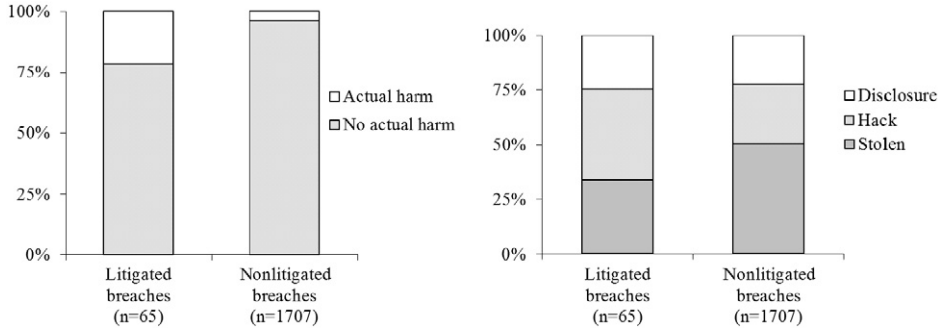
Notice that the number of reported breaches generally increased from 2005 to 2008, and decreased thereafter. Federal lawsuits, on the other hand, fluctuated slightly until 2008, after which they declined. The rise in *reported* breaches is likely a result of state data breach disclosure laws, which started becoming most popular in 2005 (Romanosky et al. 2010:Figure 6). But why have they since declined? If it were true that data breach incidents were primarily collected from news articles, then this decline might be caused by the erosion of media or consumer interest.[25] Another explanation is that U.S. data breach

---

[24]Note that while only 65 lawsuit observations are used to identify effects in this first research question, we use a larger set of observations in the second research question.

[25]However, note that while any changes in reporting may affect the proportion of breaches "reported" by Dataloss, this will not bias our regression estimates from our first research question ("Which breaches are litigated") because our inferences consider the Dataloss clearinghouse data as exogenously provided. Further, we do not use these data when examining our second research question ("Which lawsuits settle?").

Figure 3: Portion of litigated and nonlitigated breaches according to whether there was evidence of actual financial loss (left panel), and according to the cause of the breach (right panel).



NOTE: Percentages displayed sum to 100 percent across categories.
SOURCES: Datalossdb clearinghouse and Westlaw.

disclosure laws have, indeed, forced firms to internalize more of the cost of a breach, inducing them to invest more to protect personal information, and reducing the number of actual data breaches. This claim is partially substantiated by Verizon (2010:7), Symantec (2012), and the Identity Theft Resource Center[26] showing a reduction in data breaches.

As one might expect, the number of compromised records is highly correlated with breaches litigated in federal courts (i.e., federally-litigated breaches). The mean number of records compromised by non-federally-litigated breaches ($n = 1{,}708$) is just over 98,000, while the mean number of records compromised in federally-litigated breaches ($n = 103$) is over 5.3 million, providing suggestive support for H1a.

Figure 3 compares federally-litigated and non-federally-litigated breaches as a function of the presence of actual harm (left panel), and the causes of breach (right panel).

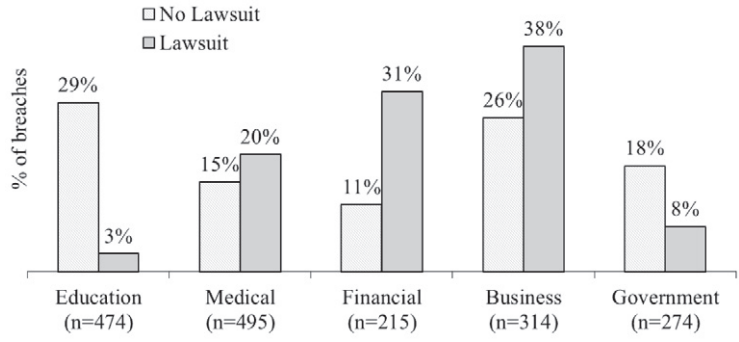
As shown in the left panel of Figure 3, 78 percent of federally-litigated breaches did not result in financial loss, while 22 percent did result in financial loss.[27] However, breaches appear less likely to be litigated in federal court absent financial harm, providing suggestive support for H1b. The right panel of Figure 3 shows that breaches resulting from the unauthorized disclosure (or disposal) of personal information and computer hack (cyber attack) are *more* likely to be litigated in federal court, while breaches due to lost/stolen hardware are *less* likely to be litigated in federal court, providing suggestive support for H1c.

---

[26]See reports available at http://www.idtheftcenter.org/artman2/publish/lib_survey/ ITRC_2008_Breach_List .shtml, which show the following breach levels: 2012: 335 (as of Oct. 16, 2012), 2011: 419, 2010: 662, 2009: 498, 2008: 657, 2007: 446.

[27]The presence of actual harm is coded as follows: for known lawsuits, we code a 1 if the complaint includes some allegation of financial loss as a direct result of the breach. For data breaches not resulting in lawsuit, we refer to news articles associated with the breach, and similarly code a 1 if the article mentions financial loss resulting from the breach. Given that it is extraordinarily difficult to obtain full information about all possible financial losses, our results very likely provide a lower threshold of loss.

Figure 4: Portion of litigated and nonlitigated data breaches according to the types of data compromised.



SOURCES: Datalossdb clearinghouse and Westlaw.

Note that these figures reflect data from all years, but that the patterns presented in both panels are robust when examining individual years.

Figure 4 compares breaches that were and were not federally litigated as a function of the types of personal information compromised. Note that a single breach may result in the compromise of multiple types of personal information.

Breaches involving financial data and credit card numbers are more likely to be litigated in federal court, which provides some support for H1c. Social Security numbers (SSN), on the other hand, compromised about 78 percent of nonlitigated breaches, though only 58 percent of litigated breaches. Medical data appear to be equally represented in federally-litigated and non-federally-litigated breaches.

C. Estimating Model

To test hypotheses H1a–H1d, we estimate a binary outcome model predicting the probability that a reported data breach will result in a federal lawsuit:[28]

$$\begin{aligned} lawsuit_i = \alpha_0 + ActualHarm_i + CreditMonitoring_i + BreachSize_i + \\ Cause_i + ProtectedPII_i + OtherPII_i + Industry_i + Year_i + \varepsilon_i \end{aligned} \quad (1)$$

where *lawsuit* is a binary variable that takes the value 1 if a reported breach, *i*, results in a federal lawsuit, and 0 otherwise.[29] Although we cannot determine with absolute certainty

[28]Equation (1) is shown as a linear probability model for clarity only. Actual regressions are estimated using logit. Also note that we limit inferences to predictions of the probability of a *known federal* lawsuit conditional on a *reported* data breach.

[29]Note again that this coding inherently pools state-litigated and nonlitigated breaches, thereby ensuring that estimates of federal lawsuits from reported breaches are unbiased.

whether financial loss had occurred following a data breach, we can proxy for this by observing any evidence from news reports following the breach. Therefore, *ActualHarm* is coded 1 if we observe any evidence of financial loss due to the breach, and 0 otherwise.[30] *CreditMonitoring* is a dummy variable coded 1 if there was any evidence that the breached firm provided any sort of credit monitoring or identity theft insurance to the individuals following the breach.[31] *BreachSize* is a continuous variable representing the log of number of records compromised. *Cause* is a vector of mutually exclusive and completely exhaustive dummies reflecting the cause of the data breach: improper disclosure or disposal, computer hack, or lost/stolen hardware.[32] *ProtectedPII* is a vector of dummies representing types of personally identifiable information (PII) that should require a heightened level of protection, as described in hypothesis H1d: Social Security number, medical, financial, credit card). *OtherPII* controls for all other data types (email address, name/address, date of birth, and miscellaneous). *Industry* is a vector of dummies representing the industry of the breached firm and whether the firm was a nonprofit or publicly traded. *Year* is a vector of year dummies (2005 to 2010) reflecting the year of the data breach and ε_i is the random error term, assumed to be independent of the observed covariates. Identification of the variables of interest comes from the portion of federally-litigated breaches. Descriptive statistics for the variables used in Equation (1) (and in Equation (2)) are shown in Table 1.

D. Results

The results of Equation (1) are presented in Table 2 and reflect the average marginal effects of the explanatory variables on the probability of federally-litigated data breach (relative to the probability of a non-federally-litigated data breach) estimated using a logit regression.[33] Model 1 presents just the variables of interest from H1a–H1d and includes

---

[30]Of the 1,772 data breaches, we were unable to find news reports for 83 of them. In the absence of evidence, we took the most conservative approach and coded these breaches as not causing actual harm. We then performed a robustness check by considering that all 83 observations did cause actual harm. All estimates maintain qualitative magnitude and significance except for *ActualHarm*, which reduces in magnitude by one-third and therefore loses statistical significance. One may also be concerned that plaintiffs may wait many years following a breach before filing suit; however, we do not find evidence of this. In a sample of 146 single-suit breaches, 78 percent were filed within one year, and 87 percent were filed within two years of public notification.

[31]This information was obtained from breach disclosure notices obtained by the Datalossdb clearinghouse, or through news reports, when available. Given that perfect information is not always available, we code this variable equal to 1 only when there is actual evidence of redress. As a result, this variable is likely an underestimate of the true frequency.

[32]As is customary with categorical variables, we will omit one of these from the regression analysis. Given that the selection is arbitrary, we omit "lost/stolen."

[33]Note that the marginal effects for logit models are nonlinear functions of the parameter estimates, and so the effect of a regressor on the probability of lawsuit can either be presented as the effect for the "average observation" (i.e., marginal effect computed at the sample mean of the regressors), or the "average effect" (i.e., computing the marginal effect for all observations and taking the average). We believe the second approach is more appropriate for our model because: (1) we avoid the confusion of subjectively determining the value of the regressor at which to compute the marginal effect, as in the case of the logged regressor, and (2) given that most explanatory variables are dummies, we do not need to justify having to calculate the marginal effect at a sample mean of a binary regressor.

Table 1: Summary Statistics for Equations (1) and (2)

	Equation (1), n = *1,772*		*Equation(2),* n = *164*	
Variable	*Mean*	SD	*Mean*	SD
Log(records compromised)	7.91	2.87	9.58	5.46
Actual harm	0.05	0.21	0.17	0.38
Breach_disclosure	0.23	0.42	0.58	0.50
Breach_hack	0.28	0.45	0.23	0.42
PII_SSN	0.77	0.42	0.37	0.48
PII_medical	0.12	0.33	0.09	0.29
PII_financial	0.09	0.28	0.27	0.45
PII_creditcard	0.12	0.32	0.26	0.44
PII_email	0.03	0.16	0.04	0.19
PII_nameaddress	0.77	0.42	0.34	0.47
PII_dateofbirth	0.16	0.37	0.15	0.35
Ind_business	0.27	0.44	0.49	0.50
Ind_education	0.28	0.45	0.02	0.15
Ind_financial	0.12	0.33	0.28	0.45
Ind_government	0.18	0.38	0.12	0.32
Non-profit	0.03	0.16	0.18	0.38
Publicly traded	0.12	0.32	0.41	0.49
Class action suits			0.76	0.43
Class certification			0.12	0.33
Statutory damages			0.54	0.50
Multisuit cases			0.18	0.38
Removed			0.14	0.35
Female judge			0.24	0.43
Settled			0.52	0.50
Standing			0.08	0.27
Log(employees)			8.73	2.80

NOTE: PII refers to "personally identifiable information."

SOURCES: Complaint and docket data were collected from PACER.

only *Year* controls, whereas Model 2 includes all data types. Models 3a and 3b control for industry variables; they are based on the same estimating equation, but Model 3b presents the results as odds ratios.

The results are robust across all models, with the third model—which controls for all variables—providing the better fit for the data and generally more conservative estimates. Though not shown, results are also robust to the exclusion of individual years 2005–2010, and to probit models. Further discussions therefore focus on results from Model 3a.[34]

In regard to the effect of the size of the breach on probability of lawsuit, our results suggest that a 10-fold increase in the number of compromised records increases the average

[34]Given the potential for certain data types (like credit card and medical) to be correlated with industry variables (such as financial and medical agencies) additional regressions were run excluding these data types. All models were robust to these alternative specifications.

Table 2: Regression Results of Equation (1), Estimating the Probability of Federally-Litigated Data Breaches

Dependent Variable: Lawsuit	*Basic Model (1)*	*All Data Types (2)*	*Full Model (3a)*	*Full Model (Odds Ratio) (3b)*
Log(records)	0.013***	0.012***	0.009***	1.592
	(0.002)	(0.002)	(0.001)	
Actual harm	0.046***	0.045***	0.025*	3.557
	(0.014)	(0.014)	(0.014)	
Credit monitoring	–0.017*	–0.017*	–0.037***	0.152
	(0.009)	(0.009)	(0.010)	
Cause_disclosure	0.025*	0.013	0.027**	3.122
	(0.013)	(0.011)	(0.013)	
Cause_hack	0.004	–0.001	0.016	2.085
	(0.010)	(0.009)	(0.012)	
PII_SSN	–0.006	–0.001	0.010	1.729
	(0.009)	(0.009)	(0.007)	
PII_medical	0.034**	0.025*	0.010	1.619
	(0.016)	(0.014)	(0.014)	
PII_financial	0.094***	0.079***	0.051***	5.875
	(0.025)	(0.023)	(0.016)	
PII_credit_card	0.019	0.018	0.005	1.263
	(0.014)	(0.013)	(0.010)	
Year controls	Y	Y	Y	
PII controls		Y	Y	
Industry controls			Y	
Observations	1772	1772	1772	
Log likelihood	–174.63145	–165.70501	–131.40823	
Pseudo R^2	0.3733	0.4053	0.5284	

NOTE: Robust standard errors are in parentheses, ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Models 1–3a represent average marginal effects of logit regressions where the dependent variable is the probability that a reported data breach will be litigated in federal court (relative to a data breach not being litigated in federal court). Models 1, 2, and 3a control for increasingly more covariates. Model 3b is the same regression model as 3a but coefficients are presented as odds ratios. Independent variables for Cause_disclosure and Cause_hack are presented relative to Cause_lost/stolen. PII refers to "personally identifiable information."

SOURCES: DatalossDB clearinghouse (breach data); Westlaw and PACER (lawsuit data).

probability of lawsuit by 8 percent (from 3.7 percent to 11.7 percent), a statistically significant amount (at the 1 percent level), which supports H1a.[35]

Supporting H1b, the presence of actual (financial) loss is associated with a 2.5 percent increase in the probability of litigation (though only significant at the 10 percent level), while the presence of credit monitoring is associated with a 3.7 percent decrease in probability of litigation (significant at the 1 percent level). Described in terms of odds-ratios (Model 3b), these results suggest that the odds of a firm being sued are 3.5 times *greater* when individuals suffer actual (financial) harm, but 6 times *lower* (1/0.152) when the firm provides free credit monitoring following a breach. Although credit monitoring is widely touted as a best practice following a data breach (e.g., GAO 2007b; DOJ 2008; Wolf 2012)

[35]A 10-fold increase represents a change of 900 percent, or 0.009 * 9 = 0.081 or 8.1 percent.

and, indeed, is included as part of a recent federal data security bill (HR2221), we provide the first statistical evidence to substantiate the practice's value in reducing an organization's ex post liability costs.

Next, we examine the relative odds of a lawsuit occurring given the different causes of the data breach (unauthorized disclosure, hack, or lost/stolen). Our results suggest that the odds of a firm being sued due to the unauthorized disclosure/disposal of consumer information are three times greater, relative to breaches caused by lost/stolen data (significant at the 5 percent level), supporting H1c. Breaches caused by cyber attack, however, are not statistically more likely to result in a suit (again, relative to breaches caused by lost/stolen data). This suggests that individuals are much more likely to punish (alternatively, attorneys are more confident in filing suits against) firms that are thought to have behaved carelessly with consumer information, relative to the firm being the unfortunate victim of computer hardware theft.

Among all types of personally identifiable information requiring greater protection, we find that only the compromise of financial data is significantly correlated with the probability of lawsuit: the compromise of financial data increased the probability of lawsuit to 5.1 percent (significant at the 1 percent level), which provides only partial support for H1d. That is, the odds of a firm being sued are almost six times greater for breaches involving the loss of financial information (relative to a breach not involving financial information).

Surprisingly, however, not all forms of data were found to be positively correlated with litigation. Indeed, breaches involving the compromise of medical or credit card data produce no significant effect. The cause for this could be that plaintiffs (and attorneys) believe that loss of financial information may more easily lead to financial harm, thereby elevating their subjective belief of a successful lawsuit. That is, they may feel that it is easier to justify bringing a claim for the breach of financial information because of the increased risk of a legally justifiable harm.

Overall, we find that our results support theoretical models of litigation. In this arena, dominated by class action practice, parties appear to behave in a rational and wealth-maximizing manner. In the context of data breaches, this translates to a higher probability of a federal lawsuit given evidence of actual financial loss, stronger claims of negligence (unauthorized disposal of information), and heightened protection of personal financial information. However, notwithstanding the statistically significant results, none were large in magnitude. That is, no marginal effect was larger than 5 percent. It is yet unclear whether the magnitude of these findings is, in itself, unexpected, though it does warrant further consideration.

Next, we examine the characteristics of data breach lawsuits leading to settlement.

V. Which Data Breach Lawsuits Settle?

A. Hypotheses

Section IV leveraged the theoretical analysis of dispute litigation to develop hypotheses explaining the probability of a federal data breach lawsuit. We continue that process to develop hypotheses regarding the probability of settlement once a suit has been filed.

Cooter and Rubinfeld (1989) consider that a plaintiff (and his or her attorney) will decide to settle when the expected gains from settlement exceed the expected gains from trial. However, the vast majority of data breach lawsuits terminate before trial, either through dismissal (motion to dismiss or summary judgment) or by settlement. Indeed, of over 230 suits in our data set, we observe only two instances of a plaintiff prevailing to a favorable ruling by a judge or jury.[36] Therefore, we can simplify the theoretical model by stating that a plaintiff (and his or her attorney) will settle when the expected benefits from a settlement award exceed the cost of further litigation. We now adapt this theory to data breach litigation by examining conditions that would increase either the probability or magnitude of settlement.

The recognition of the legal merits or "case strength" of a lawsuit has been the topic of much analysis in legal scholarship (see, generally, Boyd & Hoffman forthcoming; Eisenberg & Lanvers 2009; and see, in regard to securities class action litigation, Johnson et al. 2007; Cox et al. 2008; McShane et al. 2012; Choi 2007). The lack of clear evidence of identity theft following a data breach (GAO 2007a) is one reason for dismissal of data breach lawsuits. For example, judges often determine that the plaintiff could not show that he or she suffered harm in a sufficiently concrete way to justify proceeding in a lawsuit.[37] However, there are cases when plaintiffs *do* suffer actual harm and *are* therefore able to overcome this procedural obstacle and obtain settlement.[38] Hence, we consider that in the context of data breach lawsuits the presence of "actual harm" represents an appropriate measure of a meritorious legal claim that should affect the probability of settlement. Therefore, *the probability of settlement is positively correlated with lawsuits in which the plaintiff is able to demonstrate actual harm* (H2a).

A second factor that may affect the magnitude of the settlement award is whether, in class action lawsuits, the class achieves certification. Class certification represents the difference between damages potentially awarded to only a few named plaintiffs versus thousands or millions of plaintiffs. Indeed, "class certification stands not as a mere judicial byway on the road toward full-fledged trial on the merits but, almost invariably, as the last significant judicial checkpoint on the road toward settlement" (Nagareda 2010:152). Therefore, *the probability of settlement is positively correlated with achieving class certification* (H2b).

A final driver potentially affecting the magnitude of settlement is statutory damages. Plaintiffs bring many kinds of common-law claims (e.g., negligence, breach of contract) and statutory causes of action; for example, the Computer Fraud and Abuse Act (CFAA),

---

[36]Conner v. Tate, 130 F. Supp.2d 1370 (N.D. Ga. 2001), in which a woman allegedly disclosed an illegally wiretapped conversation to local police, and Beaven et al. v. U.S. Department of Justice, 5:03-cv-00084 (E.D. La. 2007), in which the plaintiffs alleged a violation of the Privacy Act (1974).

[37]In Shafran v. Harley-Davidson, Inc., 1:07-cv-01365 (S.D. N.Y. 2008), "[p]laintiff has failed to show an actual resulting injury that might support a claim for damages. As damages are an essential element of each of plaintiff's claims, plaintiff's claims fail as a matter of law."

[38]In Stollenwerk et al. v. Tri-West Healthcare Alliance, 05-16990 (9th Cir. 2007), "[plaintiff's] personal data was used on six occasions to open or to attempt to open unauthorized credit accounts in [plaintiff's] name. Unknown individuals successfully opened at least two credit accounts and generated more than $7,000 in unauthorized charges to these accounts."

the Fair Credit Reporting Act, and the Electronic Communications Privacy Act. A defining characteristic of these Acts is that their mere violation can justify plaintiff relief through statutory damages. For example, the Wiretap Act allows recovery up to $100 per day or $1,000, whichever is greater[39] and the CFAA allows statutory damages of $5,000 per incident (record compromised). Hence, we consider that defendants may be more likely to settle when complaints include causes of action with statutory damages. The reasons are twofold. First, these allegations shift the burden from the plaintiff having to demonstrate harm to the defendant having to prove that there was no violation of the law, increasing the defendant's cost of litigation. Indeed, "the only real significant liability threat to those companies sustaining a data breach is the advent of statutory damages—damages that would ensue with or without any showing of real harm to a plaintiff" (Paray 2011). Second, there may be a saliency effect when the defendant is forced to consider the potentially massive damage award that is the product of the statutory damage award and the size of the class. Therefore, *the probability of settlement is positively correlated with lawsuits in which the plaintiff seeks statutory damages* (H2c).

B. Descriptive Statistics

To address our second research question, we relax the restrictions imposed in Section IV and employ our full set of federal data breach lawsuits. Note that this data set is more comprehensive than that used in Section IV, in that it includes all federally-litigated breaches. Therefore, our sample now consists of 231 lawsuits filed in 50 unique district courts from 2000 to 2011.[40] Because we seek to compare settled and dismissed cases, we omit 23 pending lawsuits and two cases that ended in trial decision. We also drop 42 public actions because they are perfect predictors of settlement (i.e., the government never loses). The resulting data set of 164 observations consists of lawsuits that terminated either by settlement ($n = 86$) or dismissal ($n = 78$). The left panel in Figure 5 illustrates the number of such cases sorted by year of disposition, while the right panel shows the settlement rates, by year of disposition.

An interesting finding from this analysis is that the overall settlement rate in our data set (86/164 = 52 percent) is much higher than legal privacy scholarship would suggest (see, e.g., Solove 2010; Citron 2007; Rice 2007; Hutchins 2008; Serwin 2009), but also lower than the approximately 82 percent settlement rate found in previous work (Eisenberg & Lanvers 2009:Table 6).[41] The right panel of Figure 5 shows an early, erratic trend followed by a fairly constant settlement rate of around 50 percent after 2004.[42]

---

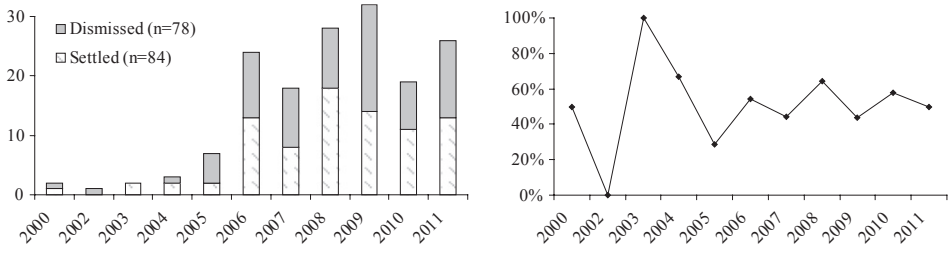
[39]18 U.S.C.A § 2520(c)(2).

[40]Note that this section employs observations only related to lawsuits, and no longer the Dataloss clearinghouse data.

[41]Although Eisenberg and Lanvers (2009) do find settlement rates as low as 50 percent for constitutional cases. Also note that the confidence interval computed by Eisenberg and Lanvers (2009), 78.1%—85.1%, $n = 448$, does not overlap with the confidence interval from our overall settlement rate, 45%—60%, $n = 164$.

[42]It is likely only a coincidence that this 50 percent settlement rate matches the theoretical settlement rate of Priest and Klein (1984) since that rate defines a *trial* settlement rate, whereas we observe no trials in this sample.

Figure 5: Settled and dismissed data breach lawsuits from 2000 to 2011.



NOTE: Right panel illustrates the percent of all suits settled in a given year.
SOURCES: Westlaw and PACER.

Figure 6 examines the proportion of cases in which plaintiffs were able to show actual damage (H2a), where the case achieved class certification (H2b), and where the plaintiff sought statutory damages (H2c). Note that in the following figures, percentages sum to 100 percent in each adjacent column pair.

The top two pair-wise comparisons in Figure 6 illustrate a similar result: the majority of cases that allege actual harm or achieved class certification settled.[43] That is, of the cases that alleged actual harm ($n = 28$), 71 percent settled, whereas only 49 percent without actual harm ($n = 135$) settled. Similarly, of the cases that achieved class certification, 85 percent settled, whereas when the class was not certified, only 48 percent settled.[44] The bottom panel, on the other hand, is more balanced. Of the cases that include causes of action with statutory damages, 59 percent settled, and only about 45 percent otherwise. Again, note that these figures reflect data from all years, and that the patterns presented in both panels are robust across individual years.

Interestingly, however, the top panels of Figure 6 show another consistent result: data breach lawsuits *lacking* actual harm or class certification are almost as equally likely to reach settlement as dismissal. That is, in cases without these characteristics, the plaintiff faces approximately a 50/50 chance of obtaining a settlement. We further discuss this observation in the following section.

C. Estimating Model

We again employ a binary outcome model to estimate the probability of settlement:[45]

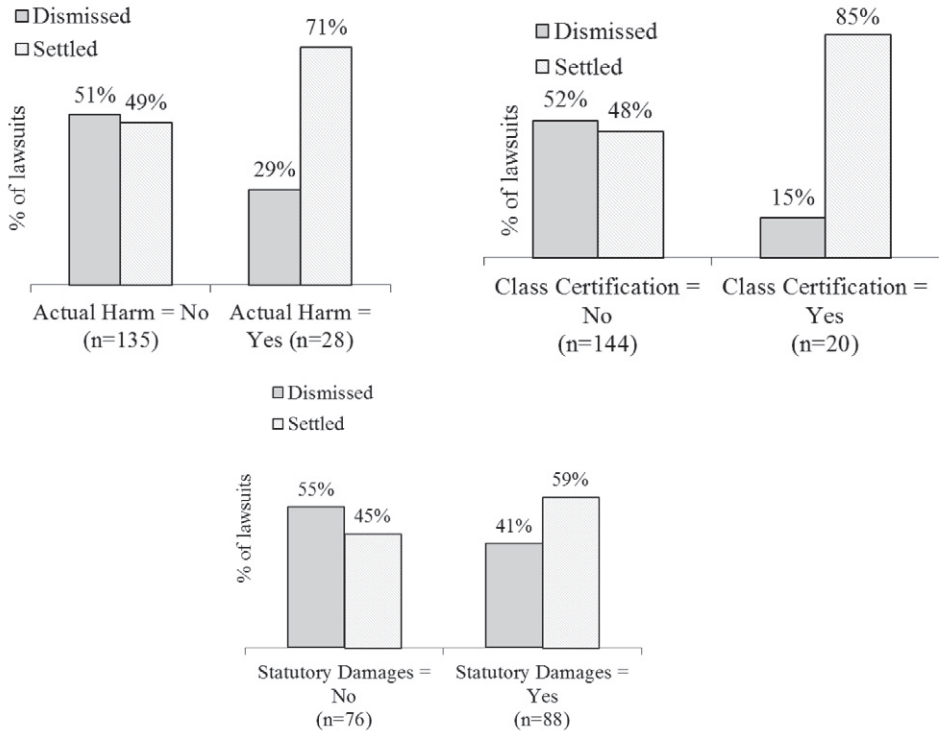
---

[43]Note that these are not the same cases. In fact, there are only two cases that both included class certification and actual harm.

[44]Note that for the purpose of this figure, we include four cases that were certified only for the purpose of settlement (i.e., not litigation). When estimating Equation (2), however, class certification for these four cases is coded as 0. We thank Paul Bond for bringing this distinction to our attention.

[45]Again, we present a linear probability model for readability, though the estimating model is nonlinear.

Figure 6: Pair-wise comparisons between dismissed and settled data breach lawsuits, as a function of actual harm, class certification, and cases alleging violation of statutes with statutory damages.



SOURCE: PACER.

$$settlement_i = \alpha_0 + ActualHarm_i + CreditMonitoring_i + ClassCertified_i + StatutoryDamages_i + Breach_i + Industry_i + Forum_i + Year_i + \varepsilon_i \quad (2)$$

where *settlement* is a binary outcome variable coded 1 if the lawsuit, *ie*, terminated in settlement and 0 otherwise.[46] *ActualHarm* is coded 1 if the plaintiff's complaint alleges an actual loss due to the breach (e.g., if the plaintiff alleges fraudulent charges on a credit card, stolen money from a checking or savings account, or other such costs incurred from criminal activity). Other forms of alleged harm such as preventive costs from credit

[46]We determine settlement either by contacting attorneys directly, or from case dockets. Also, recall that data breach lawsuits may terminate for a number of reasons and we therefore focus our analysis to address the probability of a lawsuit resulting in settlement versus being dismissed as a matter of law, summary judgment, or other technical legal reasons, such as subject matter jurisdiction.

monitoring, emotional distress, invasion of privacy, or embarrassment are coded 0.[47] *CreditMonitoring* is coded as per Equation (1). *ClassCertified* is coded 1 if the suit achieved class certification. *StatutoryDamages* is coded 1 if the complaint alleged violation of a federal statute allowing for statutory damages. *Breach* is a vector of controls for the size and cause of the data breach and types of information lost (PII). *Industry* is a vector of controls for the firm's industry and whether the firm is nonprofit or publicly traded (as in Equation (1)). *Forum* is a vector of controls for the circuit court region in which the case was heard, whether the case was removed from state court, and the sex of the judge (Boyd & Hoffman forthcoming). As a measure of the complexity of the case, we also control for the number of causes of action and number of times the complaint was amended. *Circuit* controls for the circuit region where the case was litigated. We may also be concerned with forum shopping (litigants filing cases in more favorable districts), and the possibility of a successful outcome in one case affecting the outcome in another case (formally referred to as the stable unit treatment value assumption, or SUTVA) (Rubin 1990). We partially control for both these effects by coding a variable, *Standing*, equal to 1 if a suit is filed in a district that had granted standing to a plaintiff in a previous data breach lawsuit. We also include a proxy for the size of the firm with the log of the total number of employees of the defendant. In cases with multiple defendants, we consider only the first-named defendant. We also code a variable, *Multisuit*, as 1 if a data breach resulted in more than one consolidated lawsuit. *Year* is a vector of dummies representing the year when the case was disposed of and ε_i is the random error term, assumed to be independent of observed covariates. Descriptive statistics for the variables used in Equation (2) are shown in Table 1.

Given that our analysis examines the binary outcome of lawsuits (settlement vs. dismissal), estimation of Equation (2) should not suffer from the familiar issue of case-selection bias (Priest & Klein 1984; Clermont & Eisenberg 1998). Also, recall that we are implicitly examining the determinants of lawsuit outcome, conditional on a complaint having been filed. Therefore, we are careful to interpret any parameter estimates as the marginal impact on the probability of a federal lawsuit, *not a data breach*, being settled. This is because such a model would suffer from selection bias if the set of breaches that resulted in settlement were systematically different from breaches otherwise disposed.

D. Results

Table 3 presents the results of Equation (2), reporting the average marginal effects of the explanatory variables on the probability of settlement.[48] Model 1 includes just the variables of interest and year fixed effects, while Model 2 includes subsequent controls for *Breach* and *Industry* characteristics. Models 3a and 3b include the full set of controls and estimate the same equation, with Model 3b presenting the results as odds-ratios.

---

[47]Note that we take no normative position in regard to whether such harms *should* be considered as actual harm. We merely make this distinction for the purpose of hypothesis testing.

[48]Note that some observations were dropped because of perfect prediction while some covariates were dropped because of collinearity, resulting in fewer than the full set of 164 observations being estimated.

Table 3: Regression Results of Equation (2), Estimating the Probability that a Federally-Litigated Data Breach Will Settle

Dependent Variable: Settled	*Basic Model (1)*	*With Breach and Industry Controls (2)*	*Full Model (3a)*	*Full Model (Odds-Ratio) (3b)*
Actual harm	0.275***	0.310***	0.302**	9.19
	(0.095)	(0.106)	(0.119)	
Credit monitoring	–0.041	–0.008	0.102	2.11
	(0.101)	(0.130)	(0.145)	
Class certification	0.407***	0.327**	0.304***	9.31
	(0.140)	(0.143)	(0.117)	
Statutory damages	0.163**	0.192*	0.097	2.04
	(0.078)	(0.103)	(0.096)	
Log(records)		0.003	–0.006	0.959
		(0.009)	(0.009)	
Breach_disclosure		0.085	0.170	3.63
		(0.138)	(0.135)	
Breach_hack		0.243**	0.290***	9.59
		(0.122)	(0.111)	
PII_SSN		0.113	0.078	1.79
		(0.101)	(0.108)	
PII_medical		0.310**	0.312***	15.00
		(0.142)	(0.094)	
PII_financial		-0.123	–0.072	0.589
		(0.114)	(0.096)	
PII_credit_card		–0.083	–0.045	0.715
		(0.118)	(0.109)	
Year controls	Y	Y	Y	
Circuit court and region controls		Y	Y	
PII controls		Y	Y	
Industry controls		Y	Y	
Forum controls			Y	
Observations	158	156	154	
Log likelihood	–93.475653	–78.888117	–64.067586	
Pseudo R^2	0.1456	0.2701	0.3991	

NOTE: Standard errors are in parentheses, ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$. Models 1–3a represent average marginal effects of logit regressions with the dependent variable as the probability that a reported data breach lawsuit will settle. Models 1, 2, and 3a control for increasingly more covariates. Model 3b is the same regression model as 3a but coefficients are presented as odds ratios. Independent variables for Cause_disclosure and Cause_hack are presented relative to Cause_lost/stolen. PII refers to "personally identifiable information."

SOURCES: Complaint and docket data were collected from PACER.

Note again that the results are generally robust across all models, with Model 3a providing the best fit for the data and including all covariates. Though not shown, results are also robust to the exclusion of individual years 2005–2010. Robustness checks using alternative specifications of the year of disposition (i.e., the year of the breach and the year the complaint was filed) reveal qualitatively similar results. Further discussions therefore focus on estimations from Model 3a.[49]

[49]In addition, coefficient magnitudes and signs are robust to the inclusion of the 42 public actions.

These results suggest that after controlling for all variables, plaintiff allegations of financial harm are correlated with a 30 percent increase in the probability of settlement (from 52 percent to 68 percent, significant at the 1 percent level), supporting H2a. Similarly, the certification of a class action, as Nagareda (2010) theorizes, increases the probability of settlement by 30 percent (significant at the 1 percent level), supporting H2b. In addition to each being highly statistically significant, these estimates are also large in magnitude and therefore of strong practical significance.

On the other hand, we find that causes of action asserting a violation of a federal statute with statutory damages were not positively correlated with settlement, lending no support for H2c. This finding is surprising given that this hypothesis had a strong theoretical and practical justification: these claims can help shift the burden of proof from the plaintiff having to demonstrate actual harm to the defendant having to prove it did not violate the law. A possible explanation for this result could be that the novelty of federal-statute-based privacy litigation makes it harder for the parties to arrive at a shared understanding of the merits.

Of the breach characteristics, only breaches caused by cyber attacks were found to be positively and significantly correlated with settlement (29 percent, significant at the 1 percent level), relative to lost/stolen hardware. That is, the odds of settling a litigated breach caused by cyber attack are almost 10 times greater relative to a litigated breach caused by lost or stolen hardware. The size of the breach was again not found to be positively correlated with the outcome. This is also somewhat surprising, as one might expect that defendants would be strongly induced to settle due to potentially greater publicity from larger breaches.

Of the types of information compromised, we found that breaches relating to financial and credit card information were negatively correlated with settlement (though not statistically significantly so). It is therefore interesting that while the compromise of financial information appears to lead to more litigation, it does not appear to increase a plaintiff's chance of a settlement. Instead, loss or theft of medical information is most strongly correlated with settlement (31 percent, significant at the 1 percent level).

Overall, despite our relatively small sample size, we are still able to show statistically significant results. Interestingly, while the compromise of financial data and breaches caused by improper disposal/disclosure appeared to drive litigation (Table 2), the compromise of medical data and breaches caused by cyber attack appear to drive settlement (Table 3). Moreover, Figure 6 demonstrates that even without actual harm or class certification, lawsuits still tend to settle about half the time. That is, cases with merit were much more likely to settle—yet, cases without merit still settle about half of the time.

A possible explanation could be that defendants choose to settle for reasons entirely unrelated to the merits of a case. For example, they may be rationally choosing to settle to avoid further litigation costs, publicity, or business distraction. Specifically, defendants may be balancing between the costs of an immediate and "certain" settlement versus a future "uncertain" amount (which includes a settlement award with some probability in addition to legal fees). Nevertheless, a full explanation, we believe, warrants more consideration.

VI. Limitations

The first limitation of this work stems from the lack of observed state data breach lawsuits, which limits our interferences to federal data breach actions only. However, under the Class Action Fairness Act (CAFA) (2005), we are relatively confident that all large class actions (and certainly multistate actions) would, indeed, be either filed in, or removed to, federal court. Conversations with defense attorneys strongly support this intuition. Moreover, the absence of these suits would not bias our regression estimates, as we are, in effect, drawing inferences by pooling nonlitigated and state-litigated breaches. Further, we partially controlled for this effect by estimating Equation (1) using only data breaches compromising greater than 1,000 records and found that our results were robust to even just these breaches.

Another possible limitation concerns the relatively small sample of "success" observations of Equation (1) and Equation (2). An often-cited rule of thumb is that one should observe 10 "success" observations for each parameter being estimated (Peduzzi *et al.* 1996), which may otherwise lead to biased or inefficient parameter estimates. We partially address this concern by first estimating Equation (1) and Equation (2) using only year controls (Model 1 in both Table 2 and Table 3) and thereby fulfilling this rule of thumb. We find that, in general, parameter estimates are robust to subsequent models that control for additional effects. Moreover, any concerns of efficiency would produce a lower bound on any statistical significance. Therefore, our results as presented reflect conservative significance levels.

When estimating the probability that a lawsuit will result in settlement, it may be the case that individuals (through their lawyers) contact the breached firm on their own in hopes of achieving a "back-door" settlement without first filing a formal legal complaint. However, discussions with privacy attorneys suggest that this does not happen with any practical frequency.

Finally, Westlaw and PACER have two limitations. First, as discussed in footnote 14, there is no easy way to verify Westlaw's claim that it possesses complete coverage of the kinds of litigation in which we are interested. Though we have no reason to believe that Westlaw's collection practices will introduce bias in a particular direction, we acknowledge the possibility of such bias, especially earlier in the sample period. Second, we are unable to obtain a complete record of the discovery process from PACER. Hence, it is possible that some pretrial activity occurs that may influence the outcome of a lawsuit and that we are unable to control. For example, discussions of the defendant's liability insurance coverage or IT security practices may occur that potentially drive defendants to settle, whereas otherwise they would not. We argue, however, that the vast majority (if not all) of the defendants in our sample will have basic general liability insurance policies—information that would automatically be discoverable according to Federal Rule of Civil Procedure 26(a)1(A)(iv)—thereby reducing any variation across observations.

VII. Discussion

Recent events concerning breaches of consumer personal information have prompted a flurry of lawsuits by alleged victims of identity theft. These disputes have generated consid-

erable congressional activity concerning the collection, use, and dissemination of personally identifiable consumer health, financial, and behavioral information. But is litigation an effective solution?

Consider both the probability of data breach litigation and settlement. On one hand, the overall federal litigation rate for reported data breaches is only about 4 percent, which may provide comfort to firms (potential defendants) that collect personal information. On the other hand, the settlement rate for all known privately federally-litigated breaches is much higher than one might expect (50 percent), which would alternatively be encouraging to plaintiffs.[50] Moreover, if actual harm (as defined within this article) is indeed an appropriate measure of case merit, then the results presented in Figure 6 and Table 2 may provide some assurance that data breach lawsuits are being appropriately disposed of, on average. That is, those cases that *should* settle (because of the presence of actual harm) *do* settle. In fact, the top left panel of Figure 6 suggests that defendants settle perhaps too often (i.e., in absence of actual financial harm, and therefore case merit).

In regard to settlement awards, we naturally find great variation. After contacting litigating attorneys for the 86 settlements, award details were acquired for 28 of them, with details regarding the remaining cases either held privileged (10 cases) or unknown entirely (48). The mean value of settlements awarded to plaintiffs was about $2,500 per plaintiff (min = $500, max = $15k, $n = 19$), with most awards being a nominal amount of around $500 and often awarded to named plaintiffs only. Attorney fees, on the other hand, were substantially larger, with a mean sum of $1.2m (min = $8k, max = $6.5m, $n = 15$). Importantly, however, settlements may also provide individual redress for identity theft losses and expenses, and *cy pres* awards to research, nonprofits, and charities, which have ranged from $50k to $9.5m.

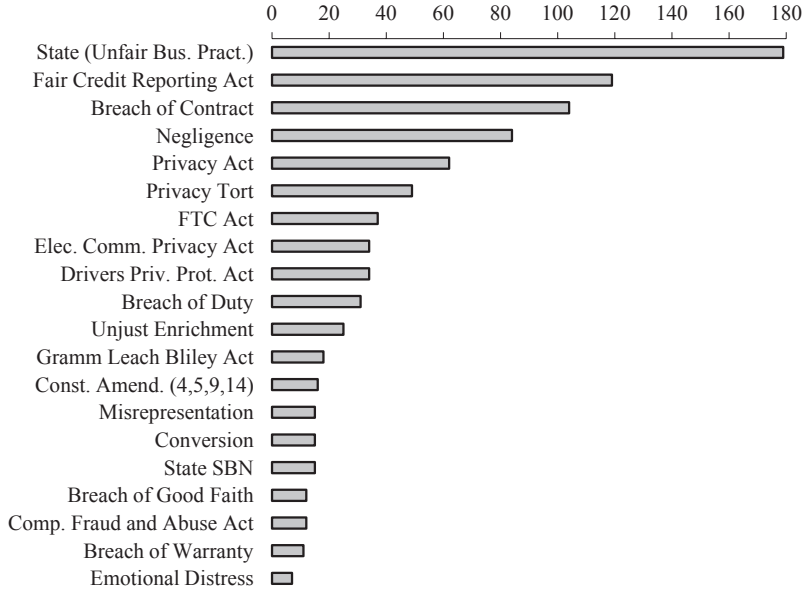
A final observation from this work lies with the diversity of the legal claims brought by plaintiffs. From our data, we identified over 86 unique causes of action (from only 231 cases) for essentially the same event: the unauthorized disclosure of personal information. We found 34 different kinds of tort causes of action, 15 contract, 4 violations of state statutes, and 33 violations of federal statutes. The 20 most common are shown in Figure 7.

What does this suggest about how well equipped the current legal system is in efficiently resolving modern data breach harms? Generally speaking, common law is well suited to address many sorts of injuries in the presence of clear physical, property, or economic loss, but privacy harms differ in important ways: there are many means by which personal information can be collected, used, and distributed, and there are equally as many ways in which consumers may feel emotionally, statistically, or financially harmed. This is no more clearly reflected than in Figure 7. Perhaps precisely because of this complexity, it may be unrealistic to expect a single privacy statute—a single private right of action—to provide redress for all privacy harms. Instead, one might consider a balanced approach that mandates appropriate care for specific kinds of data, together with a clear (federal) privacy statute.

---

[50]Note that inclusion of public actions by the FTC and SEC would, in fact, reveal a higher settlement rate of 62 percent.

Figure 7: The 20 most common causes of action identified from our sample of federal data breach lawsuits.



SOURCE: PACER.

Moreover, data breach litigation can be effective only to the extent that a plaintiff can identify the cause of a breach and subsequent harm. In the case of data brokers (such as Choicepoint, a leading data aggregator), individuals have no direct relationship. This is a market dynamic that has not yet been addressed through litigation, and may therefore justify some regulatory intervention.

VIII. CONCLUSION

The unauthorized disclosure of personal information by firms imposes externalities on consumers from medical and financial fraud and other forms of identity theft. A great deal of recent federal congressional and agency activity has been devoted to reducing the risk of data breaches and crafting legislation to empower consumers to bring federal actions. However, very little is known about the characteristics of data breach litigation and the outcomes of these cases.

This article hopes to address and inform this policy debate by examining two main research questions: Which data breaches are being litigated? and Which data breach lawsuits settle? Our first research question examines federally-litigated breaches resulting from reported data breaches, while the second question includes all known federal lawsuits related to the unauthorized disclosure of personal information.

Our results suggest that the odds of a firm being sued in federal court were 3.5 times *greater* when individuals suffered financial harm, but over 6 times *lower* when the firm involved provided free credit monitoring. Moreover, the odds of a firm being sued from improperly disposing of data were 3 times greater, relative to breaches caused by lost/stolen data, and 6 times greater when the data breach involved the loss of financial information.

Turning to data breach settlements, our results suggest that defendants settle 30 percent more often when plaintiffs allege financial loss from a data breach or when faced with a certified class action suit. Surprisingly, however, plaintiffs seeking statutory damages were not more likely to achieve a settlement. The odds of a settlement were 10 times greater when the breach was caused by a cyber attack, relative to lost or stolen hardware. Although the compromise of financial information led to more litigation, it did not appear to increase a plaintiff's chance of a settlement. Instead, compromise of medical information was most strongly correlated with settlement.

In addition to the regression analyses presented, we performed a number of robustness checks. We verified that the proportions of federally-litigated breaches resulting in actual harm were generally consistent across 2005–2010, as were the causes of litigated breaches. Similarly, we find consistent results when examining the portion of settled cases in which actual harm was alleged, which achieved class certification, and in which plaintiffs sought statutory damages. Further, regression results were robust to probit analysis, and to alternative model specifications, as shown in Table 2 and Table 3.

We also uncovered some novel descriptive data. For example, it seems that only about 4 percent of reported breaches resulted in federal litigation, and that contrary to conventional legal scholarship, the overall settlement rate of known federal lawsuits was around 50 percent. Moreover, it is perhaps staggering that of the federal actions coded, we found over 86 different causes of actions brought by plaintiffs for essentially the same kind of event.

Overall, we believe this research can be of use to various parties. First, it can help provide firms with prescriptive guidance regarding the relative chances of being sued and having to settle. This research could also be useful to insurance markets as a means of assessing a firm's risk and pricing cyber insurance policies. Moreover, we believe that this work can help inform both plaintiff and defense attorneys in better understanding overall trends of data breach litigation. Finally, we hope that our research can inform the policy debate and help create a balanced privacy framework protecting both the interests of consumers who provide personal information and organizations that collect and innovate using this information.

References

Acquisti, A., A. Friedman, & R. Telang (2006) "Is There a Cost to Privacy Breaches? An Event Study," presented at the Fifth Workshop on the Economics of Information Security, Robinson College, University of Cambridge, Cambridge, UK.

Boyd, C., & D. Hoffman (forthcoming) "Litigating Toward Settlement," *J. of Law, Economics, & Organization.*

Bureau of Justice Statistics (BJS) (2011) *Identity Theft Reported by Households, 2005–2010.* Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.

Campbell, K., L. Gordon, L. Loeb, & L. Zhou (2003) "The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market," 11(3) *J. of Computer Security* 431.

Cavusoglu, H., B. Mishra, & S. Raghunathan (2004) "The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers," 9(1) *International J. of Electronic Commerce* 69.

Choi, S. (2007) "Do the Merits Matter Less After the Private Securities Litigation Reform Act?" 23(3) *J. of Law, Economics, & Organization* 598.

Citron, D. (2007) "Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age," 80 *Southern California Law Rev.* 241.

—— (2011) "Mainstreaming Privacy Torts," 99 *California Law Rev.* 101.

Clermont, K., & T. Eisenberg (1998) "Do Case Outcomes Really Reveal Anything About the Legal System? Win Rates and Removal Jurisdiction," 83 *Cornell Law Rev.* 581.

—— (2002) "Litigation Realities," 88 *Cornell Law Rev.* 119.

Cooter, R., & D. Rubinfeld (1989) "Economic Analysis of Legal Disputes and Their Resolution," 27(3) *J. of Economic Literature* 1067.

Cooter, R., & T. Ulen (2008) *Law & Economics*, 5th ed. Pearson Education, Inc.

Cox, J., R. Thomas, & L. Bai (2008) "There Are Plaintiffs and . . . There Are Plaintiffs: An Empirical Analysis of Securities Class Action Settlements," 61 *Vanderbilt Law Rev.* 355.

Department of Commerce (2010) *Commercial Data Privacy and Innovation in the Internet Economy: A Dynamic Policy Framework*. Washington, DC: U.S. Department of Commerce.

Department of Justice (DOJ) (2008) *Incident Response Procedures for Data Breaches Involving Personally Identifiable Information*. Available at http://www.justice.gov/opcl/breach-procedures.pdf.

Dickey, T., D. Ganz, & J. Lever (2011) "Privacy Protection and Data Breaches: HR Tip of the Month," *Blog of the National Law Review*. Available at http://nationallawforum.com/ 2011/04/24/privacy-protection-and-data-breaches-hr-tip-of-the-month/.

Dunbar, F., & F. Sabry (2007) "The Propensity to Sue: Why Do People Seek Legal Actions?" 42(2) *J. of the National Association for Business Economics* 31.

Eisenberg, T., & C. Lanvers (2009) "What is the Settlement Rate and Why Should We Care?" 6(1) *J. of Empirical Legal Studies* 111.

Federal Trade Commission (FTC) (2010) *Protecting Consumer Privacy in an Era of Rapid Change*. Washington, DC: Federal Trade Commission.

Gordon, L. A., M. Loeb, & L. Zhou (2011) "The Impact of Information Security Breaches: Has There Been a Downward Shift in Costs?" 19(1) *J. of Computer Security* 33.

Government Accountability Office (GAO) (2007a) *Data Breaches Are Frequent, But Evidence of Resulting Identity Theft Is Limited; However, the Full Extent Is Unknown*. GAO publication GAO-07-737. Washington, DC: GAO.

—— (2007b) *Privacy: Lessons Learned About Data Breach Notification*. GAO publication GAO-07-657. Washington, DC: GAO.

Heise, M., & G. Sisk (2012) "Ideology 'All the Way Down'? An Empirical Study of Establishment Clause Decisions in the Federal Courts," 110 *Michigan Law Rev.* 1201.

Hensler, D., S. Marquis, A. Abrahamse, S. Berry, P. Ebener, E. Lewis, E. Lind, R. MacCoun, W. Manning, J. Rogowski, & M. Vaiana (1991) *Compensation for Accidental Injuries in the United States*. Santa Monica, CA: RAND Corporation.

Hoffman, D., A. Izenman, & J. R. Lidicker (2007) "Docketology, District Courts, and Doctrine," 85 *Washington Univ. Law Rev.* 681.

Hoffman, L. (2011) "*Twombly* and *Iqbal*'s Measure: An Assessment of the Federal Judicial Center's Study of Motions to Dismiss," 6 *Federal Courts Law Rev.* 1.

Hutchins, J. (2008) "A New Frontier in Privacy Litigation: The Advent of Private Lawsuits Over Data Security Breaches," presented at the 2008 ABA Annual Meeting, Section of Litigation, New York.

Johnson, M., K. Nelson, & A. Prichard (2007) "Do the Merits Matter More? The Impact of the Private Securities Litigation Reform Act," 23(3) *J. of Law, Economics, & Organization* 627.

Kannan, K., J. Rees, & S. Sridhar (2007) "Market Reactions to Information Security Breach Announcements," 12(1) *International J. of Electronic Commerce* 69.

Kim, P., M. Schlanger, C. Boyd, & A. Martin (2009) "How Should We Study District Judge Decision-Making?" 29(83) *J. of Law & Policy* 83.

Lerner, J. (2010) "The Litigation of Financial Innovations," 53(4) *J. of Law & Economics* 807.

Lesemann, D. J. (2012) "Once More unto the Breach: An Analysis of Legal, Technological and Policy Issues Involving Data Breach Notification Statutes," 4 *Akron Intellectual Property J.* 203.

Mammem, C. (2009) "Controlling the 'Plague': Reforming the Doctrine of Inequitable Conduct," 24 *Berkeley Technology Law J.* 1329.

McShane, B. B., O. P. Watson, T. Baker, & S. J. Griffith (2012) "Predicting Securities Fraud Settlements and Amounts: A Hierarchical Bayesian Model of Federal Securities Class Action Lawsuits," 9(3) *J. of Empirical Legal Studies* 482.

Miller, R. E., & A. Sarat (1980) "Grievances, Claims, and Disputes: Assessing Adversary Culture," 15 *Law & Society Rev.* 525.

Nagareda, R. A. (2010) "Common Answers for Class Certification," 63 *Vanderbilt Law Rev. En Banc* 149.

Paray, P. E. (2011) "Elephant in the Room–The Potential for Data Breach Statutory Damages," 2(3) *Information Security & Privacy News*, Information Security Committee, ABA Section of Science & Technology Law.

Peduzzi, P., J. Concato, E. Kemper, T. R. Holford, & A. R. Feinstein (1996) "A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis," 49 *J. of Clinical Epidemiology* 1373.

Priest, G., & B. Klein (1984) "The Selection of Disputes for Litigation," 13 *J. of Legal Studies* 1.

Rice, D. (2007) *Civil Actions for Privacy Violations 2007: Where Are We?* Howard Rice Nemerovski Canady Falk & Rabkin.

Romanosky, S., & A. Acquisti (2009) "Privacy Costs and Personal Data Protection: Economic and Legal Perspectives of Ex Ante Regulation, Ex Post Liability and Information Disclosure," 24(3) *Berkeley Technology Law J.* 1061.

Romanosky, S., R. Sharp, & A. Acquisti (2010) "Data Breaches and Identity Theft: When is Mandatory Disclosure Optimal?" paper presented at the 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?" 30(2) *J. of Policy Analysis & Management* 256.

Rubin, D. B. (1990) "Formal Modes of Statistical Inference for Causal Effects," 25 *J. of Statistical Planning & Inference* 279.

Ryan, E. (nd) *States Offer Data Breach Protection.* National Association of Attorneys General. Available at http://www.naag.org/states-offer-data-breach-protection.php.

Serwin, A. (2009) "Poised on the Precipice: A Critical Examination of Privacy Litigation," 25(4) *Santa Clara Computer & High Technology Law J.* 883.

Shavell, S. (1996) "Any Frequency of Plaintiff Victory at Trial Is Possible," 25 *J. of Legal Studies* 493.

Solove, D. (2008) "The New Vulnerability: Data Security and Personal Information," in A. Chander et al., eds., *Securing Privacy in the Internet Age*, pp. 111, 115–16. Palo Alto, CA: Stanford Univ. Press.

—— (2010) "Are People Really Harmed By a Data Security Breach?" *Concurring Opinions Legal Blog.* Available at http://www.concurringopinions.com/archives/2010/09/are-people-really-harmed-by-a-data-security-breach.html.

Symantec (2012) *Symantec Intelligence Report: August 2012.* Available at http://www.symantec.com/content/en/us/enterprise/other_resources/b-intelligence_report_08_2012.en-us.pdf.

Verizon Communications Inc. (2010) *2010 Data Breach Investigations Report.* Verizon Communications Inc.

Viscusi, W. K. (1986) "The Determinants of the Disposition of Product Liability Claims and Compensation for Bodily Injury," 15(2) *J. of Legal Studies* 321.

Wittman, D. (1988) "Dispute Resolution, Bargaining, and the Selection of Cases for Trial: A Study of the Generation of Biased and Unbiased Data," 17(2) *J. of Legal Studies* 313.

Wolf, C. (2012) *Introduction to Data Security Breach Preparedness with Model Data Security Breach Preparedness Guide.* Hogan Lovells US LLP.