Cyentia INSTITUTE

Information Risk Insights Study

A Clearer Vision for Assessing the Risk of Cyber Incidents

IRIS 2020

In partnership with Advisen Transforming • Insurance

Introduction

"You got to know when to hold 'em
Know when to fold 'em
Know when to walk away
Know when to run."

—Kenny Rogers, "The Gambler"

There are surprisingly few songs about risk-based decision making and the knowledge required to do it well. "The Gambler" is one of them, and we've probably earwormed more than one reader with that catchy refrain. But have you ever noticed how The Gambler comes to know the things he needs to know? According to the lyrics, he stakes it all on "the way they held their eyes." If that sounds like a less-than-ideal basis for risky decisions to you, we agree. Perhaps that's why he later confesses "the best that you can hope for is to die in your sleep."

We're writing this report to help cyber risk takers avoid The Gambler's fate of futility. The cards might indeed be stacked against defenders, and adversaries have grown adept at hiding their tells. But there are ways to improve the odds of winning. In short, those ways involve leveraging better data to gain better knowledge to build better models that ultimately lead to better decisions for successfully managing cyber risk.

This report links together that chain of "better," starting with a vast dataset spanning tens of thousands of cyber loss events over the last decade. Our analysis of those events yields important lessons—and baseline model inputs—about the frequency and impact of breaches to organizations of all types and sizes. We've included some of those findings on the next page, but they're just a taste of what's in store in the pages that follow.

Are you ready to make cyber risk less of a gamble? Excellent! We are too. Let's do this.

Table of Contents

The Cyentia Institute is a research and data science firm working to advance cybersecurity knowledge and practice. We do this by partnering with vendors and other organizations to publish a range of high-quality, data-driven content like this study.

Advisen is the leading provider of data, media, and technology solutions for the commercial property and casualty insurance market. Advisen's proprietary data sets and applications focus on large, specialty risks.

IRIS 20/20 **Key Findings**

60% Over ***60%*** of the Fortune 1000 had at least one cyber incident over the last decade. On an annual basis, we estimate one in four Fortune 1000 firms will suffer a loss event.

2% Moving beyond mega-corporations, the probability of incidents drops substantially. SMBs have rates below ***2%*** and are orders of magnitude less likely to suffer several breaches in a year.

30X The likelihood of incidents varies up to ***30x*** by industry. Government agencies, administrative support, information services, and financial firms, have the highest rates.

$1.7T The traditional method of estimating breach losses—using a flat cost per record—is flat-out wrong. It results in ***$1.7 trillion*** in error from overestimating losses. We offer a better option.

50% We can use the number of records breached to estimate losses, but it's probabilistic rather than deterministic. An exposure of 1,000 records has a 6% chance of exceeding $10M. By comparison, an exposure of 100M records has a better than ***50%*** chance of racking up at least $10M in losses.

$20M Financial losses following a cyber event typically run about $200K, but 10% of them exceed ***$20M***. The cost of extreme events (95th percentile) to the mega corporations in the Fortune 250 approaches $100M (or more).

10X Typical and extreme losses differ greatly among industries. The information services and retail sectors show abnormally high losses that exceed many other sectors by a ***factor of 10***.

25% A $100B enterprise that experiences a typical cyber event should expect a cost that represents 0.000003% of annual revenues. A mom and pop shop that brings in $100K per year, on the other hand, will likely lose ***one-quarter*** of their earnings or more.

6% Based on these frequency and loss estimations, we assess that there's a ***6%*** chance that a Fortune 1000 firm will lose $100M or more in a 12-month period due to cyber events. These are the type of probabilistic cyber risk projections we're aiming to support in this study.

Like what you see? Join the vision!

We intend to continue the IRIS in the future to discover even more insights for managing information risk. If you'd like to join in that effort by contributing relevant data or sponsoring, please reach out to us at research@cyentia.com.

119
Cy
cyentia

What is "Information Risk?"

The 2020 Information Risk Insights Study (IRIS 20/20) aims to clear the fog of FUD (fear, uncertainty, and doubt) surrounding information (aka cyber) risk and help managers see their way to better data-driven decisions. Since we suspect this study will be read by audiences from different backgrounds with different working definitions of "risk," we thought it necessary to make sure we're all on the same page.

A quick web search will find many definitions of risk, and we're not going to attempt to pick just one or proffer yet another. While definitional variations abound, most agree at some level that risk involves the frequency and impact of adverse events. Thus, information risk deals with the occurrence and cost of events that adversely affect information systems.

Unfortunately, reliable data about the frequency and impact of cyber incidents has been historically difficult to obtain. This lack of data presents a serious challenge for decision makers, causing many to fall back on subjective judgements and qualitative ratings. We know that struggle well and that's why we're so excited about the IRIS. Our extensive analysis yields objective data on the frequency and financial impact of cybersecurity breaches[1] to organizations of all types and sizes. We hope it helps many escape the qualitative quagmire of information risk assessments.

"Our extensive analysis yields objective data on the frequency and financial impact of breaches to organizations of all types and sizes. We hope it helps many escape the qualitative quagmire of information risk assessments.

About the data used in this study

This first-of-its-kind study leverages a vast dataset spanning 56,000 cyber events experienced by 35,000 organizations over the last decade. That dataset comes courtesy of Advisen's Cyber Loss Data, which contains nearly 100,000 cyber events collected from publicly verifiable sources. This dataset is widely used, with three features that make it ideal for this research: 1) It is the most comprehensive list of historical cyber incidents we've found. 2) It tracks losses publicly disclosed in the wake of those incidents. And 3) it includes supplemental firmographic information on the organizations affected by cyber events and the broader economy. Additional information about Advisen's Cyber Loss Data and our analysis of it can be found in the Methodology & Firmographics section.

Find out more: www.advisenltd.com/data/cyber-loss-data

[1]We've chosen to use the term "breach" in this study as our standard way of broadly referring to adverse events that impact the confidentiality, integrity, or availability of a firm's information assets. To add some variety, we also interchangeably use terms like "incidents," "loss events," and "cyber events." This terminology encompasses common events such as data theft, ransomware infections, DDoS attacks, lost or stolen laptops.

Event Frequency Analysis

"What's the frequency, Kenneth?"

—R.E.M.

In our journey to better assess the risk posed by cybersecurity breaches, we first explore how often they occur. Our initial step is to provide some high-level parameters for the number of breaches reported by firms over the time frame of our study. We then estimate the likelihood that a given organization will experience a breach within a twelve-month period. Our ultimate goal is to derive an annualized frequency distribution for the number of breaches various types and sizes of firms should expect. Let's get started.

> **From this, we can infer two simple but important facts. First, some firms had multiple breaches during the timeframe. Second, breaches are not an annual event for most firms.**

Some Priors on Breach Probability

Before getting deep into breach probabilities, it's worth taking some baby steps by examining the historical frequency of breaches. Our dataset includes more than 56K breaches across 35K organizations over 10 years. From that, we can infer two simple but important facts. First, some firms had multiple breaches during the timeframe. Second, cyber events are not an annual event for most firms. We'll expand on the first point now and pick up the second in the next section.

Figure 1: Distribution of the number of publicly-known breaches per firm

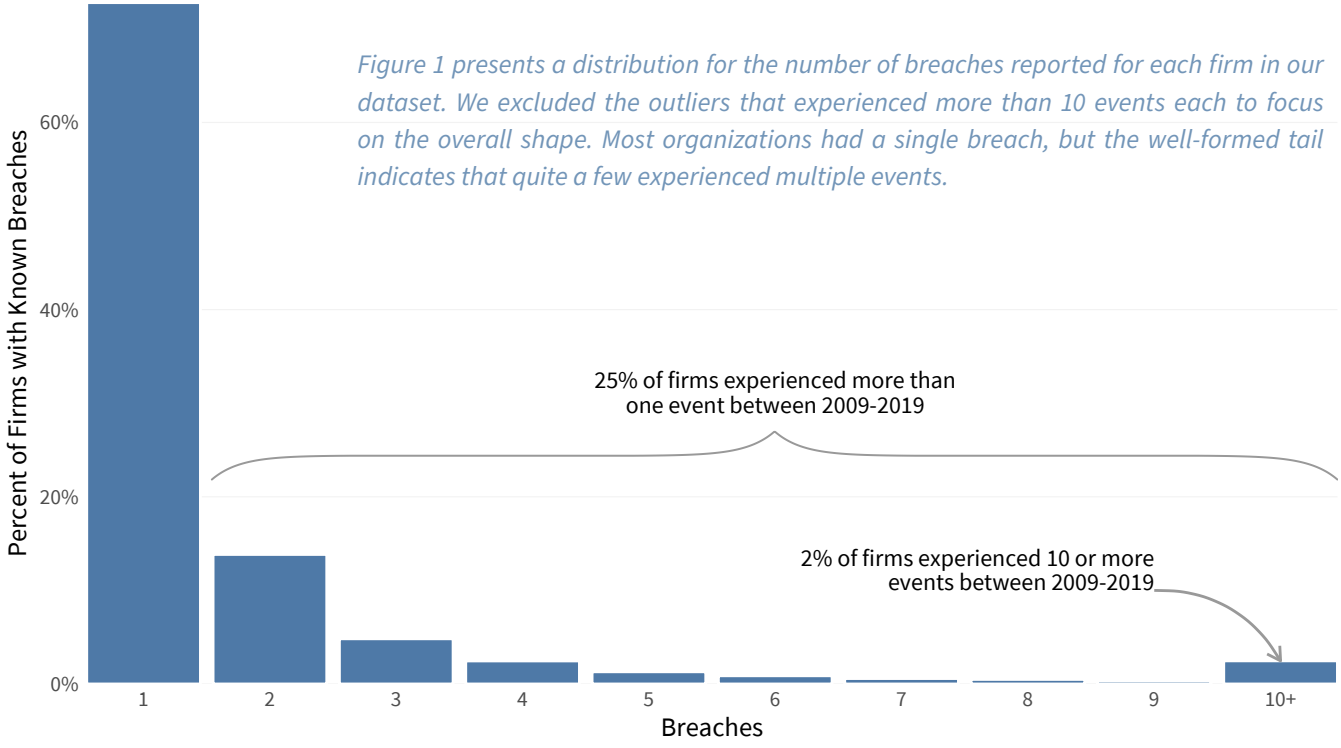


Figure 1 presents a distribution for the number of breaches reported for each firm in our dataset. We excluded the outliers that experienced more than 10 events each to focus on the overall shape. Most organizations had a single breach, but the well-formed tail indicates that quite a few experienced multiple events.

Keep in mind that all organizations in this study had at least one known breach—otherwise they wouldn't be in our dataset. Per Figure 1, most of those kept it to just the one. About 25% of firms went on to have another incident, and just over 2% experienced ten or more during our study period. Those facts alone serve as a good reminder that most breaches aren't the rare business-ending disasters many fear them to be. Indeed, some seem to have made "once more unto the breach, dear friends!" their company motto.

While that helps shape our assumptions about the raw count of breaches, we cannot use it to determine the probability that any particular firm will experience an incident. To do that, we need a reliable estimate of the total number of firms that exist in the segment of the population we're trying to measure. We take a step in that direction in the next section.

Breaches in the Fortune 1000

Calculating the probability of a firm experiencing a breach requires that we have a reliable estimate of the total number of firms operating in our target population. Obtaining that on a global basis, where firms constantly appear/disappear and the availability/validity of statistics varies from country to country, is difficult to say the least. As a proxy, the Fortune 1000 offers a very well-defined population of organizations to study. Later we will apply the techniques introduced here to the larger population beyond the Fortune 1000.

Every year, Fortune Magazine publishes a list of the largest 1,000 US companies as measured by their annual revenues.[2] Commonly referred to subsets of this list include the Fortune 500 (the top 500 firms) and the Fortune 100 (you guessed it—the top 100). Companies slide up and down the rankings each year and enter and exit regularly. To have a static set of organizations for our analysis, we use the Fortune 1000 as published in 2019.

> By using the Fortune 1000, we are in no way implying that U.S. companies are more important or that cyber events occur only in the U.S. The Fortune 1000 aligns well with our breach dataset, which is most comprehensive for the U.S.

In addition to providing a known set of firms, another perk of using the Fortune 1000 is that breaches involving them are more likely to become publicly known (and thus recorded by Advisen or others). These corporate giants make up less than 0.08% of all US firms, employ more than a quarter of the US work-force and account for 20% of all the breaches in our data set. That's 250 times more events than we would expect based on the number of companies alone!

Figure 2: Percent of Fortune 1000 with known breaches

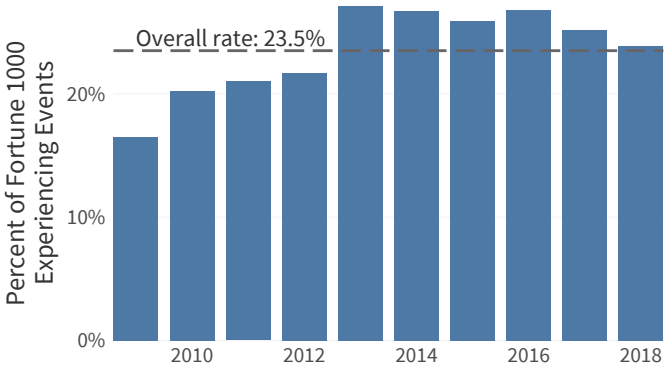


Figure 2 pegs the annual percentage of Fortune 1000 firms with known breaches at 24%. That rate rose steadily for the first five years and leveled off thereafter.

Over 60% of the Fortune 1000 had at least one breach during the decade we examined. As seen in Figure 2, the percent of Fortune 1000 firms with known incidents each year averaged 24%. That rate varied somewhat over time, rising steadily for the first five years and then leveling off thereafter.

[2] *The list is officially titled the Fortune 500, but includes the top 1000 organizations (https://www.fortune.com/fortune500/).*

Breach likelihood in the Fortune 1000

Looking at breach rates across prior years is interesting for establishing trends, but we've yet to see an organization make risk decisions according to a strict Gregorian calendar. It's more common to ask something like "what's the chance we'll have a breach in the next 12 months?" Answering questions like that for the Fortune 1000 is where we're off to next.

Rather than fixed, calendar-based timeframes, we now employ a rolling 12-month window to estimate breach likelihood. To construct those windows, we examine the period of January 2010 to December 2010, February 2010 to January 2011, March 2010 to February 2011, and so on. Doing this smooths out spikes in the data and boosts the confidence of our estimates, creating a better "window" (see what we did there?) into long-term trends.

Some readers may be forming "buts" about the perils of predicting the future based on the past. An argument using the word "ergodic" is probably being formulated. Someone's thinking about the 2008 financial crisis or the COVID-19 pandemic. We get all that. View these numbers as helpful starting points for risk models—informed priors for our Bayesian friends—rather than absolute predictions about an uncertain future.

Using this approach, we calculated the proportion of the Fortune 1000 that had at least one breach within each 12-month rolling window. Figure 3 shows how the rate changes over these rolling "years" and ranges from 18% to 28%. Similar to Figure 2, incident likelihood rises through 2013 and then transitions to a gradual decline. The overall likelihood of 25% is based on the average across all of these periods. Thus, we expect approximately 1 in 4 Fortune 1000 firms to have at least one publicly attributable breach in a one-year timeframe.

Figure 3: Overall annual breach likelihood among Fortune 1000 firms using a 12-month rolling window

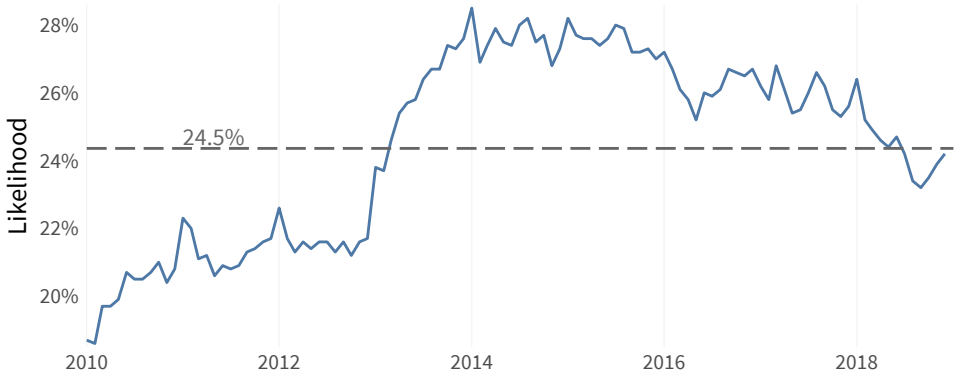


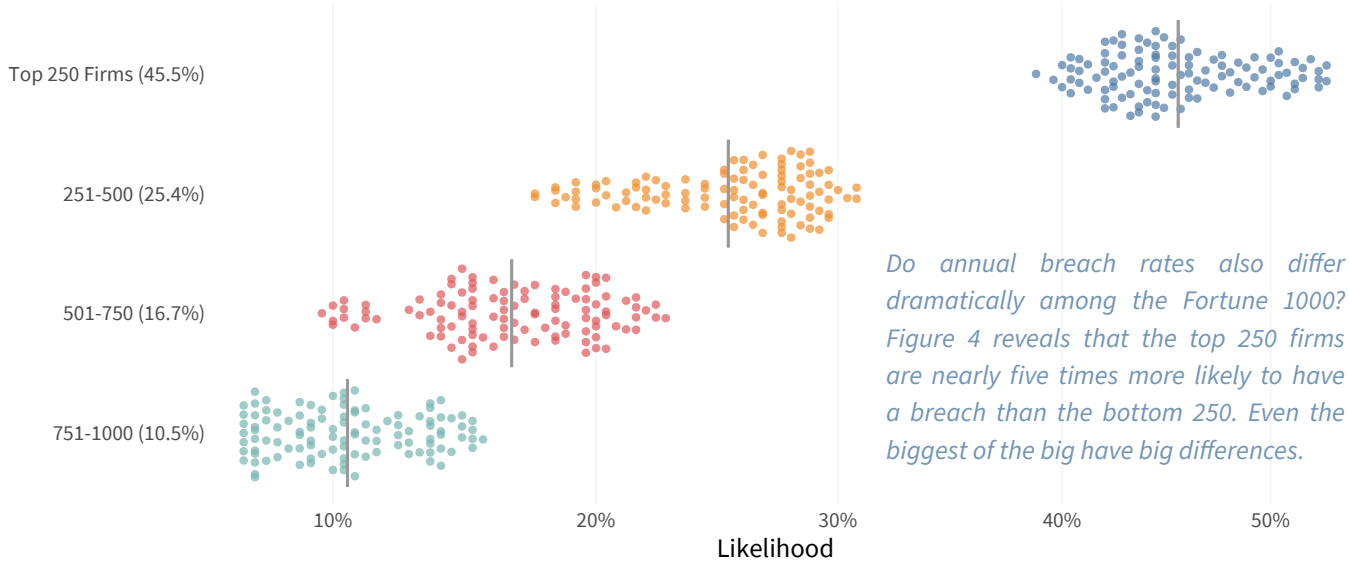
Figure 3 shows breach rates among the Fortune 1000 over a rolling one-year window. Overall, we estimate the annual likelihood of any given firm on that list having a breach at 25%.

Though the Fortune 1000 are, by definition, the biggest of the big, it would be a mistake to consider them all as equally big. Annual revenues for the largest on the list dwarf the smallest among them by a factor of 250. Do annual breach rates also differ dramatically among them? Figure 4 reveals the answer. (Hint: it's a solid "yes.")

> **"It's common to ask something like "what's the chance we'll have a breach in the next 12 months?" Answering questions like that is where we're off to next.**

Figure 4 applies the same methods used in Figure 3, except that it compares breach rates among ranked quartiles of the Fortune 1000. Each dot represents a rate observed for one of the rolling 12-month windows we constructed based on the event data, and the gray bar marks the average among them. The top 250 firms in the Fortune 1000 are five times more likely to have a breach than the bottom 250. Differences aren't quite so dramatic for the second and third quartiles, but the clear trend remains. The bigger they are, the more likely they fall prey to cyber incidents.

Figure 4: Comparison of annual breach likelihood among quartiles of Fortune 1000 firms



Do annual breach rates also differ dramatically among the Fortune 1000? Figure 4 reveals that the top 250 firms are nearly five times more likely to have a breach than the bottom 250. Even the biggest of the big have big differences.

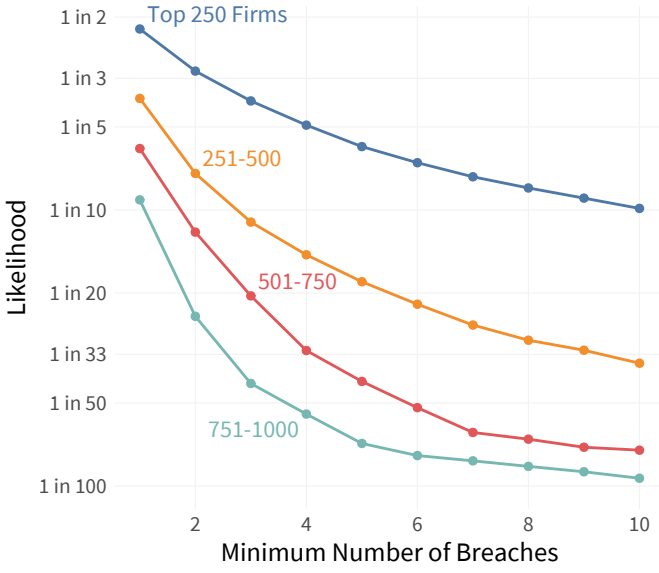
Breach Frequency in the Fortune 1000

Since we've already established that some organizations suffer multiple incidents in a single year, one might assume that holds true for the Fortune 1000 as well. Allow us to confirm and then expand on that assumption. In 2005, 5% of firms in the Fortune 1000 had more than one breach. That peaked a decade later, with 20% registering multiple breaches in 2015. The multi-breach rate has stabilized and even declined since then.

That's nice to know, but what we really need to know for proper risk modeling is exactly how many cyber events an organization should expect over the course of a year. As you can see in Figure 5, that number varies quite a bit depending on where firms sit within the Fortune 1000.

Figure 5 depicts the chances of firms in different quartiles of the Fortune 1000 having from one to 10 breaches in a single year. The leftmost estimates for one incident mirror those in Figure 4. From there, tracing the likelihood of successive numbers of breaches is fairly straightforward. For instance, there's about a 20% chance that one of the top 250 organizations will report four or more cyber loss events. By comparison, the likelihood of the bottom tier suffering that many breaches drops to just under 2%.

Figure 5: Comparison of annual breach frequency among quartiles of Fortune 1000 firms



The shape of the curves in Figure 5 is also interesting. Notice how they become steeper and more convex while moving from top to bottom of the Fortune 1000. This means the largest organizations are not only more likely to have a breach; they're also much more likely to have larger numbers of breaches. Breach frequencies for firms in the lowest tier of the Fortune 1000 drop quickly at first but then level off. They have about the same chance of having six incidents as having 10.

While helpful in visually communicating the concept, the format of Figure 5 makes it hard on readers wanting to interpolate precise data points to drive risk models. For those overachieving readers, we offer the more extractable format of Table 1. From now on, we'll skip the frequency curves and jump straight to the tables.

Fortune 1000 Group	Number of Incidents					
	1 or more	2 or more	3 or more	4 or more	5 or more	10 or more
Top 250	1 in 2	1 in 3	1 in 4	1 in 5	1 in 6	1 in 10
251-500	1 in 4	1 in 7	1 in 11	1 in 15	1 in 18	1 in 36
501-750	1 in 6	1 in 12	1 in 20	1 in 32	1 in 42	1 in 74
751-1000	1 in 9	1 in 24	1 in 43	1 in 55	1 in 70	1 in 94

Table 1: Likelihood of a Fortune 1000 firm experiencing a certain number of breaches

And with that, we now have a repeatable way of estimating the frequency of breaches for a given firm across segments of interest (e.g., Fortune 1000 tiers) within a specific timeframe. Now let's leave the exclusive district of mega corporations behind and turn our attention to the 99.9% of organizations outside the Fortune 1000. Get ready for a culture shock!

FAIR Use: Those familiar with Factor Analysis of Information Risk (FAIR™) may see Figure 5 and Table 1 as Loss Event Frequency (LEF) estimates. Baseline inputs for LEF by industry and size can be found throughout this section.

Breach Frequency by Sector

Having measured the frequency of breaches in the relatively tidy world of the Fortune 1000, we can now expand our vision to all sectors of the economy.[3] Things get a little more tricky here because we do not have a definite count of organizations in each sector. Thankfully, our Advisen dataset also includes the number of registered businesses by size and sector. It provides a reasonably accurate denominator for our purposes here.[4]

> **Now let's leave the exclusive district of mega corporations behind and turn attention to the 99.9% of organizations outside the Fortune 1000.**

[3]We focus on U.S. firms because breach notification laws have helped to grow a large set of historical breach records that, while not exhaustive, is the best available for our purposes here. It should also be noted that we excluded organizations with less than $1M in revenue from this analysis because they're far more likely to fall below the radar of breach reporting and collection efforts.

[4]We've compared the Advisen industry statistics feed (in part derived from Dun & Bradstreet information) with data from the U.S. Census Bureau and confirmed the overall accuracy of these numbers.

Figure 6 follows the same process and format of Figure 4, except we now compare top-level sectors using the North American Industry Classification System (NAICS). Once again, dots show the rates calculated for each rolling one-year window, and the vertical gray bars mark the overall likelihood for each sector.

Figure 6: Comparison of annual breach likelihood among firms by sector

A couple key observations jump out here. First, the overall rates are very small. All but the public sector show less than a 1% annual probability that any given organization will experience a breach. This is potentially confusing, especially given the much larger likelihoods we saw for the Fortune 1000. After all, don't the sectors shown here include those exact same Fortune 1000 firms? Yes, they do. We view these as very conservative, lower bound estimates due to (possibly substantial) under-reporting of incidents. Beyond that, think of this as the base likelihood estimate you'd give if you knew nothing else about an organization other than industry. If you also learned that a financial firm was ranked in the Fortune 100, you'd revise that estimate up substantially. We'll examine the effect of size on breach probability in a moment, but let's first make a few more observations regarding Figure 6.

Circling back to the comparatively high breach likelihood shown for the public sector, some explanation is warranted. It's tempting to conclude that government agencies must be way more vulnerable (or attacked) than all other organizations. But stricter breach reporting requirements and transparency efforts are a more likely explanation. Supporting this hypothesis, Verizon's Data Breach Investigations Report consistently shows a much larger number of incidents for the public sector since US-CERT began contributing statistics they collect from agencies.

> **We view these as conservative, lower-bound estimates due to under-reporting of incidents. Beyond that, think of this as the base likelihood estimate you'd give if you knew nothing else about an organization other than industry.**

Moving beyond public offices, the administrative, information, financial, and management sectors exhibit the highest annual breach rates per firm. The first two represent rich targets for data and money-hungry cybercriminals and the latter serves as the poster child of a target rich environment. Notice also how some sectors show wide ranging estimates (e.g., management), while others bunch tightly around the mean (e.g., construction). The degree of spread may indicate the diversity of risk factors among firms in each sector. Peruse the remaining sectors in Figure 6 as you see fit; we're continuing on to frequency-based estimates in Table 2.

Sector	Number of Incidents					
	1 or more	2 or more	3 or more	4 or more	5 or more	10 or more
Administrative	1 in 123	1 in 435	1 in 839	1 in 1.4k	1 in 2k	1 in 5.9k
Education	1 in 185	1 in 940	1 in 2.6k	1 in 5.8k	1 in 10k	1 in 31k
Entertainment	1 in 438	1 in 2.2k	1 in 5.5k	1 in 7k	1 in 7.3k	1 in 8.4k
Financial	1 in 128	1 in 363	1 in 602	1 in 799	1 in 994	1 in 1.9k
Healthcare	1 in 249	1 in 1.3k	1 in 3.5k	1 in 6.3k	1 in 8.6k	1 in 19k
Hospitality	1 in 310	1 in 1.9k	1 in 6.8k	1 in 13k	1 in 18k	1 in 36k
Information	1 in 126	1 in 400	1 in 694	1 in 1k	1 in 1.3k	1 in 2.1k
Management	1 in 142	1 in 503	1 in 976	1 in 1.7k	1 in 2.1k	1 in 3k
Manufacturing	1 in 977	1 in 4.3k	1 in 12k	1 in 23k	1 in 38k	1 in 78k
Other	1 in 610	1 in 4.2k	1 in 12k	1 in 23k	1 in 29k	1 in 49k
Professional	1 in 326	1 in 1.8k	1 in 5.4k	1 in 11k	1 in 19k	1 in 65k
Public	1 in 18	1 in 66	1 in 124	1 in 182	1 in 235	1 in 392
Real Estate	1 in 742	1 in 3.8k	1 in 8.8k	1 in 13k	1 in 16k	1 in 23k
Retail	1 in 763	1 in 2.5k	1 in 4.3k	1 in 7.9k	1 in 10k	1 in 22k
Trade	1 in 1.7k	1 in 8.2k	1 in 20k	1 in 34k	1 in 44k	1 in 72k
Transportation	1 in 1.2k	1 in 4.9k	1 in 10k	1 in 15k	1 in 19k	1 in 25k
Utilities	1 in 412	1 in 1.7k	1 in 2.4k	1 in 2.9k	1 in 3.2k	1 in 3.5k

Table 2: Number of breaches per firm across sectors

The public sector dominates all others for the number of breaches experienced. Other hot spots for event frequency include the financial, information, and management sectors.

Table 2 summarizes the probability that firms in each sector will experience a certain number of breaches in a year. It's clear the expected frequency of incidents varies substantially across industries. For example, the chance that a given financial firm will suffer 10 or more breaches is nearly 12x that of a retailer and over 41x that of the typical manufacturer.

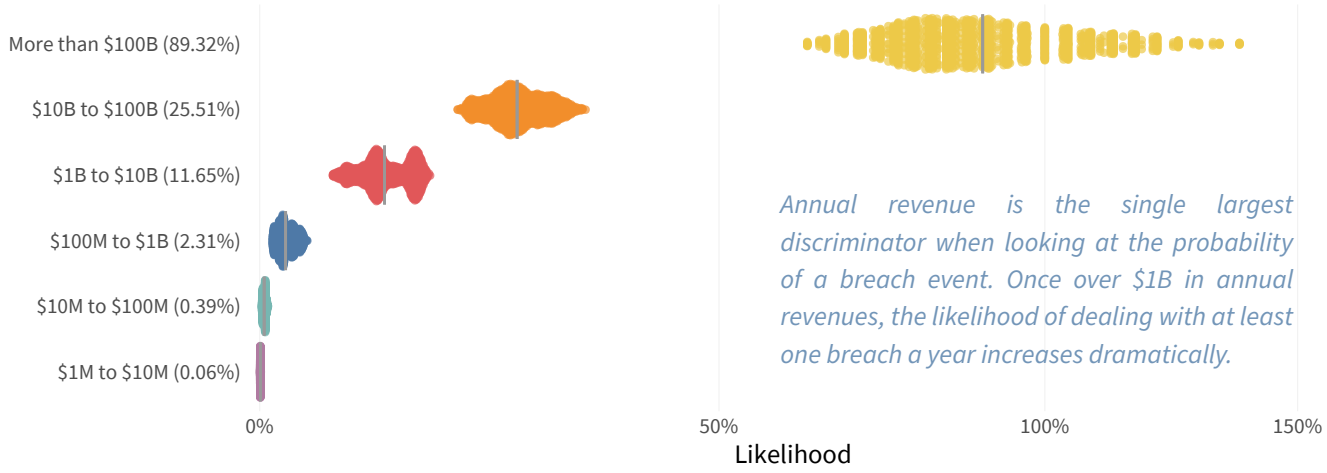
There's a lot packed into Table 2, and we could call out any number of other interesting observations. But we'll leave you to choose your own adventure instead. Do keep in mind that industry classifications, regulation, reporting requirements, and data collection all have major effects on these results. Certain industries undoubtedly have more reliable estimates than others based on a variety of reasons. While we acknowledge that many managers are looking to compare their firms with others based on industry, we caution against using these numbers as a single stand-alone estimate. We're not convinced that a firm's sector is the most valuable signal in determining likelihood of breach and hope to delve into this topic further in future work.

> **Moving beyond public offices, the information, financial, and education sectors exhibit the highest annual breach rates per firm. The first two represent rich targets and the latter serves as the poster child of a target rich environment.**

Breach Frequency by Revenue

Sector-based estimates are interesting, but analysis of the Fortune 1000 taught us that organizational size has a major effect on breach estimates. Figure 7 expands on that analysis by comparing incident likelihood among firms grouped by annual revenues. As with similar charts above, dots represent estimates from successive one-year windows and bars mark the overall likelihood.

Figure 7: Comparison of annual breach likelihood among firms by revenue



It's clear right away that size really does matter when it comes to incident likelihood. Firms under $1B in annual revenues (where most organizations live) have less than a 2% chance of experiencing a breach in a given year. Beyond that, rates rise quickly from 9.6% ($1B to $10B) to 22.6% ($10B to $100B) to just over 75% for the exclusive $100 billion dollar club. Those largest enterprises are over 1,000 times more likely to report a breach than small (<$10M) businesses.[5]

[5]*This phrasing applies to all our analysis but seems particularly important to note here. It's conceivable that large firms aren't any more likely to have a breach; their breaches are just far more likely to become publicly known.*

Not only are the biggest firms much more likely to have a breach, but Table 3 shows their expected number of events in a year is much larger too. It's exceedingly rare for SMBs, on the other hand, to experience multi-incident years. The chance of a $100B+ enterprise reporting 10+ breaches is 3.5 orders of magnitude higher than a <$100M firm. We saw zero examples of firms with annual revenues under $10M reporting incidents in the double digits.

	Number of Incidents					
Revenue Category	1 or more	2 or more	3 or more	4 or more	5 or more	10 or more
More than $100B	8 in 9	3 in 4	2 in 3	5 in 8	4 in 7	2 in 5
$10B to $100B	1 in 4	1 in 6	1 in 9	1 in 11	1 in 14	1 in 30
$1B to $10B	1 in 8	1 in 19	1 in 34	1 in 52	1 in 72	1 in 156
$100M to $1B	1 in 42	1 in 157	1 in 343	1 in 579	1 in 798	1 in 1.8k
$10M to $100M	1 in 252	1 in 1k	1 in 2.2k	1 in 3.6k	1 in 5.1k	1 in 14k
$1M to $10M	1 in 1.6k	1 in 9.1k	1 in 27k	1 in 56k	1 in 97k	1 in 279k

Table 3: Number of breaches per company by size (revenue)

All in all, it appears that each ten-fold increase in a firm's revenue brings approximately a ten-fold increase in the frequency of cyber events. It seems too easy to the point of being cliché to quote Biggie[6] on this one, but we'll do it anyway: "It's like the more money we come across, the more problems we see." Now let's see exactly how much money those problems actually cost.

"All in all, it appears that each ten-fold increase in a firm's revenue brings approximately a ten-fold increase in the frequency of cyber events.

Looking for more specific data on event frequency?

We understand (and hope) that readers will likely be left wanting more after reading this section. Some may wish to see estimates for their specific sub-sector in NAICS (i.e., banks and insurance carries both fall under the financial sector). Some will want to combine sector and size estimates to, for example, get the expected rate of breaches for financial firms with revenues between $1B and $10B. Other interesting questions and directions abound. Unfortunately, we can't possibly include all those views in this report. But you're not wholly out of luck, because that's exactly what Advisen's cyber loss data is designed to provide. If you're longing to slice and dice data on cyber events at will, then you should definitely check that out.

[6] *Before you "Well, AACCTTUUALY..." us for attributing this to Notorious B.I.G.—yes—we're aware this refrain is actually sung by Kelly Price. It's Biggie's song. Now, who's hot, who's not?*

Loss Magnitude Analysis

"There's a bit of magic in everything.
And then some loss to even things out."

—Lou Reed, "Magic and Loss"

Now that we've looked at how often breaches occur across several firmographic dimensions, we'll switch over to the financial losses incurred when a firm is unfortunate enough to experience a cyber event. Our first task is defining how losses are distributed and then to establish the cost of a typical event. We also challenge the popular approach of using a flat cost per record metric to estimate losses. The section ends with loss estimates by industry and size groups.

While reading this section, keep in mind that not all losses for all incidents become public. Certain types of losses are easier to identify from public records, such as court filings, SEC filings, etc. We suspect the losses from highly public, major events are more complete than those from minor events due to increased scrutiny and public records (e.g., lawsuits, corporate filings, etc.).

With those caveats in mind, there is no reason to believe loss data collected in this manner is any less accurate than loss data collected via survey. Asking someone "How much did your breach cost?" is subject to similar (and even more) issues:

- The respondent may not know all losses
- All losses may not have yet come to pass
- Extreme losses are more memorable
- Intangible costs may not be known, quantified, or included
- Cognitive biases lead to over- or underestimation

Thus, we hold that our sample of losses suitably reflects "known losses" from cyber incidents. If biased in any direction, we suspect it's toward larger breaches. Smaller, minor, or otherwise less costly breaches that don't require disclosure, don't drive headlines, and aren't material enough for annual reports are less likely to be in this dataset. It's also a much larger and more comprehensive sample of breaches than any known loss survey or study.

The Shape of Cyber Loss Events

We need to make a simple yet important point at the outset of this section. While any single breach will have a specific loss magnitude, our purpose here is to evaluate losses from many breaches. No single value can do this sufficiently because of the wide variation that exists in reported losses—even among breaches of similar type and extent. Instead, we need to define a range of values that describe losses we observe across those breaches.

Let's start by creating a distribution based on historical losses from breaches in our dataset over the last 10 years. That's done in Figure 8. The first thing to note about any distribution is the basic shape. The shape sets the rules for what kind of statistics are valid for analyzing and describing it. (More on that later.)

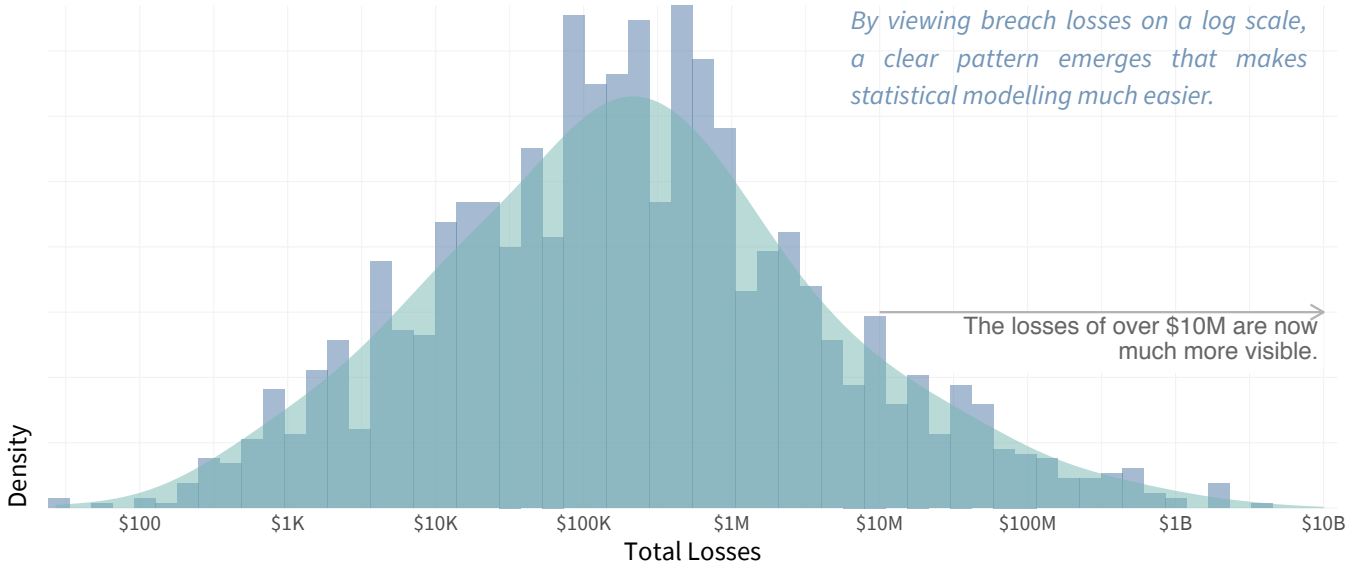
> **Our first task is defining how losses are distributed and then establishing the cost of a typical event. We also challenge the popular approach of using a flat cost per record metric.**

It's clear that Figure 8 is not the classic bell-shaped "normal" distribution you may remember from Statistics 101. The vast majority of breach losses occur at the lower end of the spectrum and we see a long tail of rare-but-extreme values extending to the right. That tail actually extends much further, but we cut it off at $50M to prevent zooming out so far that the chart becomes just a single visible bar on the left. Such distributions are often referred to as "heavy tailed" and they break a lot of the assumptions from that Stats 101 course.

Figure 8: Distribution of breach losses on a linear scale (truncated at $50M)

When dealing with heavy-tailed distributions, it often helps to view them on a non-linear scale. Figure 9 presents losses on a log scale, and you can immediately see why this transformation is so helpful. Things now appear much more symmetrical. Fitting a continuous distribution around the values smooths out the spikes in the histogram[7] and creates the familiar bell curve. That tells us something very important about breach losses: They follow a lognormal distribution.[8] That's super useful information for folks seeking to model cyber risk. But to make fancy statistical statements like "losses are lognormally distributed around $typicalLoss," we need to know what a typical cyber event costs these days.

Figure 9: Distribution of breach losses on a log scale



[7] *The spikes in this distribution are due to repeated loss values. We suspect this is a rounding effect, where organizations tend to report rounded loss figures of $25K, $100K, $1M, etc. We also see evidence of grouped cases that show the same loss value (i.e., a fine or lawsuit evenly apportioned to multiple parties).*

[8] *If mention of a lognormal distribution has you running off to Wikipedia or to (or away from) your college statistics textbook, remember that a lognormal distribution is just a normal (or Gausian, if you're fancy) distribution but over the log of all the values. This means that if we take the log of every point in our dataset, we can apply all the same properties and techniques from a normal distribution to this collection of log-transformed points. Isn't math fun?*

In Search of the "Typical" Loss

Now that we know the overall shape of breach losses, the next logical step is to determine what "typical" losses look like. This is harder than it seems because simply taking the average of all losses does not yield a reliable measure of what's typical in heavy-tailed datasets like this one. Another candidate, the median, always marks the exact middle point in a set of values, which may or may not be what we want. The geometric mean offers a third option that we (and many others) view as the best all-around measure of typicality for many datasets because of its shapeshifting super powers. In near-normal distributions, it behaves like the arithmetic mean. In skewed distributions, it mimics the median.

FAIR Use: This section corresponds to what is known as Loss Magnitude in the FAIR framework. FAIR™ breaks that down further by bucketing losses into six forms: productivity, response, replacement, competitive advantage, fines and judgements, and reputation. Advisen tracks similar categories of losses, but we do not differentiate among them in this study. Some of these are probably better represented by our data (e.g., fines and judgements) than others (e.g., productivity and reputation).

With all this in mind, let's add some measures of central tendency to the distribution of breach losses in Figure 10. Within our sample, we get an "average" loss of $19M (arithmetic mean), while the median and geometric mean peg it around $200K. A disparity between the mean and median should be expected because we've already established that cyber losses are highly-skewed, but some may be surprised at the magnitude of that disparity. Adding some context helps demonstrate how unfit the mean is to be the champion statistic for the common loss: 90% of breaches cost less than the mean of $19.1M. Does that feel "typical" or "average" to you? Not to us either.

Figure 10: Distribution of breach losses on a log scale with measures of central tendency

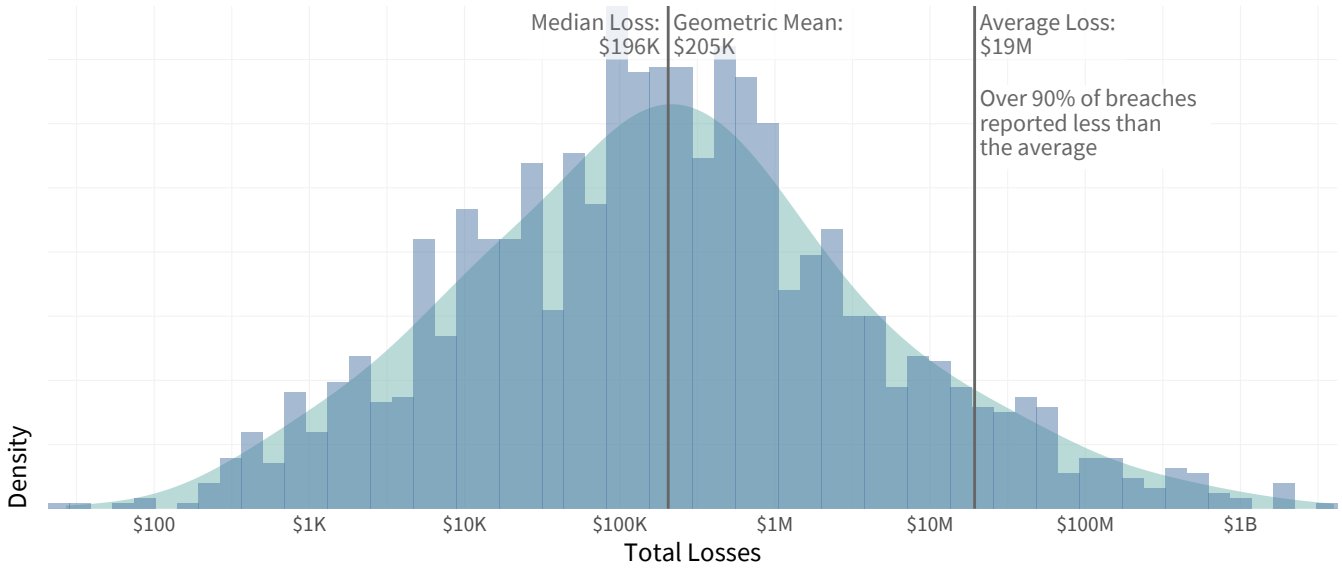


Figure 10 pegs the average cost of a breach at $19M, but 90% of breaches fall below that amount. That's because losses are very heavy tailed, which skews the mean. The median and geometric mean offer a better estimate of typical breach losses at around $200K.

Again, that inflated mean results from a relatively small number of extremely large loss events that simply don't reflect the magnitude of a typical breach. This is why FUD-mongers love touting the average loss rather than the median or geometric mean (and rarely show more than one statistic). Now that you know their scheme, demand they do better.

> **Next time you're asked what a breach will cost, "A couple hundred thousand dollars" is a simple answer backed by lots of evidence. It's also totally appropriate to add, "But there's a 10% chance it could be 100x higher than that (or more)."**

If you don't want to take the Chicken Little approach to loss estimates, try this. Next time you're asked what a breach will likely cost, "A couple hundred thousand dollars" is a simple and sound answer backed by lots of evidence. It's also totally appropriate to add, "But there's a 10% chance it could be 100x higher than that (or more)," to cover your assumption.

Straight Talk on Cost-Per-Record Estimates

"Why, sometimes I've believed as many as six impossible things before breakfast."

—Alice in Wonderland

There's a long history of efforts to establish a "cost per record" for data breaches. This is understandable since logic suggests that total losses would correlate with the size of the breach, usually measured as the number of records compromised. But does the evidence suggest the same? We've found that approaches based on a flat cost per record are more harmful than helpful for the purpose of accurately estimating or predicting cyber losses. Let's explore why.

In Figure 11 we show a basic scatter plot of the number of records compromised versus the total financial losses associated with each incident. If there were such a thing as a linear cost per record, we'd see all these dots forming a tightly grouped diagonal trendline across the grid. Instead the plot is sparse to the point of unreadability, with all but a scattering of nonconformists concentrating in the lower range of both axes. We can fix that.

Figure 11: Records versus losses on a linear scale

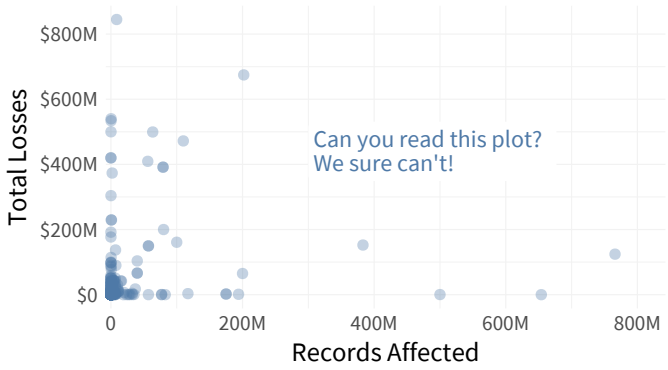


Figure 12: Records versus losses on a log scale

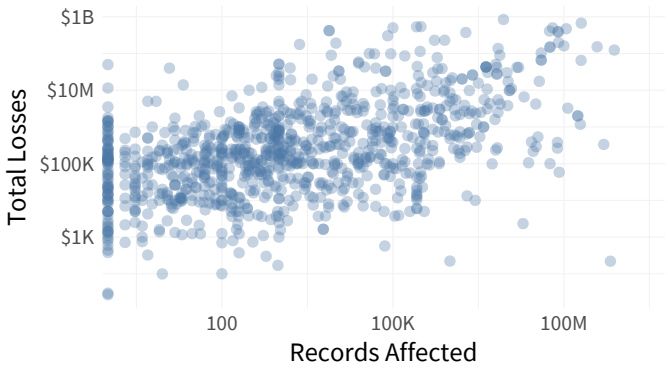
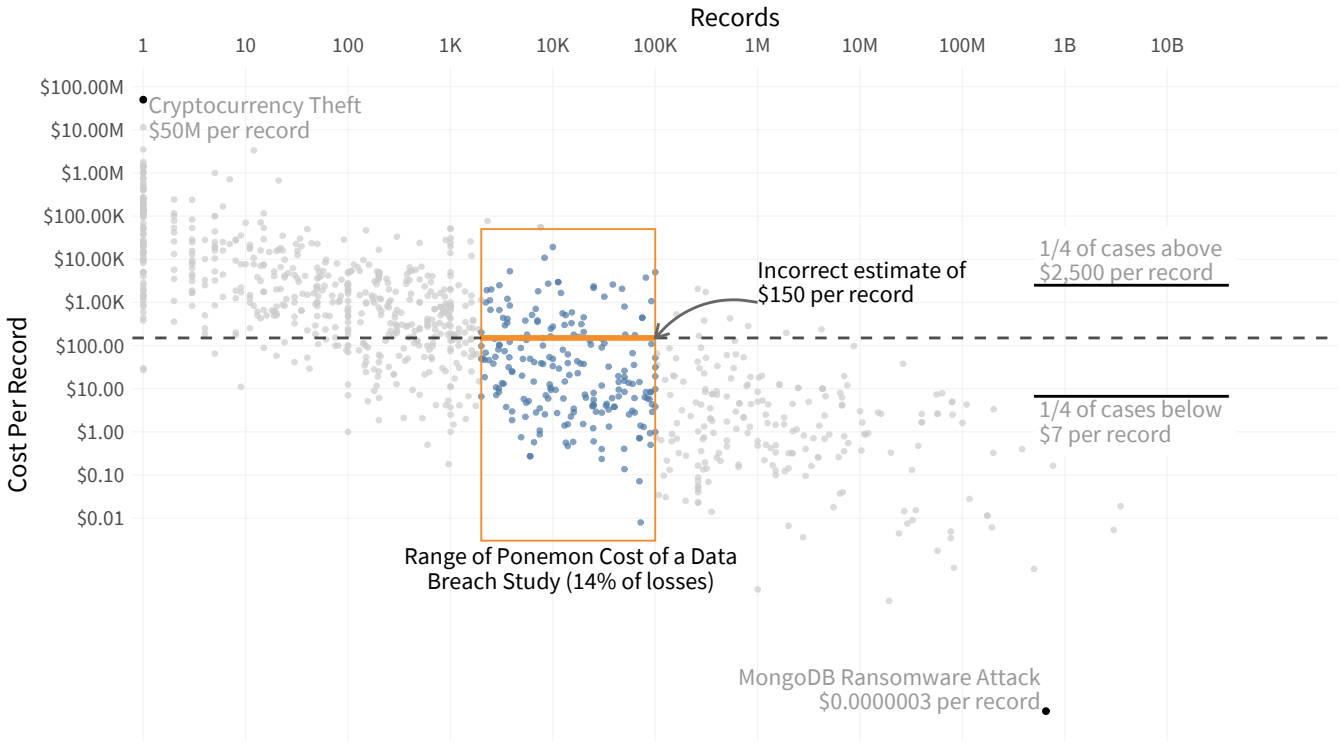


Figure 12 displays the same data but changes the scale from linear to a logarithmic. This immediately results in a clearer relationship between the two variables whereby losses increase by some percentage as the records affected increase exponentially. So there is evidence that larger breaches cost more, but it's definitely not a linear relationship. That means the popular method of multiplying the number of records times an average cost won't yield valid estimates. Let's examine why such a calculation simply doesn't hold water.

Figure 13 calculates the cost per record for each loss event in our dataset. If the relationship between records and losses was linear, these events would converge around the horizontal dotted line running through the chart at the average cost per record. Clearly this is not the case. Instead, losses relative to records affected appear much higher for small breaches before economies of scale kick in to drop the cost to pennies per record for large events. Thus, a single cost-per-record metric simply doesn't work and shouldn't be used. It underestimates the cost of smaller events and (vastly) overestimates large events.

> **A single cost-per-record metric simply doesn't work and shouldn't be used. It underestimates the cost of smaller events and (vastly) overestimates large events.**

Figure 13: The fallacy of a flat cost per record for estimating breach losses



You might be wondering about the orange box in Figure 13. Let's talk about that. That marks the range of breaches to which Ponemon's annual Cost of a Data Breach Study claims its "average cost per record" metric can be applied to calculate total losses. Two things should jump out: 1) The range covers only a small portion of all breaches, and 2) even within that range, the relationship between records and losses isn't linear. Clearly, there's a problem with the data collection and/or calculation method used in that research. Regardless, we can use the latest estimate of $150 per record from that study[9] to test how well it predicts losses for breaches reported over the last decade. Figure 14 does exactly that.

The Ponemon report clearly states that their cost-per-record metric does not apply to breaches over 100K records, but that doesn't stop numerous people from abusing it. One recent egregious example[10] claimed $5T in losses from cloud misconfigurations. If you're wondering how that number was derived, multiply $150 by 33B exposed records and see what you get. No study can fully prevent the misuse of its findings, but the fallacy and scope of this statistic make it inevitable.

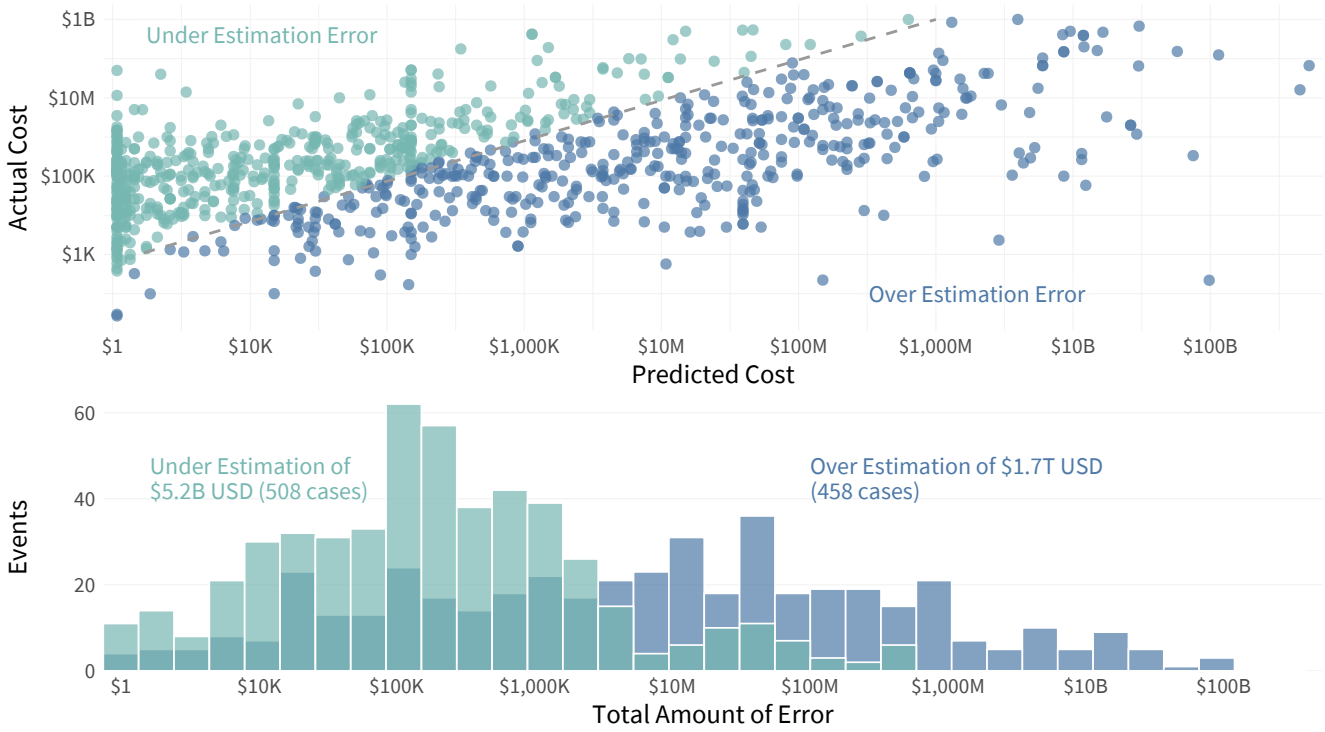
[9] *Cost of a Data Breach Study 2019, Ponemon Institute and IBM Security, 2019*

[10] *2020 Cloud Misconfigurations Report, DivvyCloud, 2020*

Figure 14 is actually two charts that give different views of the same data. Let's start with the scatterplot on top. A dot right on the diagonal dashed line shows where using the proposed $150/record metric perfectly predicts the actual losses from a historical breach.[11] There aren't too many of them. Teal dots above the line indicate breaches where the Ponemon method underestimates actual losses (it's more costly than predicted). Blue dots below the line mark events where estimated losses were higher than actual losses (it's less costly than predicted).

If you measured the vertical distance from each dot (actual loss) in Figure 14 to the line (predicted loss based on $150/record) and then made a histogram of those distances, you'd get something like the bottom chart in Figure 14. Overestimates are tallied in blue, while underestimates appear in teal. This makes it easier to see, for instance, that predicted losses using the Ponemon model routinely exceed actual losses by $10M and those overestimates stretch upwards of $100B for a single loss event!

Figure 14: Single cost per record error rates



If Figure 14 still fails to convey just how far off those predictions are from the truth, try this on for size: The total error (over and under) from those estimates is more than $1.7 trillion dollars. To put that in perspective, that's the equivalent to the 10th largest economy in the world (Canada). It's just way less nice, eh? Another way to look at it is that the typical error from using that model (~$600K) is three times higher than the typical cost of a breach ($200K). We hope this exposes the folly (and puts the last nail in the coffin) of loss estimates based on a simple average cost per record derived from a limited range of data.

> **The total error from those estimates is more than $1.7 trillion dollars. We hope this exposes the folly (and puts the last nail in the coffin) of loss estimates based on a simple average cost per record derived from a limited range of data.**

[11] *For example: A breach of 50K records at $150 per record would be expected to cost $7.5M.*

Using Records to Estimate Loss Less Badly

Does this mean we should altogether avoid using the number of records compromised to estimate breach losses? No, it does not. While a model based on a flat cost of $150 per record has no predictive value at all (it has a negative R^2),[12] record count does meaningfully contribute to a proper loss model that includes other factors. Building such a model falls outside the scope of this historical analysis of loss events, but we do plan to take that up in future research.

But what should we do in the meantime? Some simple statistical techniques can make loss estimates based on record count a lot less bad (and even pretty good). Next time you need a quick estimate for the cost of breach affecting some number of affected records, we recommend using Table 4.

Records	Probability of At Least This Much Loss					
	$10K	$100K	$1M	$10M	$100M	$1B
100	82.0%	49.9%	17.8%	3.3%	0.3%	0.0%
1K	88.4%	60.9%	26.0%	5.9%	0.7%	0.0%
10K	93.0%	71.1%	35.8%	10.0%	1.4%	0.1%
100K	96.0%	79.8%	46.7%	15.8%	2.7%	0.2%
1M	97.9%	86.7%	57.7%	23.5%	5.0%	0.5%
10M	99.0%	91.8%	68.2%	32.8%	8.6%	1.1%
100M	99.5%	95.3%	77.4%	43.4%	13.9%	2.3%
1B	99.8%	97.4%	84.9%	54.5%	21.0%	4.2%
10B	99.9%	98.7%	90.5%	65.3%	30.0%	7.4%

Table 4 should help those wanting to estimate losses based on the number of records affected by a cyber event. a breach of 100K records will almost certainly (96% chance) cost at least $10K but probably won't (2.7% chance) exceed $100M. Not as easy as multiplying by $150, but it'll go a long way toward better risk assessments.

Table 4: Probable losses based on records affected in a breach

Table 4 works like this: Pick the number of records for which you're trying to estimate losses. The percentages in that row denote the probability of losses in the amounts shown in each column. So, for example, a breach of 100K records will almost certainly (96% chance) cost at least $10K but probably won't (2.7% chance) exceed $100M. If you want to estimate something between those ranges (e.g., 500K records), you'll need to read (estimate) between the lines. No, it's not as easy as just multiplying by $150, but it's a whole lot more accurate, and that will go a long way toward better risk assessments.

> **"It's logical that the cost of a breach would be different for a municipal brewer than a multinational broker. The next section examines how losses vary across firmographic dimensions.**

Breach Losses by Revenue and Sector

Thus far, we've managed to establish what breach losses look like overall but made no distinction among organizations of different types and sizes. It's logical that the cost of a breach would be different for a municipal brewer than a multinational broker. This section revisits the firmographic dimensions analyzed in the frequency section to understand how losses vary across the Fortune 1000, industry sectors, and annual revenues. Buckle up—lots of long-tailed distributions ahead!

[12] *Yes; there is such a thing as a negative R^2. It says using the model is worse than just quoting the mean. See http://www.fairlynerdy.com/what-is-r-squared/*

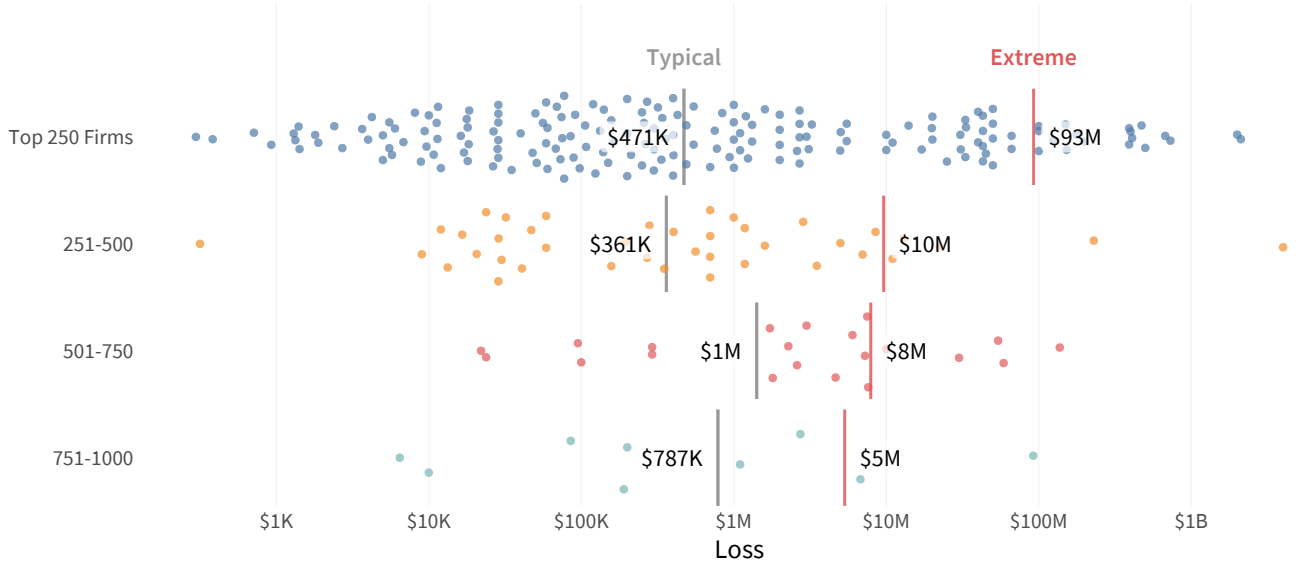
Breach Losses in the Fortune 1000

Following the precedent set in earlier sections, we begin our examination of cyber event losses with the Fortune 1000. We also once again divide the Fortune 1000 into ranked quartiles for easy comparison. Figure 15 is a rather information-intensive chart that conveys three key aspects of loss magnitude:

- It shows the full range of historical loss events affecting each segment. Every dot is a loss event and its location on the horizontal axis denotes the total reported cost.
- It estimates the typical cost for breaches in each segment. The gray bar marks the geometric mean of observed losses.
- It estimates an extreme loss for breaches in each segment. The red bar marks the 95th percentile of observed losses.

Armed with that explanation, we're now ready to see what Figure 15 actually says about breach losses within the Fortune 1000 quartiles. Whereas the frequency of cyber events rises in concert with corporate rankings, the average impact of those events does not appear to exhibit the same predictable trend. Expected losses for the bottom half of the Fortune 1000 exceed that of the top half, though it should be noted that many more data points exist for the top 250 firms.[13] We find this somewhat encouraging as it suggests there may be a natural upward limit for the cost of ordinary breaches that isn't simply a reflection of a firm's revenue.

Figure 15: Distribution of breach losses for the Fortune 1000 with estimates for typical and extreme events



Whereas the frequency of cyber events rises in concert with corporate rankings, typical losses do not appear to follow that trend. Extreme events, however, do rise with the rankings. The 95th percentile for the top 250 is almost 10x that of the next bracket (251–500) and nearly 20x the bottom 250.

Extreme events, on the other hand, do rise with the rankings. The 95th percentile for the top 250 is almost 10x that of the next 250 (251–500) and nearly 20x the bottom 250. From this, we infer that the importance of managing tail risk from cyber events increases with organization size. This will become even more apparent in the next section as we look at losses outside the Fortune 1000.

> **From this, we infer that the importance of managing tail risk from cyber events increases with organization size. This becomes even more apparent outside the Fortune 1000.**

[13] *The fact that we have many more cyber events with reported losses for the top 250 firms in the Fortune 1000 corroborates our previous statement that loss statistics are biased toward larger organizations with a higher public profile.*

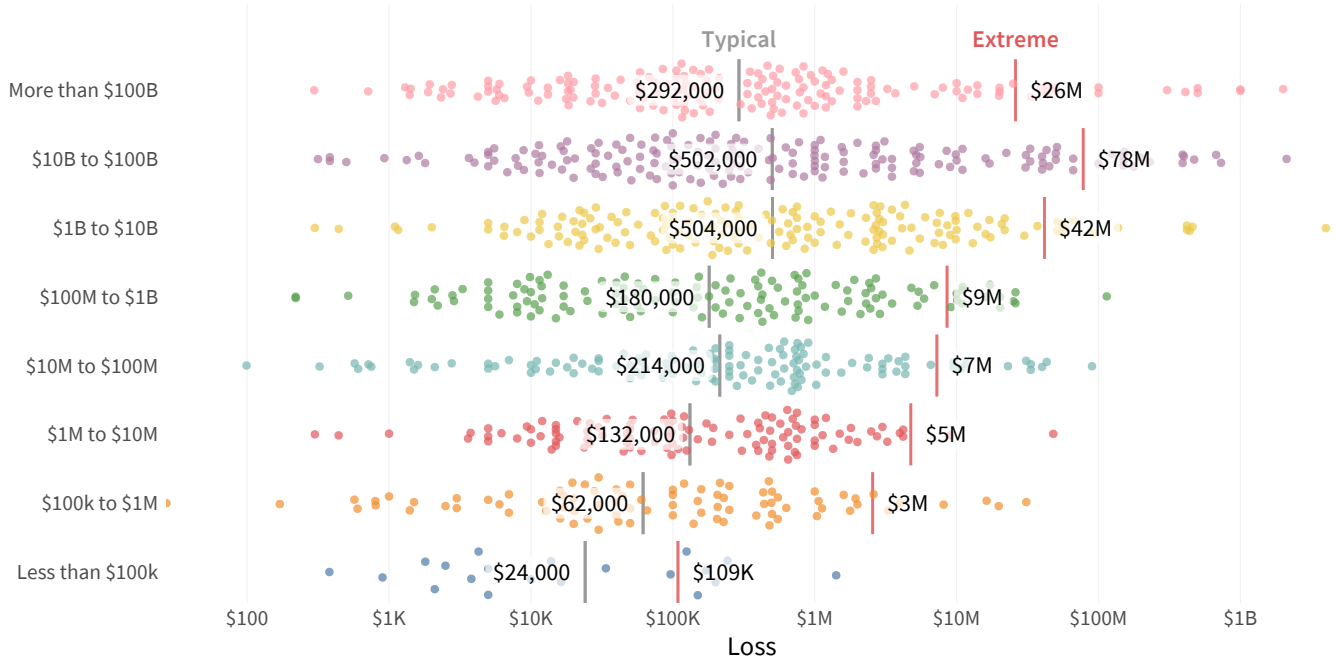
Breach Losses by Revenue Tier

Moving downstream from the Fortune 1000, we do see a more apparent (but not surging) upward trend in typical losses with increasing revenues. Firms under $1M in revenues generally report losses below $100K. The cost of the average cyber events doesn't exceed $200K until revenues climb above the $1B line. After that, typical losses look more or less similar to one another and to the pattern we saw for the Fortune 1000.

In line with our earlier observation, tail costs become even more inflated as organizations grow in size. Extreme losses for the largest revenue brackets are over 100x that of typical costs. A similar typical-to-extreme ratio applies to all firms below $100M in revenues (though it falls to around 30x for <$10M). This once again emphasizes the importance of managing the long tail of cyber risk.

> **A $100B enterprise should expect a cost that's 0.000003% of annual revenues for a typical breach. A mom and pop shop, on the other hand, will likely lose 1/4 of their earnings.**

Figure 16: Distribution of breach losses by firm size (in revenue) with estimates for typical and extreme events



In Figure 16, we see an upward trend in typical losses with increasing revenues. Tail costs become even more inflated as organizations grow in size. Extreme losses for the largest revenue brackets are over 100x that of typical costs.This once again emphasizes the importance of managing the long tail of cyber risk.

There's one more thing we'd like to call out related to Figure 17. While the extremity of losses do indeed grow with organization size, their relative impact does not. It's difficult to do the proper math with the revenue ranges provided, but a rough estimate will demonstrate the point. A $100B enterprise that experiences a typical cyber event ($292K) should expect a cost that represents 0.000003% of annual revenues. A mom and pop shop that brings in $100K per year, on the other hand, will likely lose one-quarter of their earnings ($24K) or more. A key takeaway here is that while a small firm may not experience one of these publicly discoverable events very often as compared to their much larger peers, when they do, the relative impact to their bottom line is huge. Economies of scale seem to very much apply to cyber risk.

Breach Losses by Sector

That brings us to the third and final dimension of comparative cost analysis for cyber events, NAICS sectors. We've filtered out industries for which we have less than 10 observations in Figure 17 to avoid drawing conclusions on overly sparse data. Beyond that, let the number of dots be a guide for how much weight to give loss statistics for your sector(s) of interest.

Here again we see that typical loss events are usually within an order of magnitude of one another. The Information, Manufacturing, and Retail sectors incur the highest average costs, but it's not as though the other industries skip town when the breach bill hits.

Figure 17: Distribution of breach losses by sector with estimates for typical and extreme events

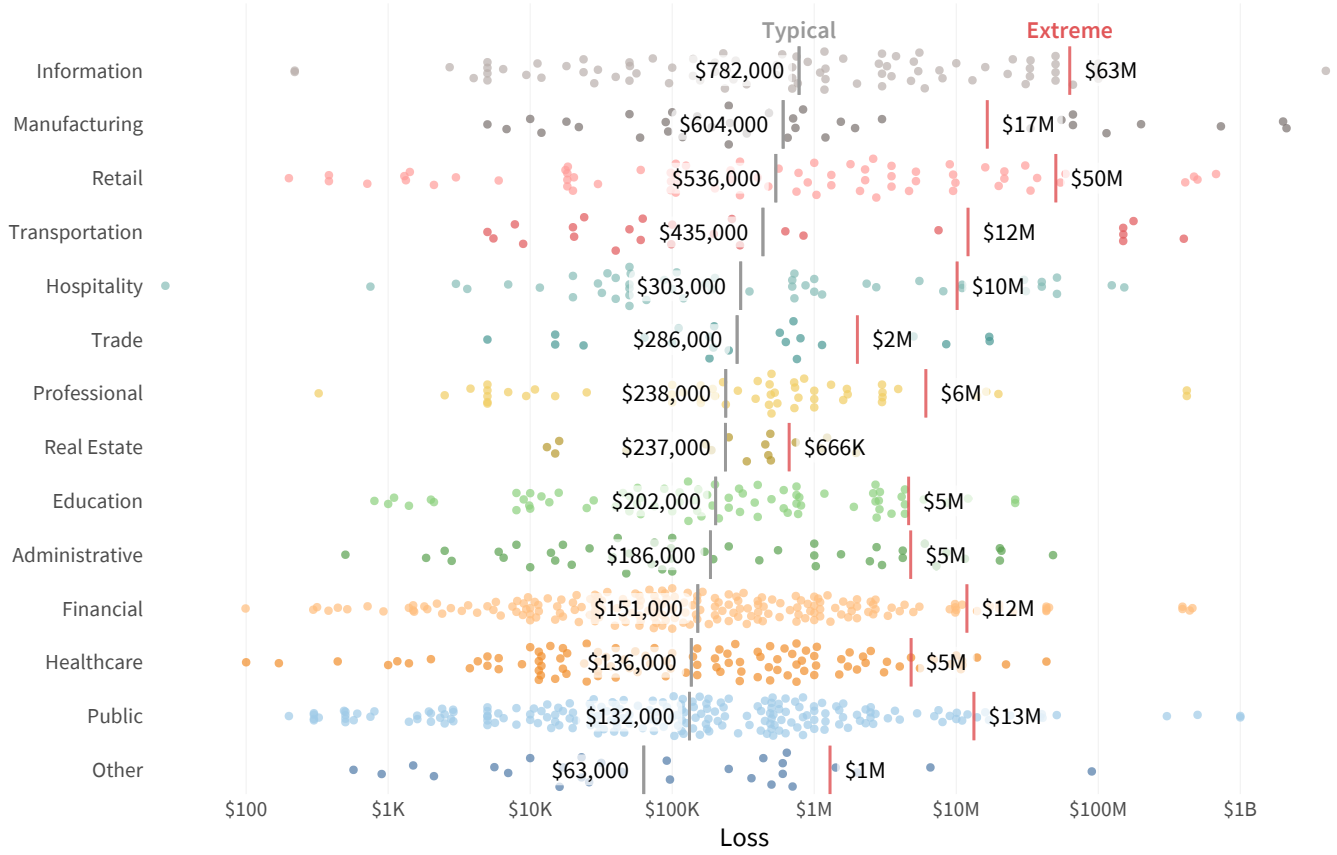


Figure 17 shows that typical loss events are usually within an order of magnitude of one another. The Information, Manufacturing, and Retail sectors incur the highest average costs. That trio also leads for extreme events, except their costs now exceed many other sectors by a factor of 10.

Focusing now on extreme events, we see more substantial variation. The same trio of sectors leads once again, except now their 95th percentile costs exceed many other sectors by a factor of 10. The data-intensive nature of these organizations combined with extensive supply chain dependencies may have something to do with that state of affairs.

As we consider the findings presented in Figure 17, the financial sector strikes us as a bit odd. One might assume that financial firms would bleed cash after a breach, but that doesn't seem to be the case. They apparently curtail losses quite well, despite being among the sectors most likely to suffer a breach. We suspect their long history dealing with risk and regulation helps them mitigate loss events better than the average organization.

Exceeding the Risk Curve

"And you may ask yourself, 'How do I work this?'"

—Talking Heads, "Once in a Lifetime"

Having taken an extended tour through both the frequency of breaches and the magnitude of their impacts, we hope you're eager to apply this analysis to your own environment. Here at Cyentia, we're big fans of quantitative risk analysis based on cold, hard facts. The estimates of event frequency and loss distributions we share in this report can be used as baseline parameters in combination with frameworks like FAIR to measure and manage information risk for your organization.

One way of getting to the bottom line for risk managers is to create something called an exceedance probability curve (EP Curve) or loss exceedance curve (LEC). We'll use the latter term since it seems to be more common in the domain of cyber risk. LECs combine the frequency of adverse events with potential impact in a single curve showing the probability that losses will exceed a certain threshold within a given amount of time. Since preventing all losses is generally not feasible for most firms, understanding the chance that losses will exceed a risk threshold (set by each firm based upon their own risk appetite) is a great way of understanding if additional investment in risk mitigation is warranted.

"LECs combine the frequency of adverse events with potential impact in a single curve showing the probability that losses will exceed a certain threshold within a given amount of time.

A full treatment on loss exceedance curves would require more space than we have left in this study, but it's absolutely something we plan to do a lot more of in future research. As an example of how you can use the frequency and loss estimates we shared (and as a preview of things to come), we've created an example LEC for the Fortune 1000 based on our dataset. Using a data source such as Advisen, a risk manager could create loss exceedance curves to match their industry, annual revenue, and other firmographics.

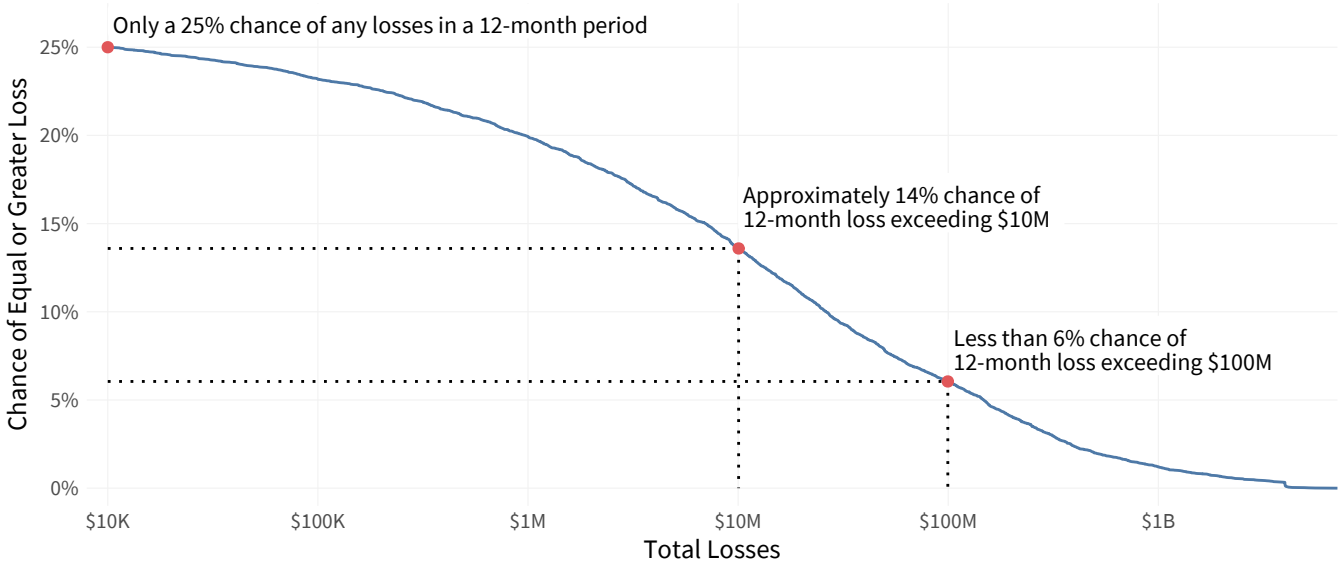
Wondering how to create your own LECs?

To create Figure 18 (see next page), we took the data behind Table 1 showing the probability of a given Fortune 1000 firm having a certain number of breaches in a 12-month period. We then used those parameters to perform a Monte Carlo simulation of 10,000 possible outcomes for future years. In conjunction with this modeled event frequency, we used that lognormal loss distribution derived earlier to simulate how much each event would have cost our hypothetical Fortune 1000 firm. Summing up the total annual losses in these 10,000 hypothetical periods gives us a handy cumulative distribution of the chance of losing an equal or greater amount in a year's time.

If that sounds like a heavy lift, don't despair! There are several simpler approaches to consider. The simplest is to do nothing for now and wait for our future reports that will delve into this concept in more depth. Beyond that, it would be pretty simple to use Table 4 to do a quick "what if" cost analysis for a breach affecting some range of records. Combining that with the estimates from Figure 6 or 7 will help determine the likelihood of a loss event. A step further would be to construct some basic "min-max-most likely" parameters for simple simulation packages. And there's always the option of going straight to the data source (Advisen in this case) to see what kind of tools they offer.

Figure 18 is surprisingly straightforward once you get your bearings. There's an incredible amount of data and calculation behind them, but on the surface they effectively distill a complicated risk assessment into a single curve. That curve answers a common question: "What's the chance we'll lose more than $X next year?" The ability of LECs to succinctly communicate that answer is what makes them so powerful. Let's look at what Figure 18 tells us about the Fortune 1000.

Figure 18: Loss Exceedance Curve for a typical Fortune 1000 firm based on historical data



Earlier in this study, we learned that there's a 25% chance that a typical Fortune 1000 firm will suffer a breach in a 12-month window of time. That's marked at the starting point of the curve in the upper left. And keep in mind that suggests the most likely scenario is that the organization won't experience any cyber events or losses. From there, we slide down the curve to determine that the probability an organization will lose $10M+ in a year is 14%. Less than 6% of the Fortune 1000 are facing probable losses exceeding $100M. The risk of cyber events crossing the billion-dollar threshold is slim, but certainly not impossible. In fact, it's statistically likely that one of the Fortune 1000 will incur costs of this magnitude from cyber events before 2020 ends.

Let's hope this study helps them—and organizations of all types and sizes—to better anticipate and manage that risk.

Join the information risk management community!

If you'd like to get to know and learn from others who assess and manage cyber risk, there are two professional communities you should consider joining (the Cyentia team participates in both):

Society of Information Risk Analysts (SIRA): The Society of Information Risk Analysts (SIRA), established in 2011, is the go-to resource for decision makers & practitioners of information risk management. We endeavor to do this by supporting the collaborative efforts of our members through research, knowledge sharing, and member-driven education. Find out more at www.societyinforisk.org

FAIR™ Institute: The FAIR™ Institute is a non-profit professional organization dedicated to advancing the discipline of measuring and managing information risk. The FAIR™ Institute and its community focus on innovation, education and sharing of best practices to advance the FAIR™ framework and the information risk management profession. Find out more at www.fairinstitute.org

Find value in this report? Join us for the next one!

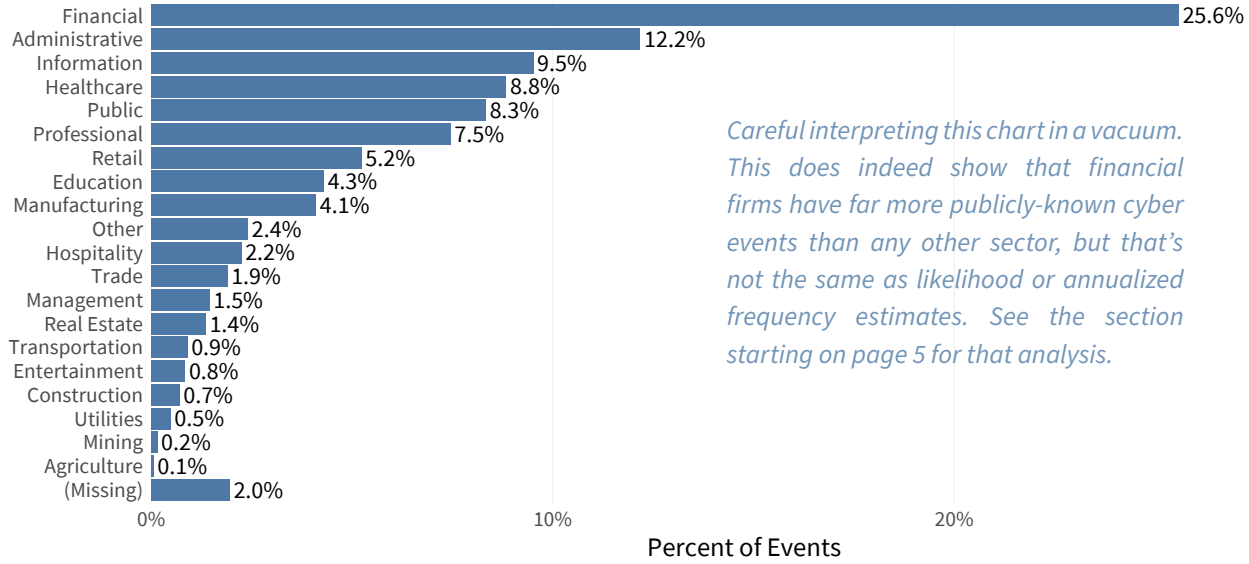
Methodology & Firmographics

All incidents and losses analyzed in this report come from Advisen's cyber loss data. For those new to this data set, Advisen maintains a repository nearing 100,000 cyber events, with events ranging back as far as the mid-20th century. They compile this valuable information through publicly-available sources such as breach disclosures, company filings, litigation details, Freedom of Information Act requests, etc. The dataset also goes through a rigorous process of matching events to known company IDs (e.g. D&B and S&P). This enables the many firmographic views we share in this report. See Advisen's cyber risk data methodology for more detail.

For this report, we are working off of the November 2019 release of the Advisen data feed, focusing on a ten-year window ranging from 2009 to 2019. Advisen tracks several different types of cases such as ransomware, privacy, denial of service, etc. We removed incidents that are exclusively privacy related as these are dominated by issues that most information security practitioners do not typically include in their response plans (items such as telephone privacy, etc.). Our ten-year observation period includes 56K cyber events, of which 1,900 record financial losses associated with the event and nearly 12K have counts for the number of records involved. There is an overlap of just under 1,000 events which contain both financial losses and the number of data records compromised.

Looking at just the breached firms ignores all the firms that do not have publicly disclosed breaches. To get a sense for how large of a net we're casting, we use Advisen's comprehensive feed on the number of firms in the global economy, segmented by several different dimensions such as revenues, number of employees, and industry. Specifically, we retrieved the data for the most recent completed year (2019), filtered down to those companies with headquarters in the US, and applied categories to the revenue and employee counts. This gives us 37,352 breached firms in our observation window, with about three quarters (28,041) headquartered in the US. The charts below provide basic firmographic information on organizations included in this dataset.

Figure 19: Organizations represented in this study by sector



Careful interpreting this chart in a vacuum. This does indeed show that financial firms have far more publicly-known cyber events than any other sector, but that's not the same as likelihood or annualized frequency estimates. See the section starting on page 5 for that analysis.

Figure 20: Organizations represented in this study by annual revenue (left) and employee count (right)