Regression Model for the Impact of a Data Breach for a Financial Institution

Thomas Lee[2], Jason Hegland[1] , Spencer Graves[2]

1. Stanford University Law School, Stanford Security Litigation Analytics, Palo Alto California
2. VivoSecurity Inc., Los Altos CA,

Introduction

- An impactful data breach is a tail event often addressed in scenario analysis using expert judgment.
- The questioner-studies, which are a fixed amount per record, are often used.
- We have performed a regression analysis to aid expert judgment.
- We find variables considered but eliminated, as well as variables retained provide insights to expert.

Sources & Methods

Summary

Sources

Source	Purpose
Advisen	• Principle source • Data breach cost • Lawsuit probability
EDGAR	• Research cost in 10-K SEC filings
California Attorney General	• Comparison analysis
Maryland Attorney General	• Comparison analysis. • Estimate total number of data breaches
HHS	• Comparison analysis.
Interviews	• Challenge assumptions

Tools

Tool	Purpose
SQL Server	• Filtering • Aggregation
R-Studio	• Modeling
Python	• Additional variables
Excel	• Manual corrections

Process

Critical

Step	Purpose
Data Correction	• Created independent and meaningful variables
BMA	• Screen many variables
LM	• Final cost modeling • Establish SE
GLM	• Lawsuit Probability
Monte Carlo	• Combine cost model with lawsuit probability model

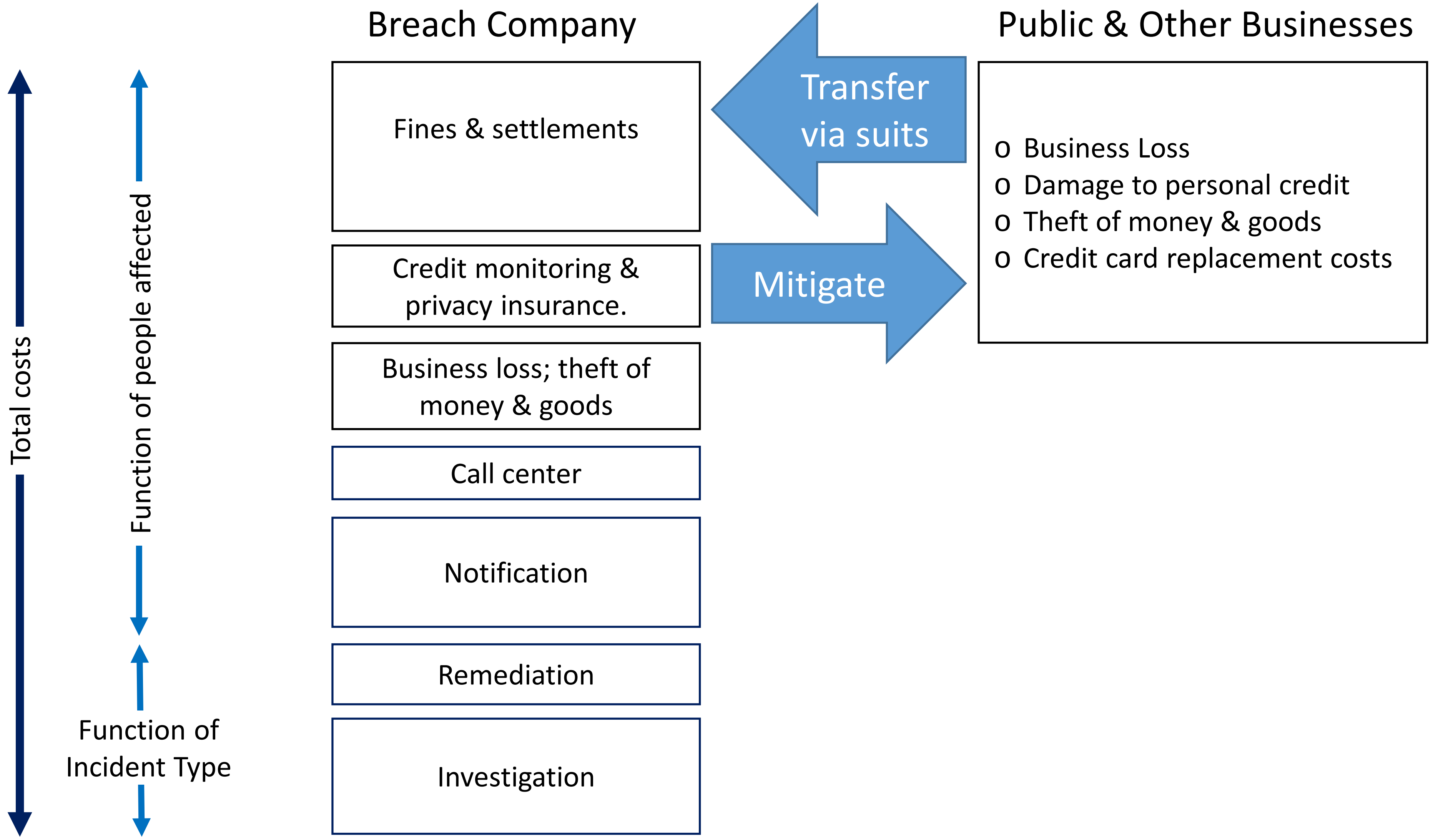
Establish Scope

Total Cost of PII Data Breach

- Impact from breach of PII that triggers legal reporting requirements:
  - PHI, PII, CHD, PFI

- Does not include:
  - Impact from intellectual property loss
  - Denial of services
  - Ransomware
  - Data corruption or loss
  - Fraud

Total Cost of PII Data Breach

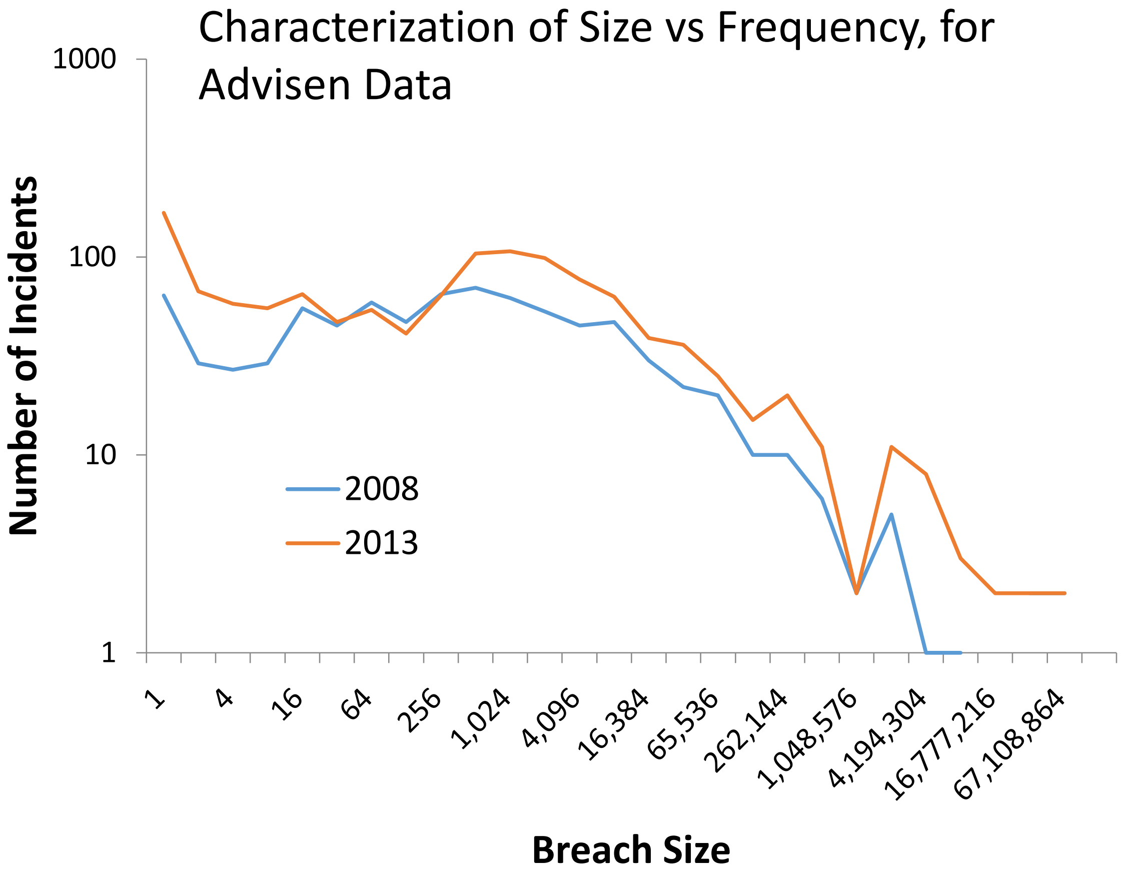


Breach Company
Public & Other Businesses
Total costs
Function of people affected
Function of Incident Type
Fines & settlements
Credit monitoring & privacy insurance.
Business loss; theft of money & goods
Call center
Notification
Remediation
Investigation
Transfer via suits
Mitigate

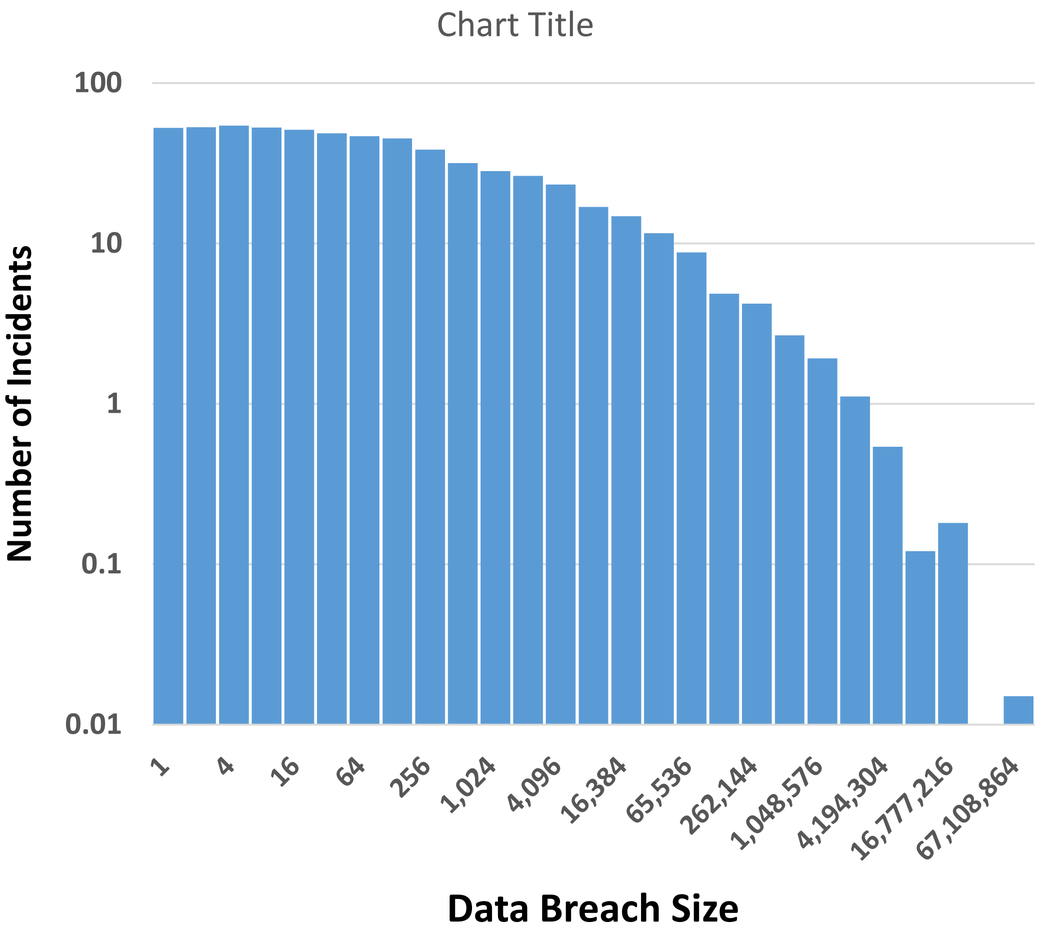
- Business Loss
- Damage to personal credit
- Theft of money & goods
- Credit card replacement costs

Suitability of Data

No Significant Size vs Frequency Bias



Characterization of Size vs Frequency, whole United States, based upon Maryland Attorney General

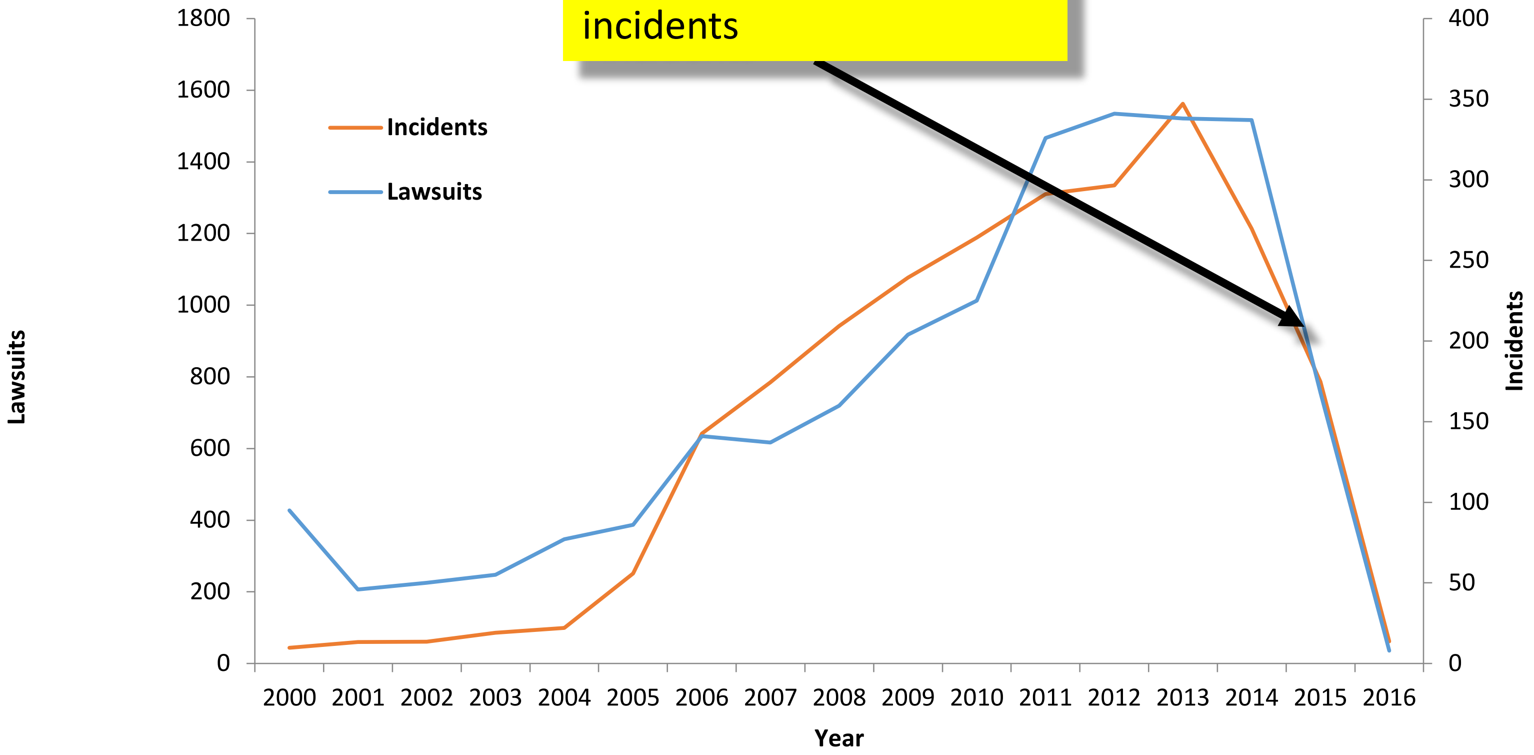


Suitability of Data

Significant drop in Incidents per Year

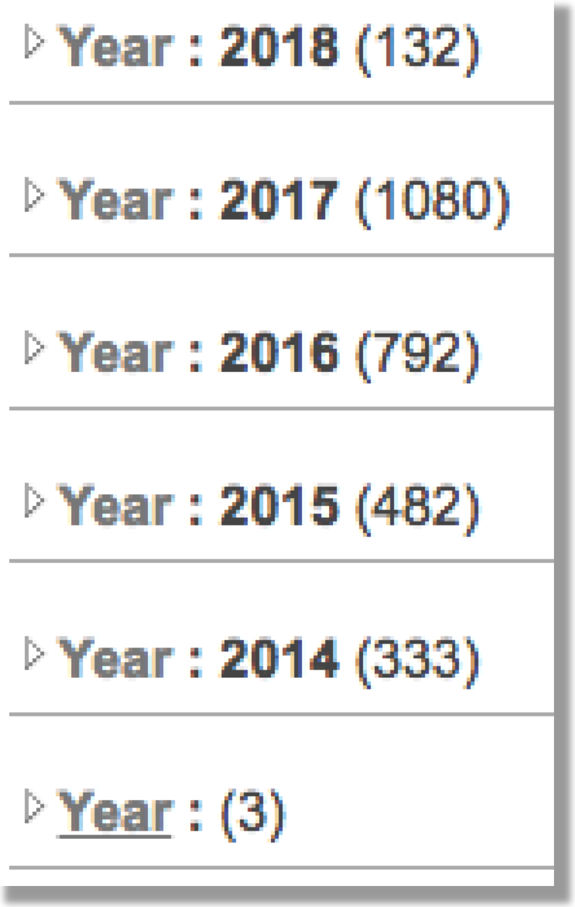
Could Affect Lawsuit Probability

Unexplained drop-off in incidents



Maryland Breach Notices

Data breaches are increasing


▷ Year : 2018 (132)

▷ Year : 2017 (1080)

▷ Year : 2016 (792)

▷ Year : 2015 (482)

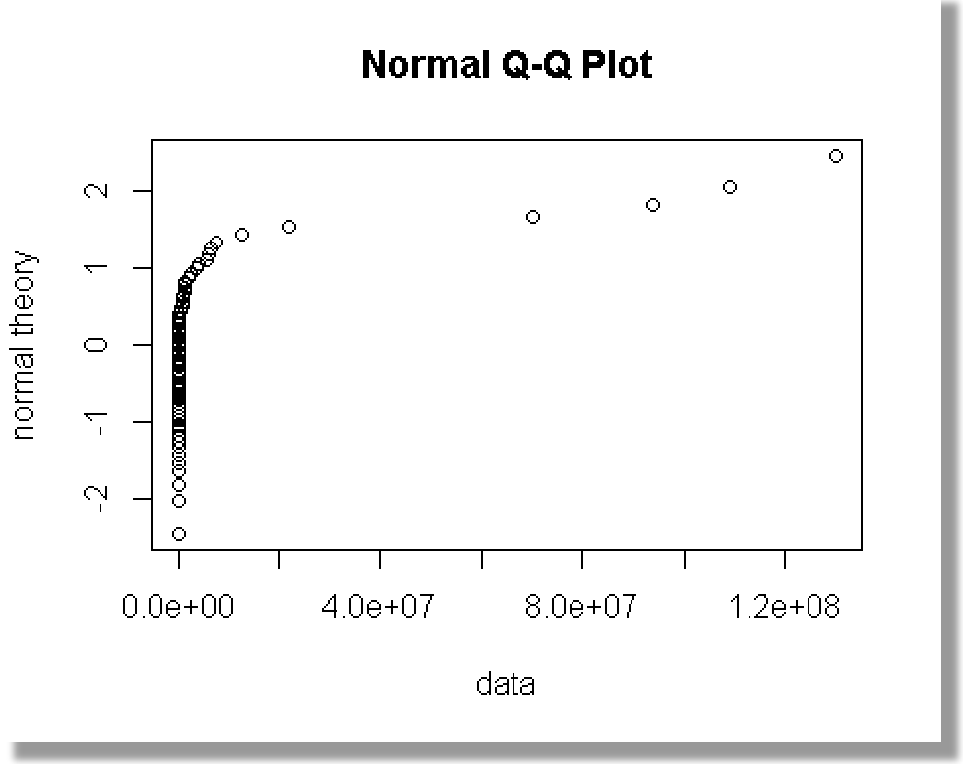
▷ Year : 2014 (333)

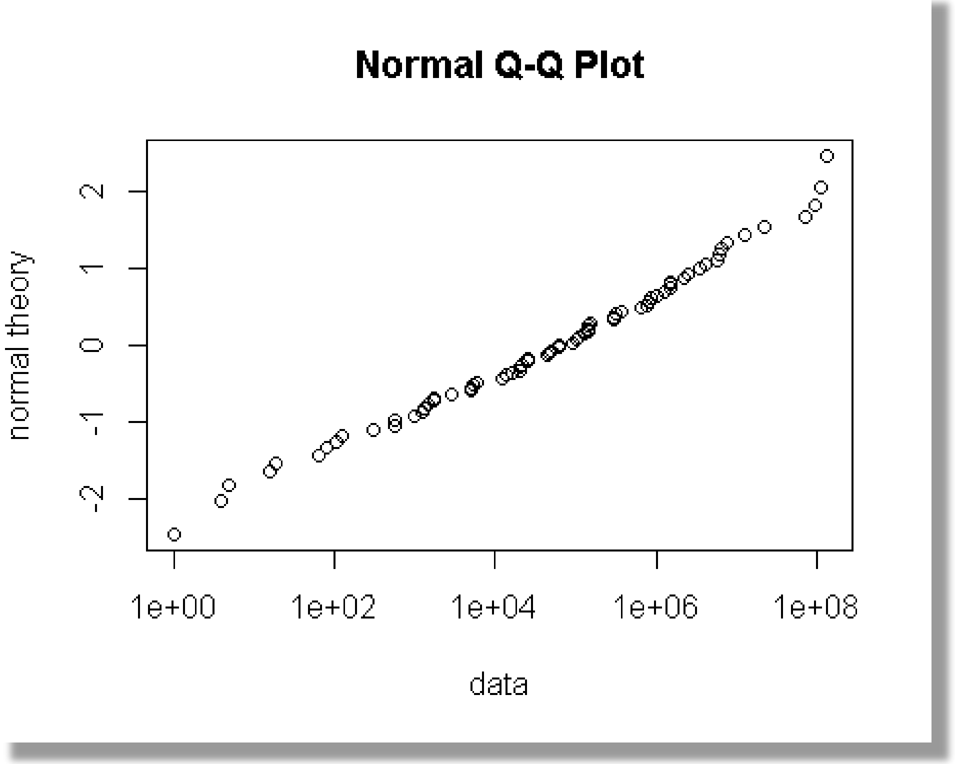
▷ Year : (3)

Data Transformation

Affected Count

Best transformed as log

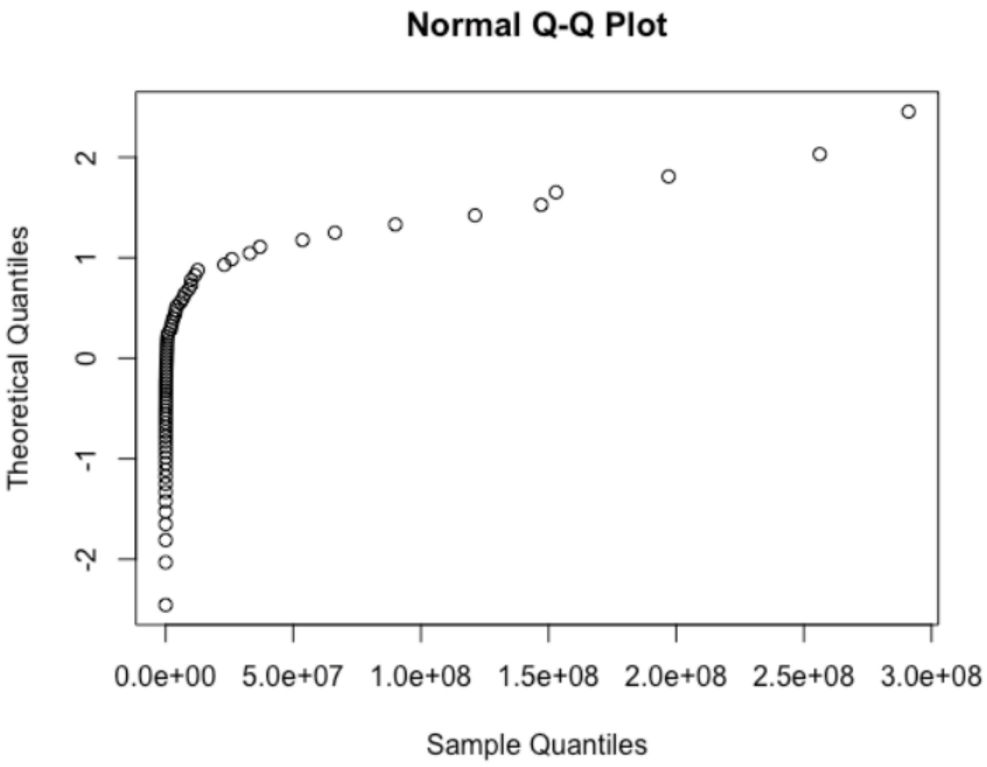


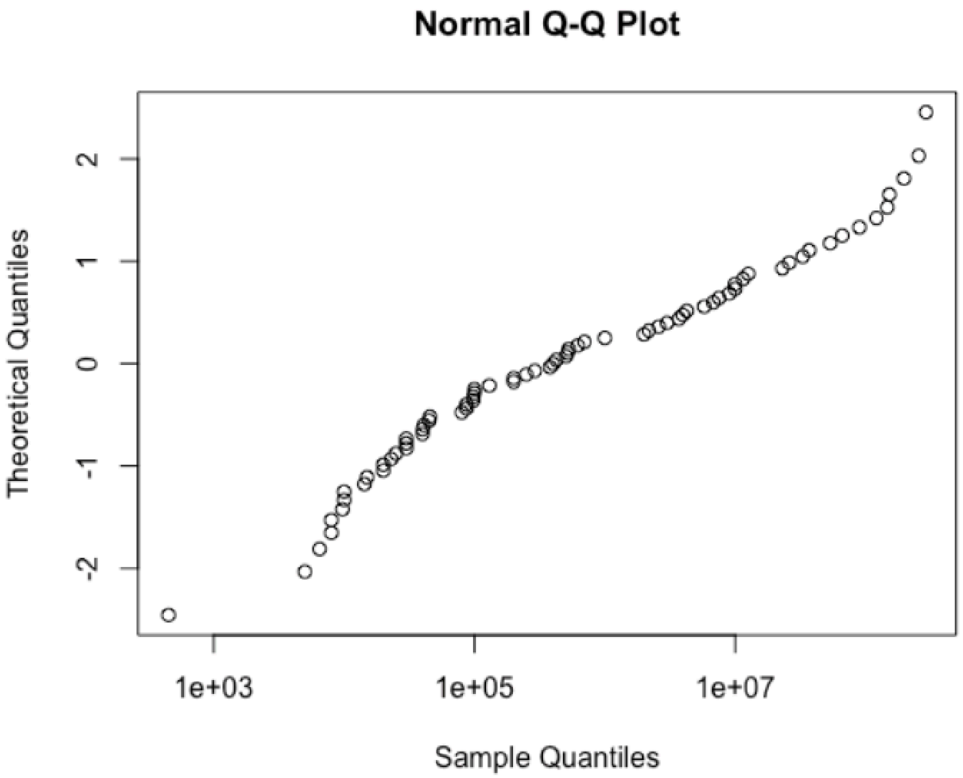


Data Transformation

Calculated Total

Best transformed as log

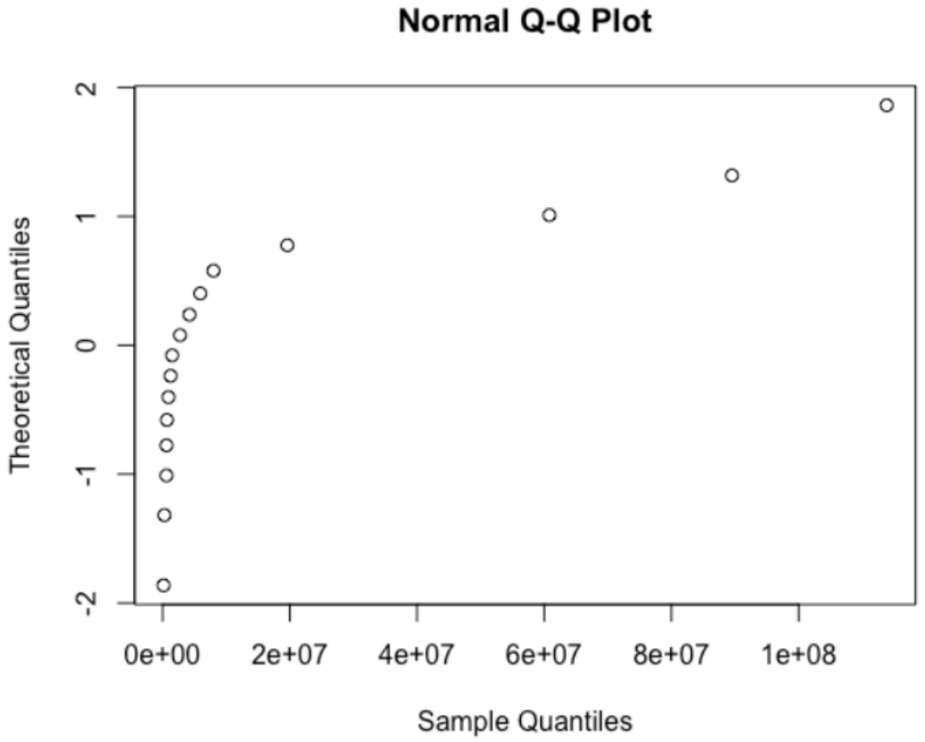


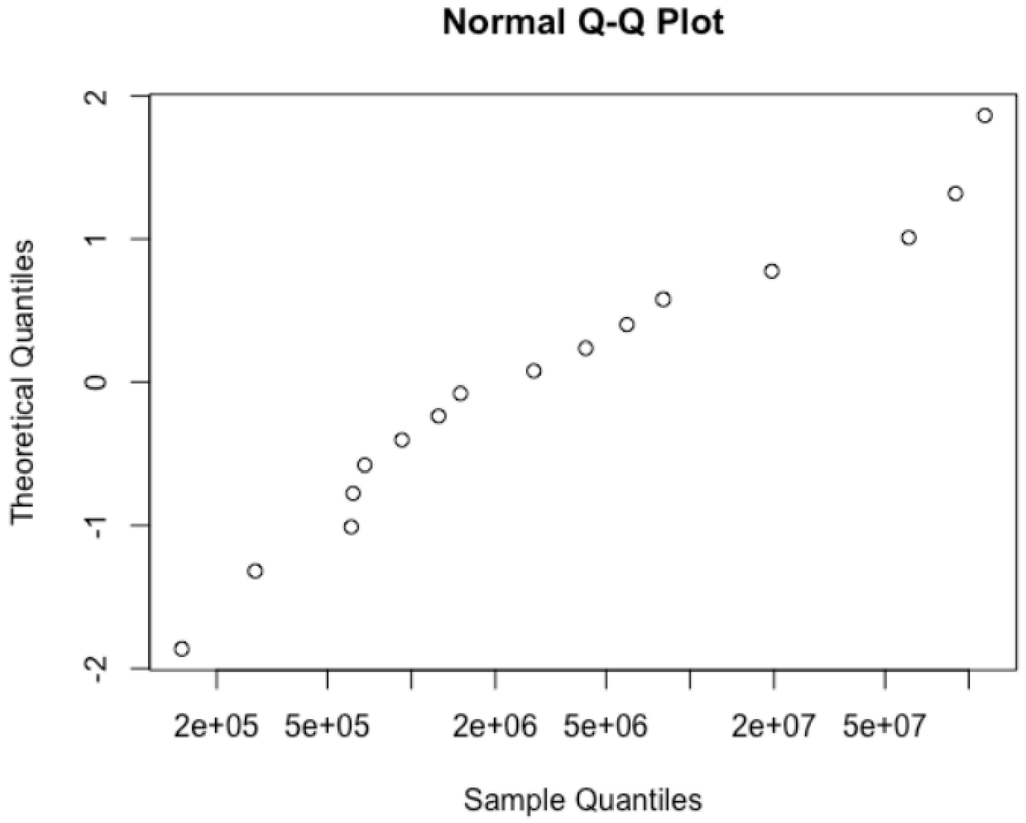


Data Transformation

Lawsuits

Best transformed as log





Best Cost Model

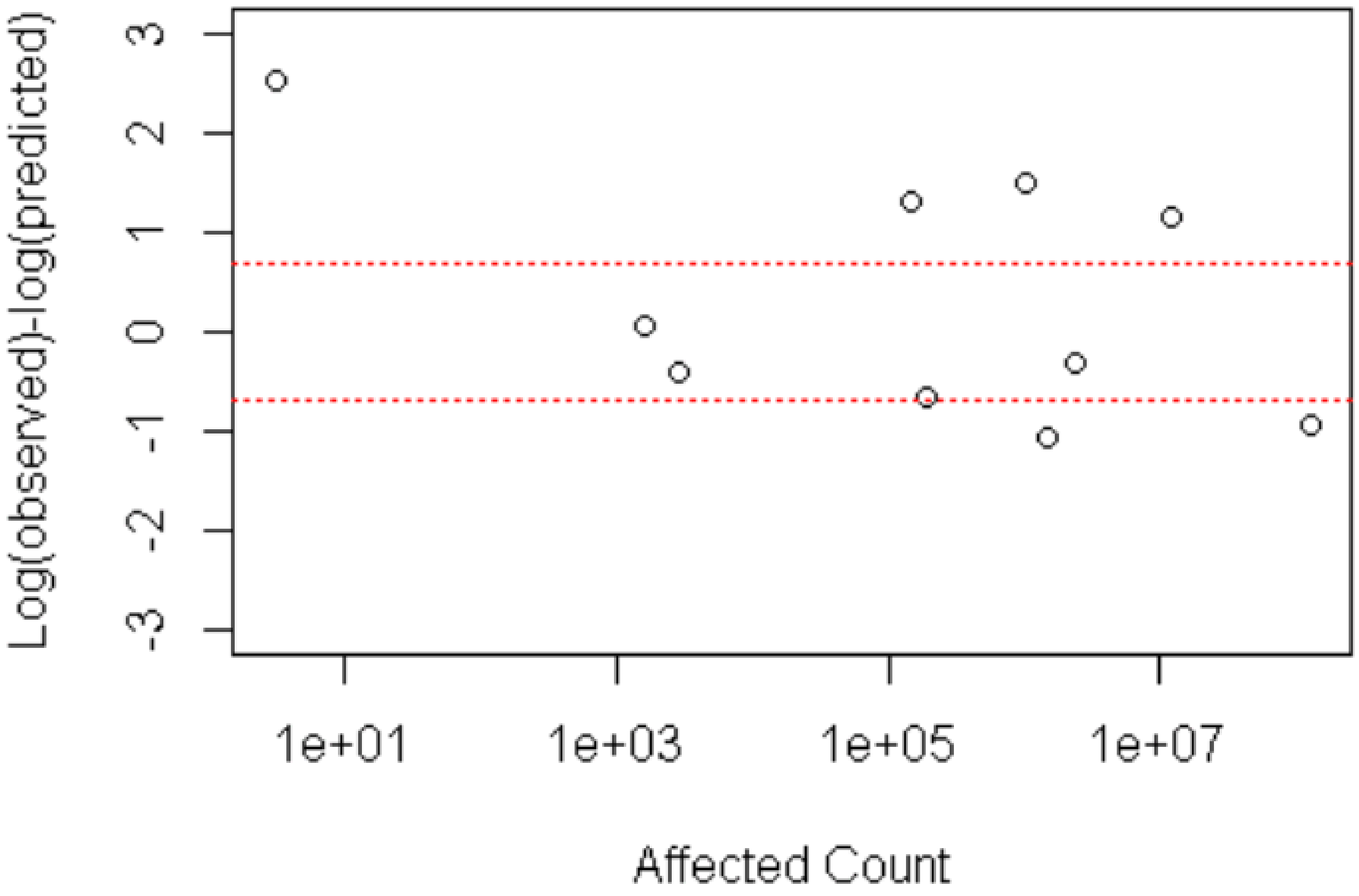
Full Formula

$$Cost = e^{\left(12 - 3.6 \times BrchY - 1.4 \times (IncidentOther) + 0.7 \times \ln(1 + Lawsuits) + \frac{BrchY \times \ln(Affctd)}{2}\right)}$$

Variable	Description	Pr(>|t|)
BrchY	Data was breach, leading to 1) notification, 2) call center, 3) privacy insurance, 4) possible lawsuits	4.9e-4
Lawsuits	Number of lawsuits filed as a result of the data breach.	5.1e-2
Affctd	Number of people affected by the data breach (number of people with exposed PII data)	1.7e-9
IncidentOther	Data breach was caused by any the following: Malicious Insider, Lost or Stolen Device or Accident. Data breach was NOT caused by a Malicious Outsider.	6.0e-3
	R-Sqrd	0.69

Forecast Accuracy

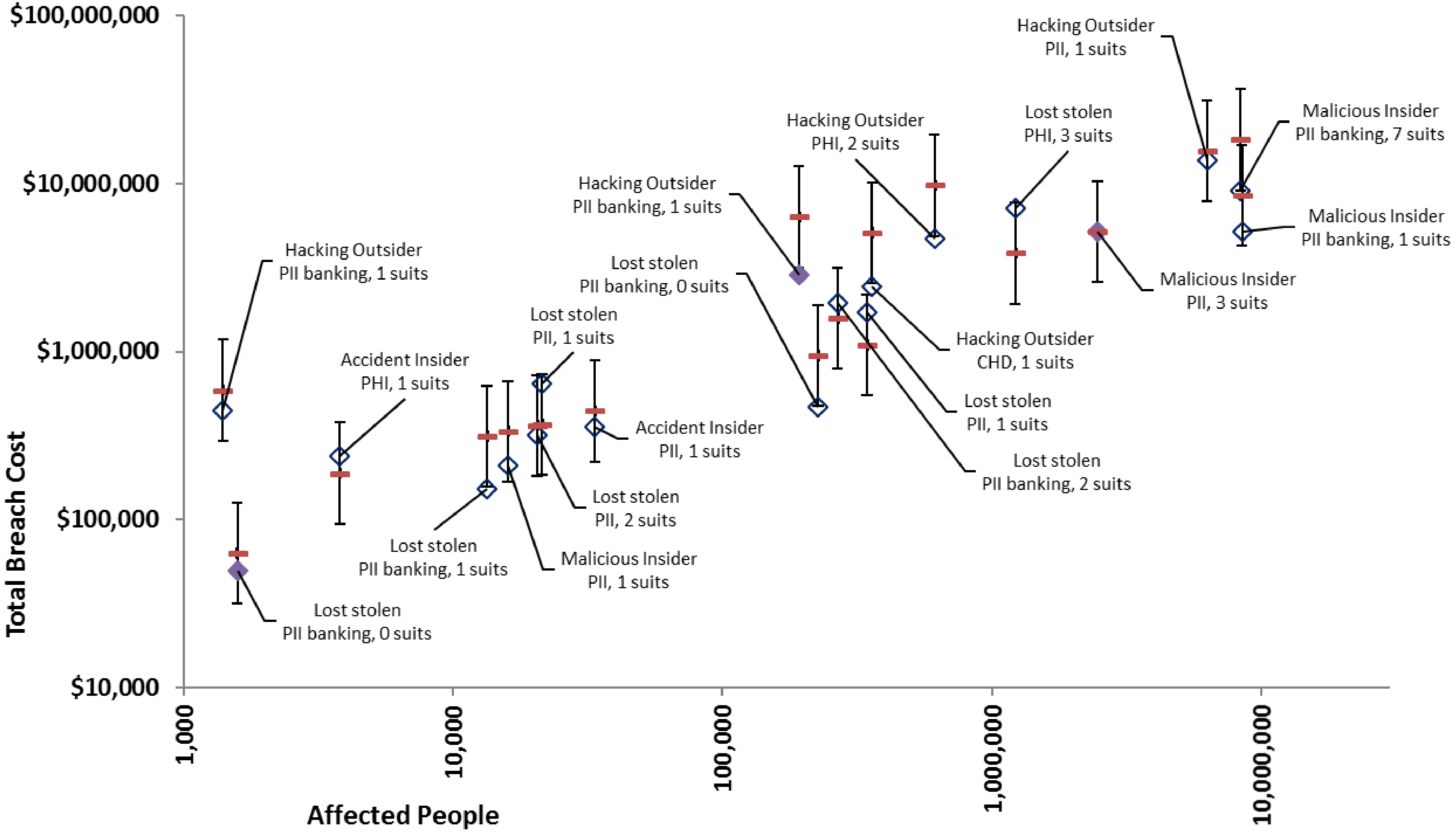
10 observations from training set, NAICS=52



2x range

Forecast Accuracy

- 19 Observations found after model development
- Red bars are median costs,
- error bars are 2x range



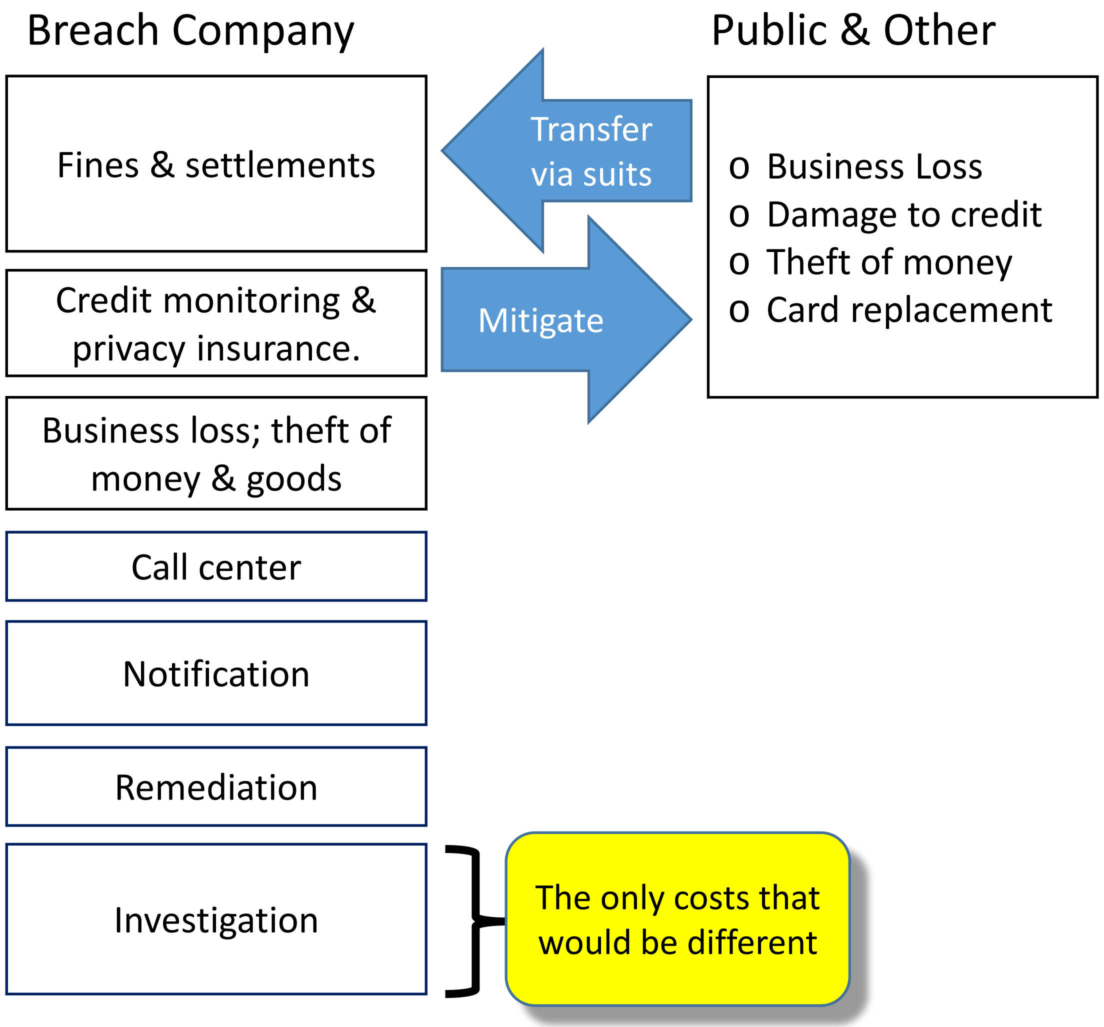
Interpretation

- Based upon Variables
- Based upon Confidence Interval

Interpretation

Condensing of Incident Type

Final Model	Initially Modeled
Malicious Outsider	Malicious Outsider
Other	Malicious Insider
	Lost/Stolen
	Accident



Breach Company

Fines & settlements

Credit monitoring & privacy insurance.

Business loss; theft of money & goods

Call center

Notification

Remediation

Investigation

Transfer via suits

Mitigate

Public & Other

- Business Loss
- Damage to credit
- Theft of money
- Card replacement

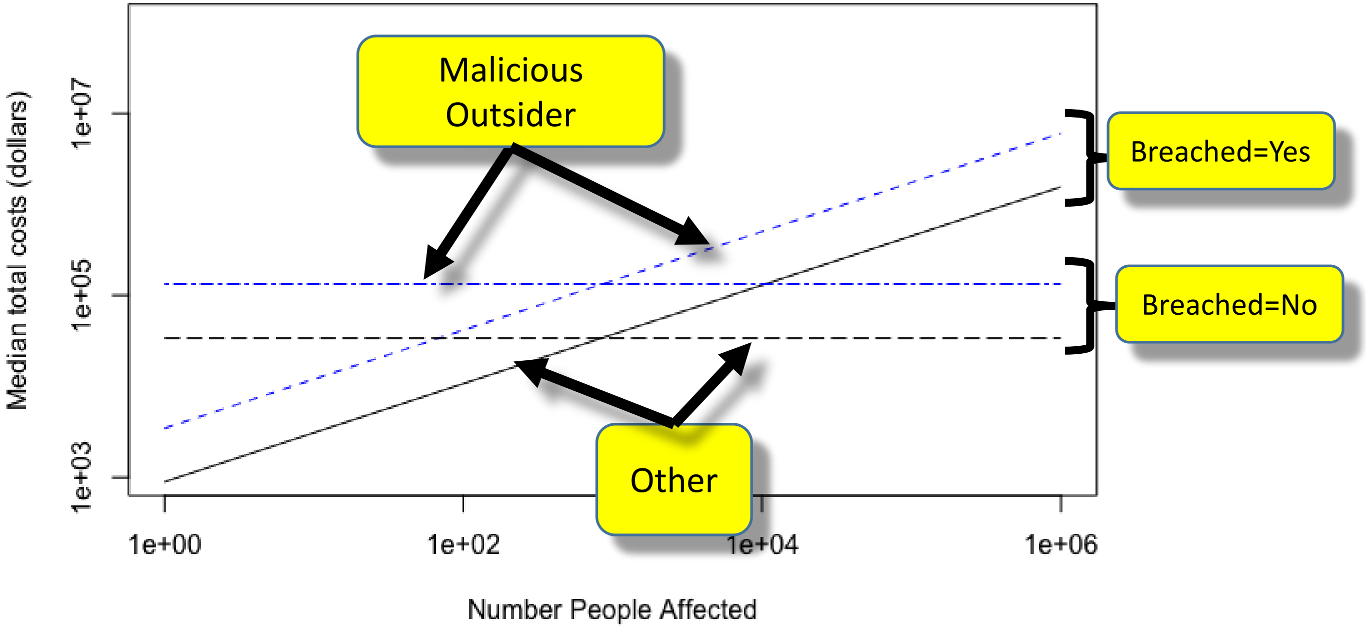
The only costs that would be different

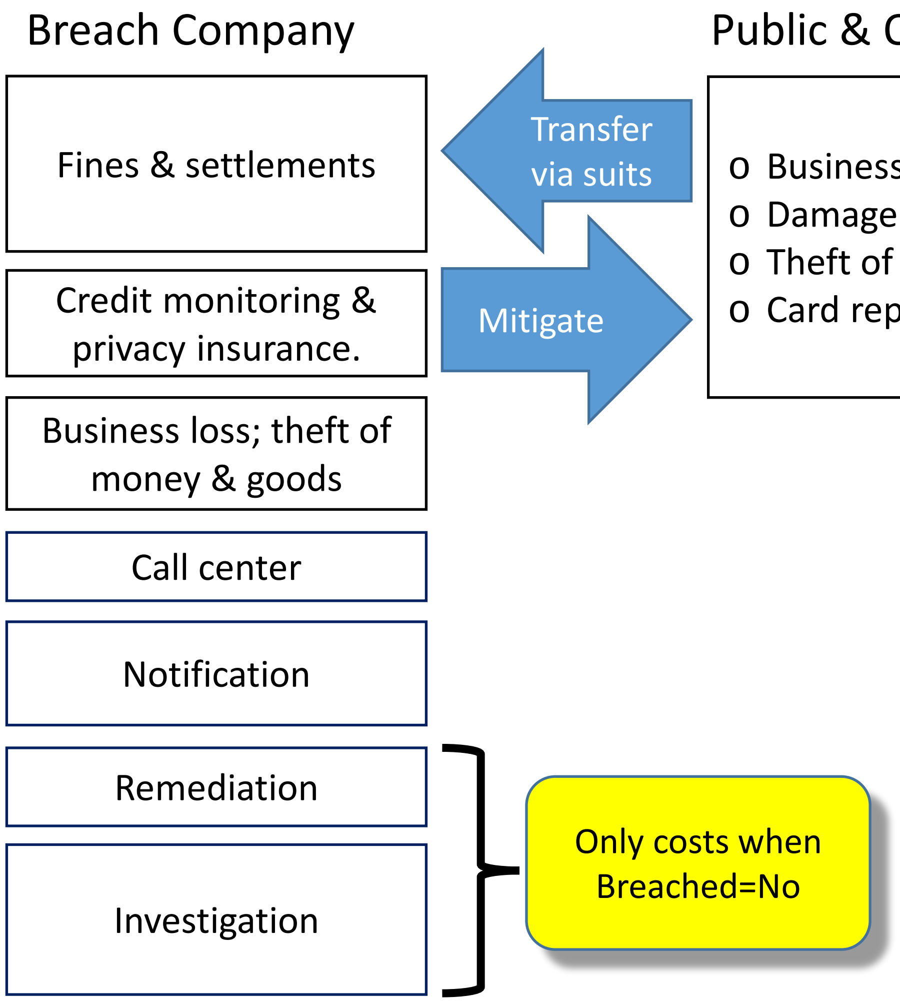
Interpretation

Based upon Variables

Breached=Y vs Breached=N

Same cost relationship between MO and Other suggests investigation cost is an importance difference between MO and Other





Breach Company

Public & C

Fines & settlements

Transfer via suits

Mitigate

Credit monitoring & privacy insurance.

Business loss; theft of money & goods

Call center

Notification

Remediation

Investigation

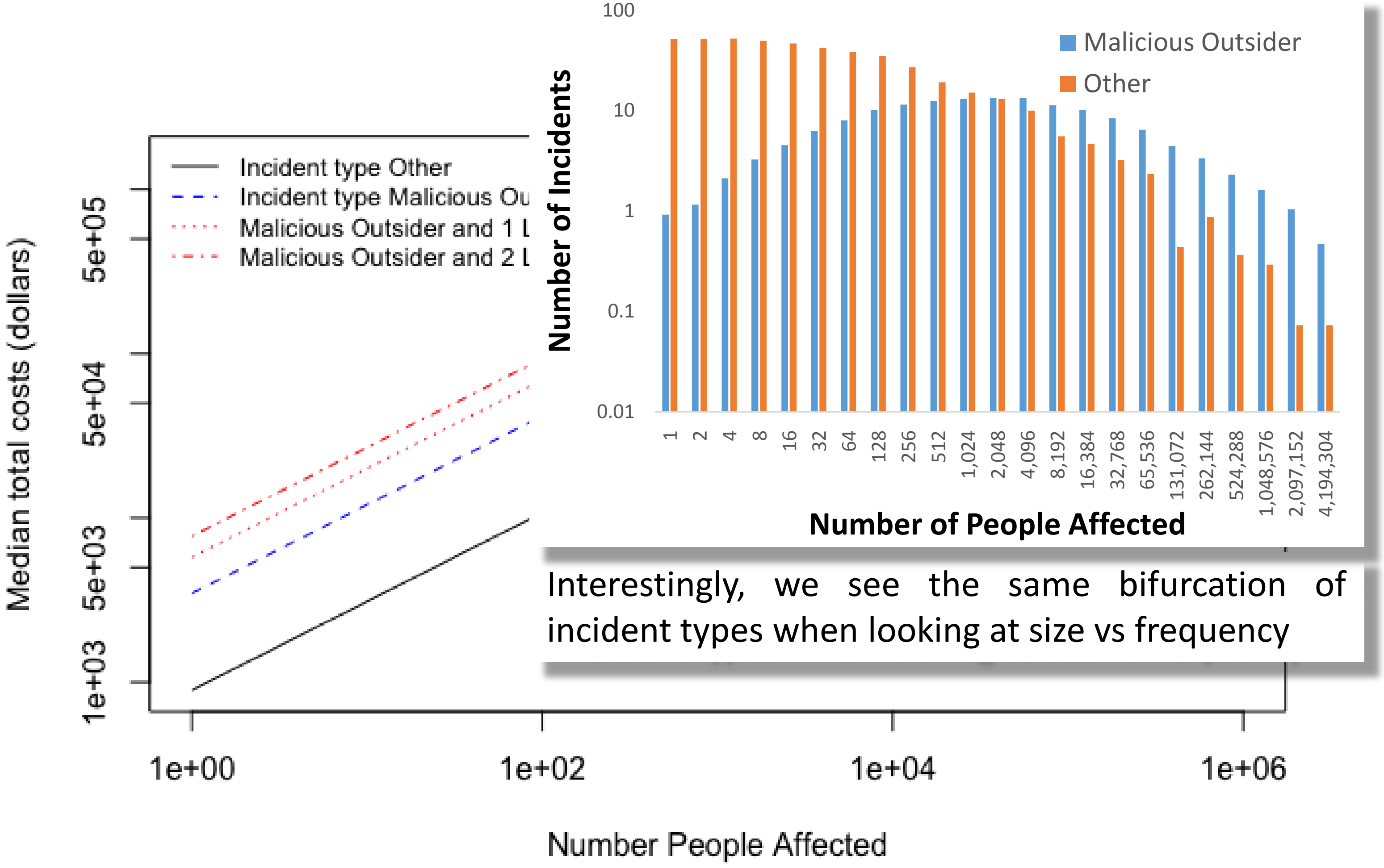
Only costs when Breached=No

- Business
- Damage
- Theft of
- Card rep

Overview

Breached=Y

- Two types of PII breach:
  - *Malicious Outsider*
  - *Other*
- *Malicious Outsider* is 4x costlier
- Cost increases by sqrt of people affected
- One lawsuit doubles the cost



Interestingly, we see the same bifurcation of incident types when looking at size vs frequency

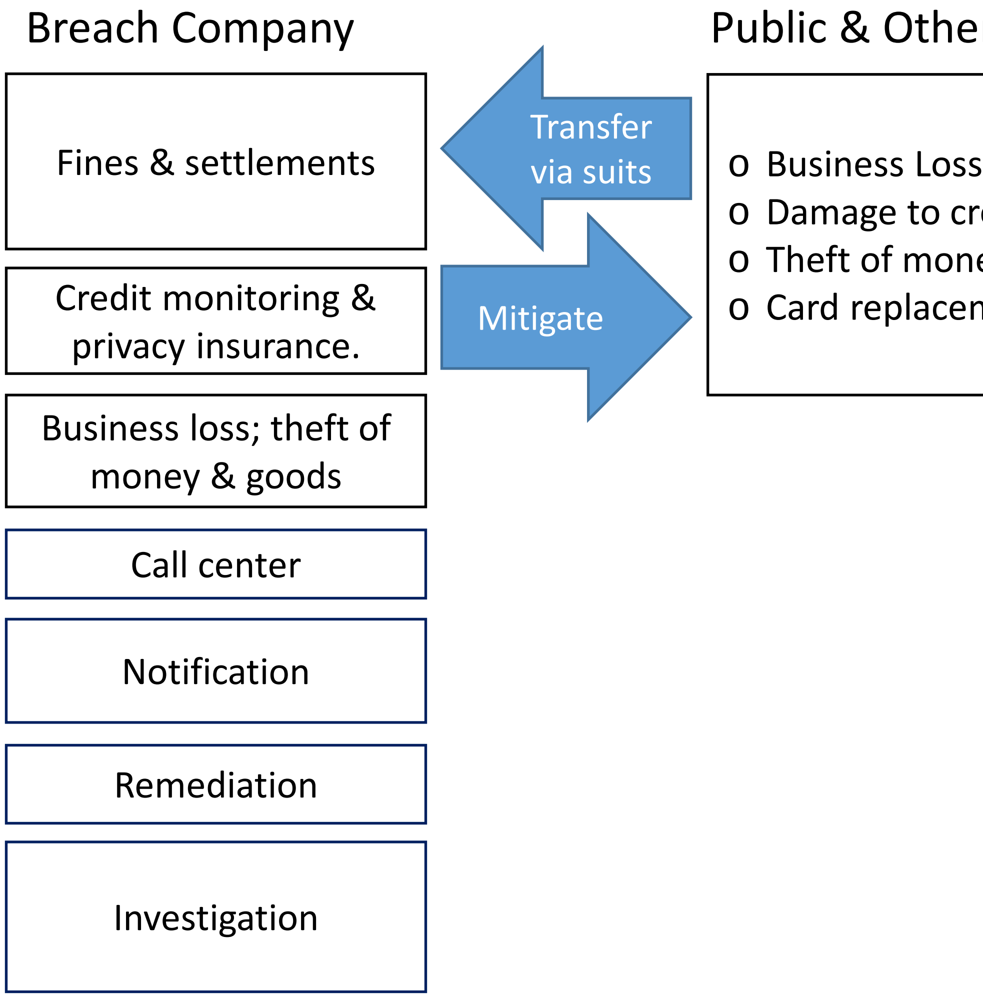
$$Cost_{MO} = 4{,}500 \times (1 + Lawsuits)^{0.7} \times \sqrt{Affected}$$

$$Cost_{Other} = 1{,}100 \times (1 + Lawsuits)^{0.7} \times \sqrt{Affected}$$

Interpretation

Based upon Variables

Variable	Importance	Interpretation, Guide the Expert
Year	Small	We can draw upon lessons learned from past data breaches.
NAICS=52	Small	We can draw upon lessons learned across industries. Look towards costs that are the same across industries: investigation, notification, credit monitoring, call center
Company Demographic	Small	Suggests reputation damage may be less important
Data.Breached	**Large**	Consistent with Investigation cost being significant when variable is considered with Incident Type,
Data Type	Small	Suggests damage to customers is small. Look towards costs that are independent of data exposed: investigation, notification, call center.
Incident Type	**Large**	Consistent with investigation costs being significant.
Affected Count	**Large**, (Sqrt)	Should not use a constant multiplier per person affected since there is efficiency with scale. Don’t use record count, use people affected. Consistent with Investigation and Notification costs being major costs.
Lawsuits	**Large**	Focus on reducing probability of lawsuits for large data breaches. Most costs will be experienced over several years.



Breach Company

Fines & settlements

Credit monitoring & privacy insurance.

Business loss; theft of money & goods

Call center

Notification

Remediation

Investigation

Transfer via suits

Mitigate

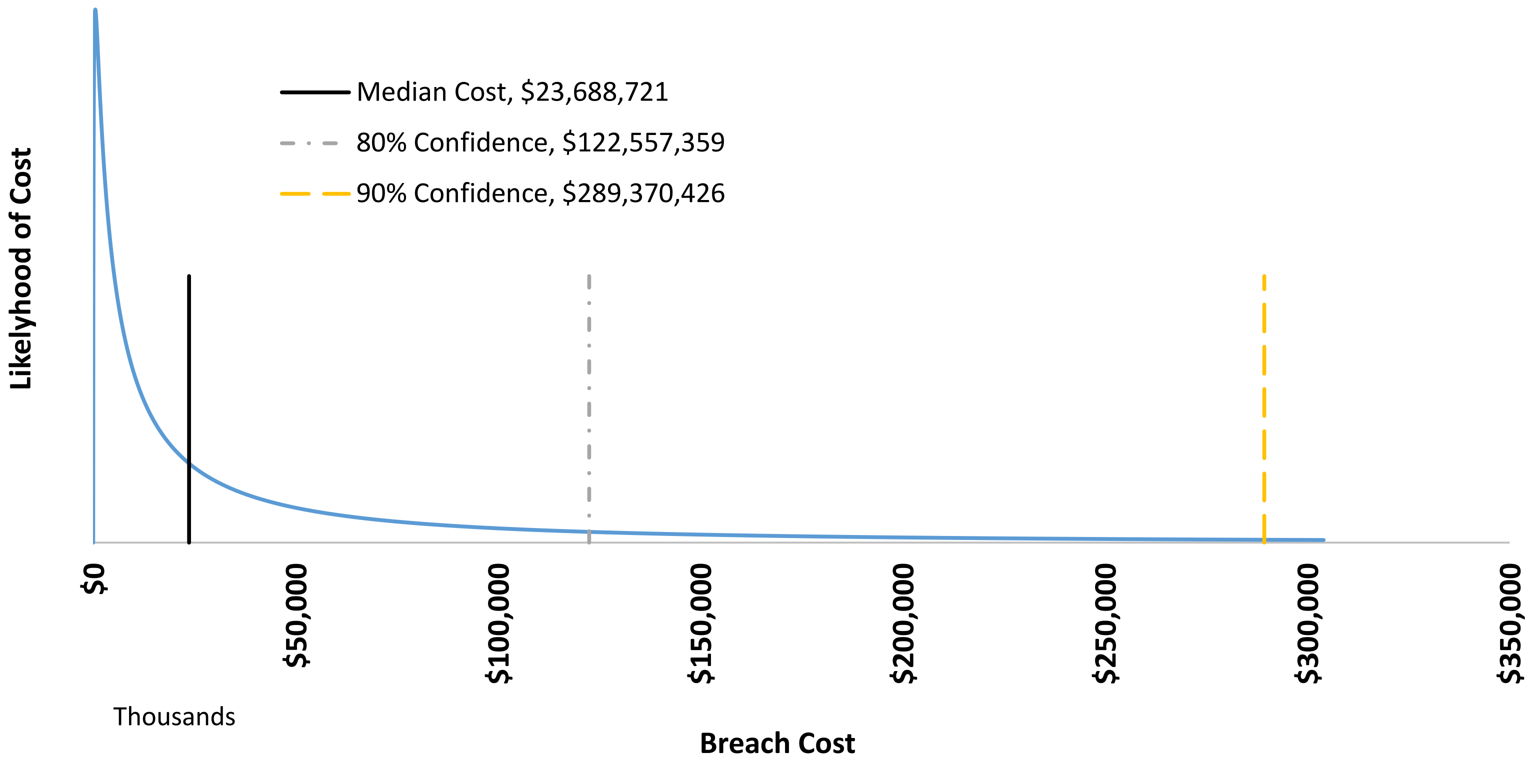
Public & Othe

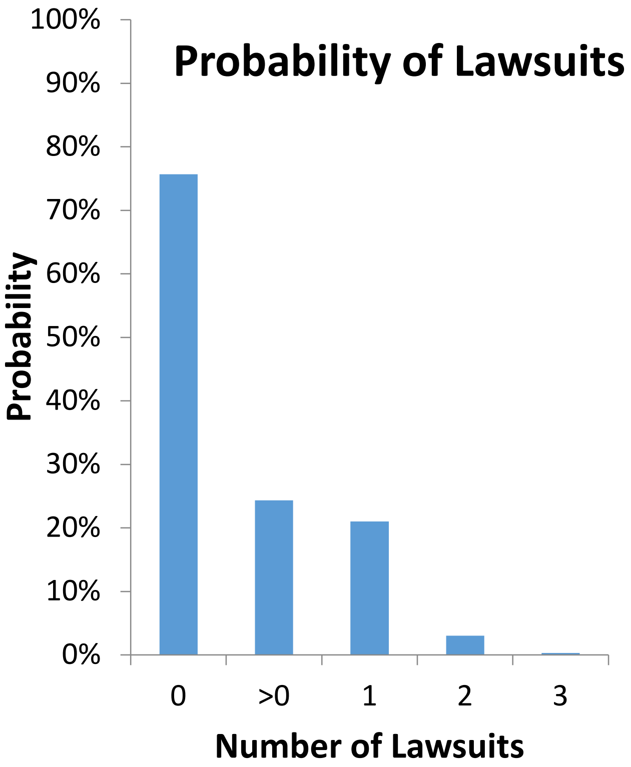
- Business Loss
- Damage to cr
- Theft of mone
- Card replacen

Interpretation

Based upon Confidence Interval

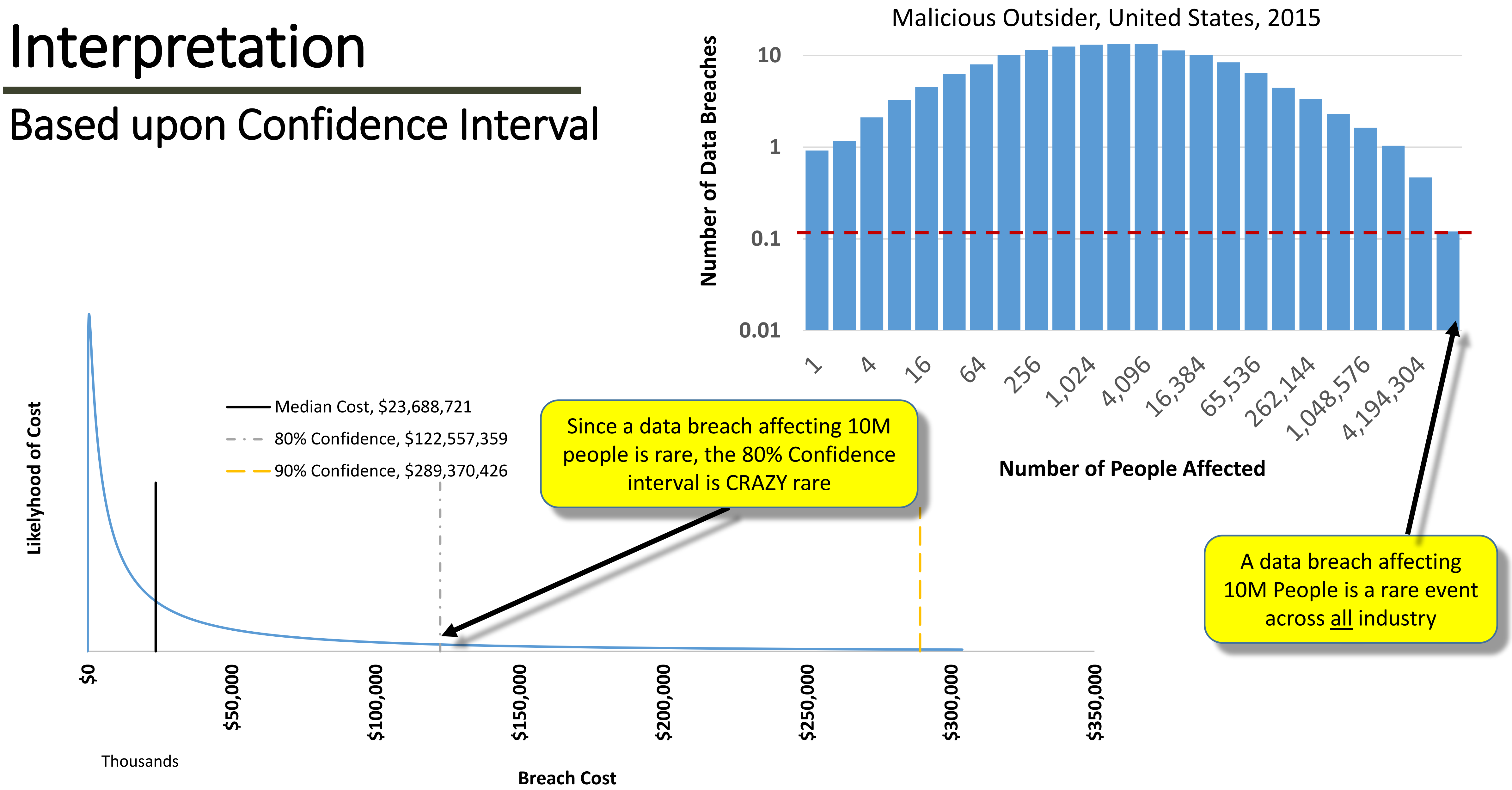
Cost of a Data Breach Affecting 10M People, caused by Malicious Outsider





Interpretation

Based upon Confidence Interval



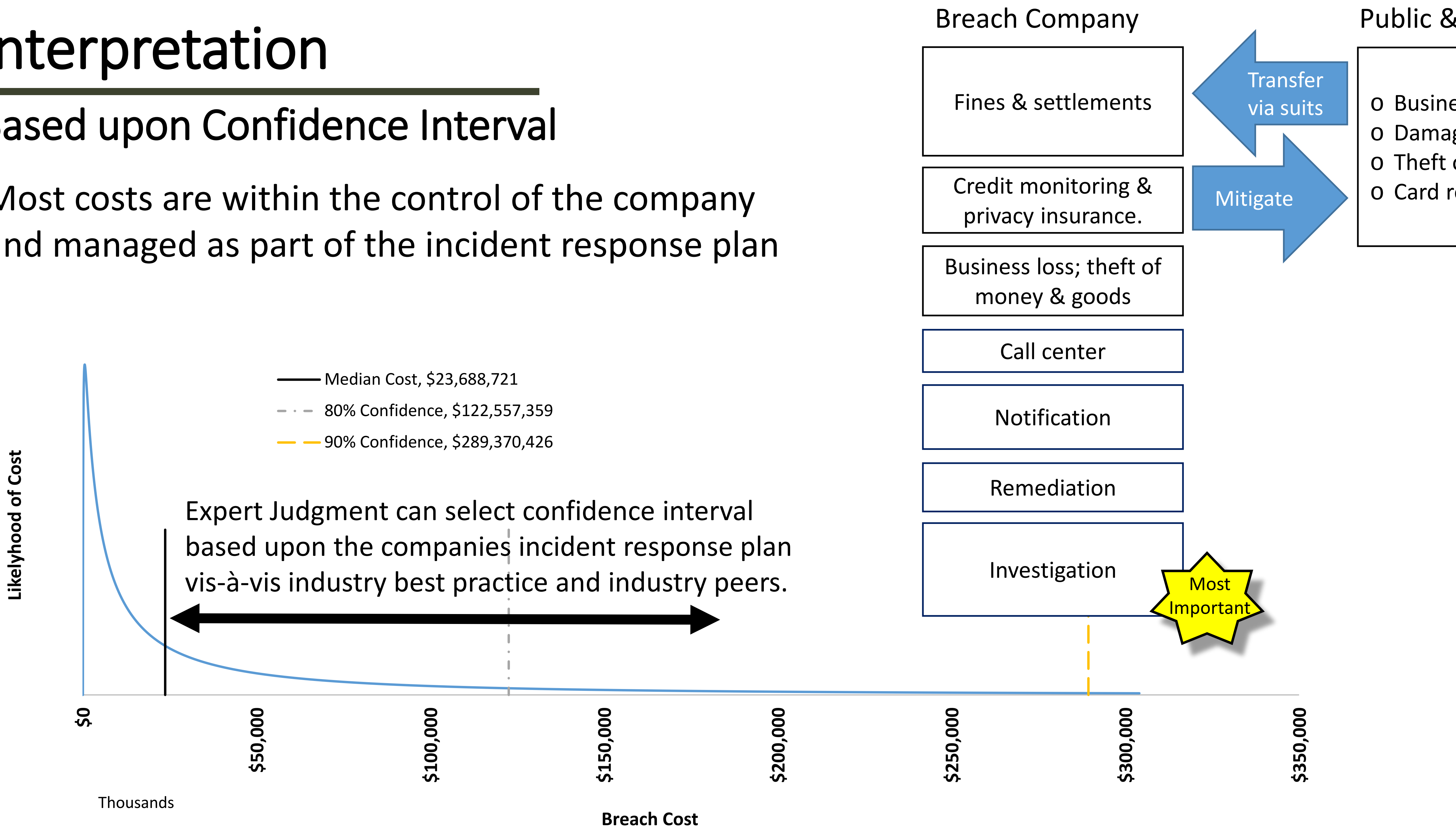
Since a data breach affecting 10M people is rare, the 80% Confidence interval is CRAZY rare

A data breach affecting 10M People is a rare event across all industry

Interpretation

Based upon Confidence Interval

Most costs are within the control of the company and managed as part of the incident response plan



Expert Judgment can select confidence interval based upon the companies incident response plan vis-à-vis industry best practice and industry peers.

Interpretation

Based upon Confidence Interval

***Incident Response* best practice:**

- All access logs turned on
- Access logs saved in a read-only manner
- Access logs saved in a uniform format
- Supporting policies, procedures, training and records
- Tools to aid investigation: Carbon Black, end-point detection and response technology
- Experienced third party evaluation of readiness

Conclusion

- It is possible to develop a model that characterizes the cost of a PII data breach
- The forecasting accuracy of such a model is acceptable, over a large range of *Affected Count* and incident types
- It is important to characterizes incident type independent of methods used to cause the data breach:
  - Malicious Insider
  - Malicious Outsider
  - Accidents
  - Lost Stolen
- Both variables eliminated and variables kept can inform expert judgment
- Expert judgment can assess most reasonable confidence interval by evaluating incident response plan