

Forecasting Data Breaches



ISACA®

Philadelphia Chapter

ThomasL@VivoSecurity.com

Agenda

ThomasL@VivoSecurity.com

- What is probability
- Developing a model
- What is the model; what does it **teach**.
- What is the model-risk
- The most effective way to reduce data breach

What is Probability

How we sell security

What business wants to know

$$\text{Risk} = \text{Impact} \times \text{Probability}$$

After controls are deployed

$$\text{Residual-risk} = \text{Impact} \times \text{Residual-probability}$$

$$\text{Enterprise residual-risk} = \text{Impact} \times \sum \text{Residual-probability}$$

This talk

$$\$1,000,000 \times 0.5 = \$500,000$$

$$\$1,000,000 \times 0.0001 = \$100$$

$$\$1,000,000 \times (0.0001 + 0.0003) = \$400$$

Another way the breach can happen, and a control has been deployed

What is Probability

4% (annual probability)



0.04

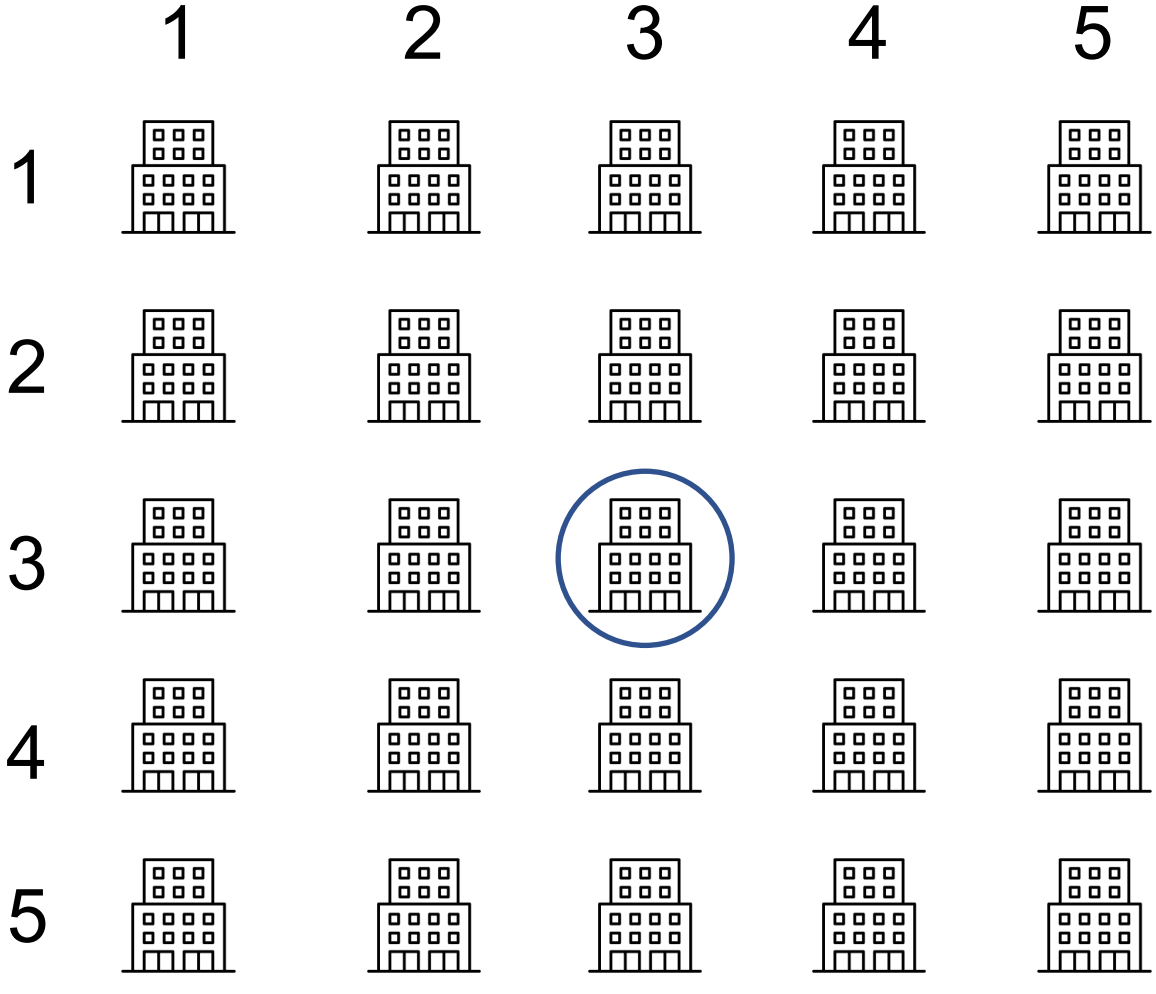


Once in 25-years ($1 \div 0.04$)



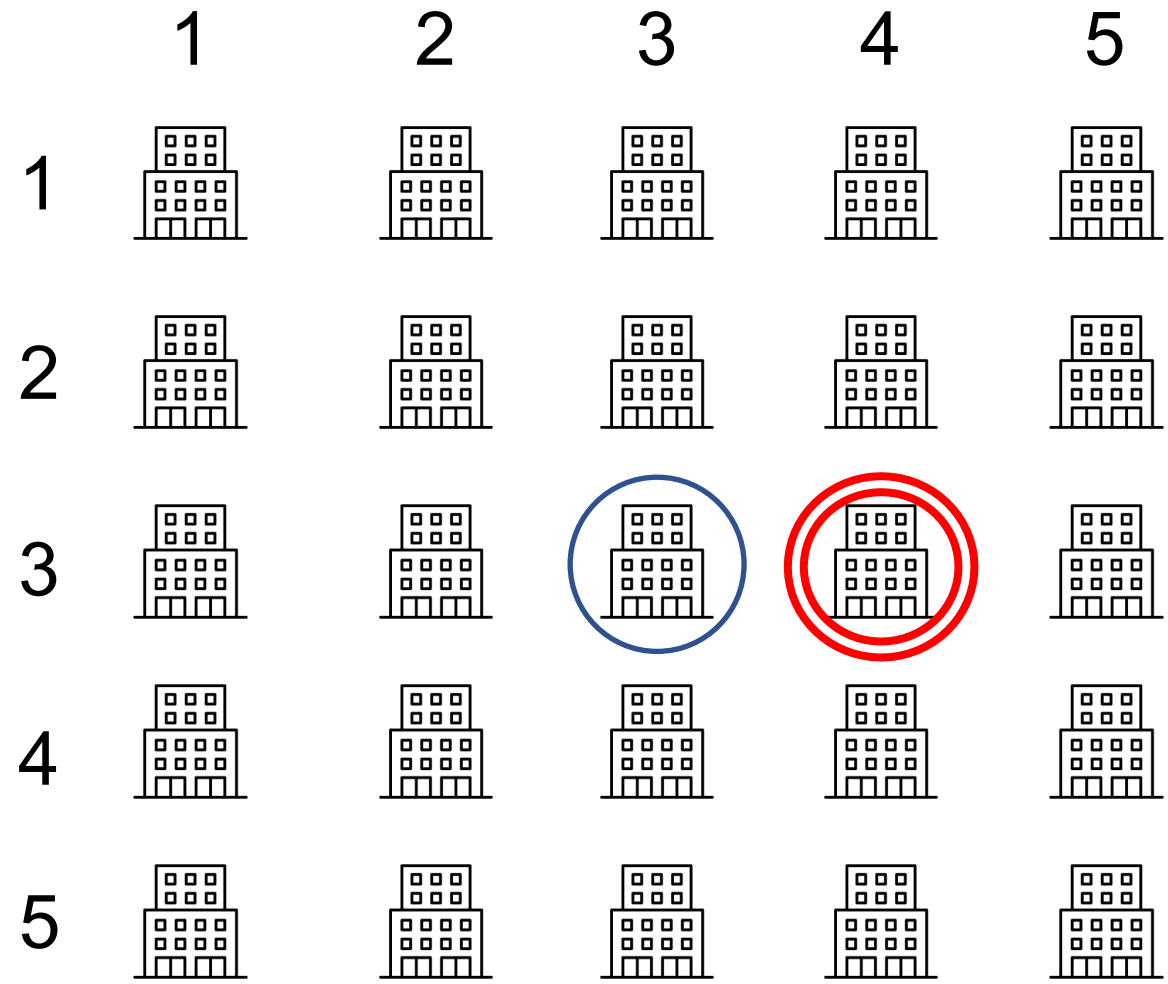
All indicate the same probability

What once in 25-years (4%) really means



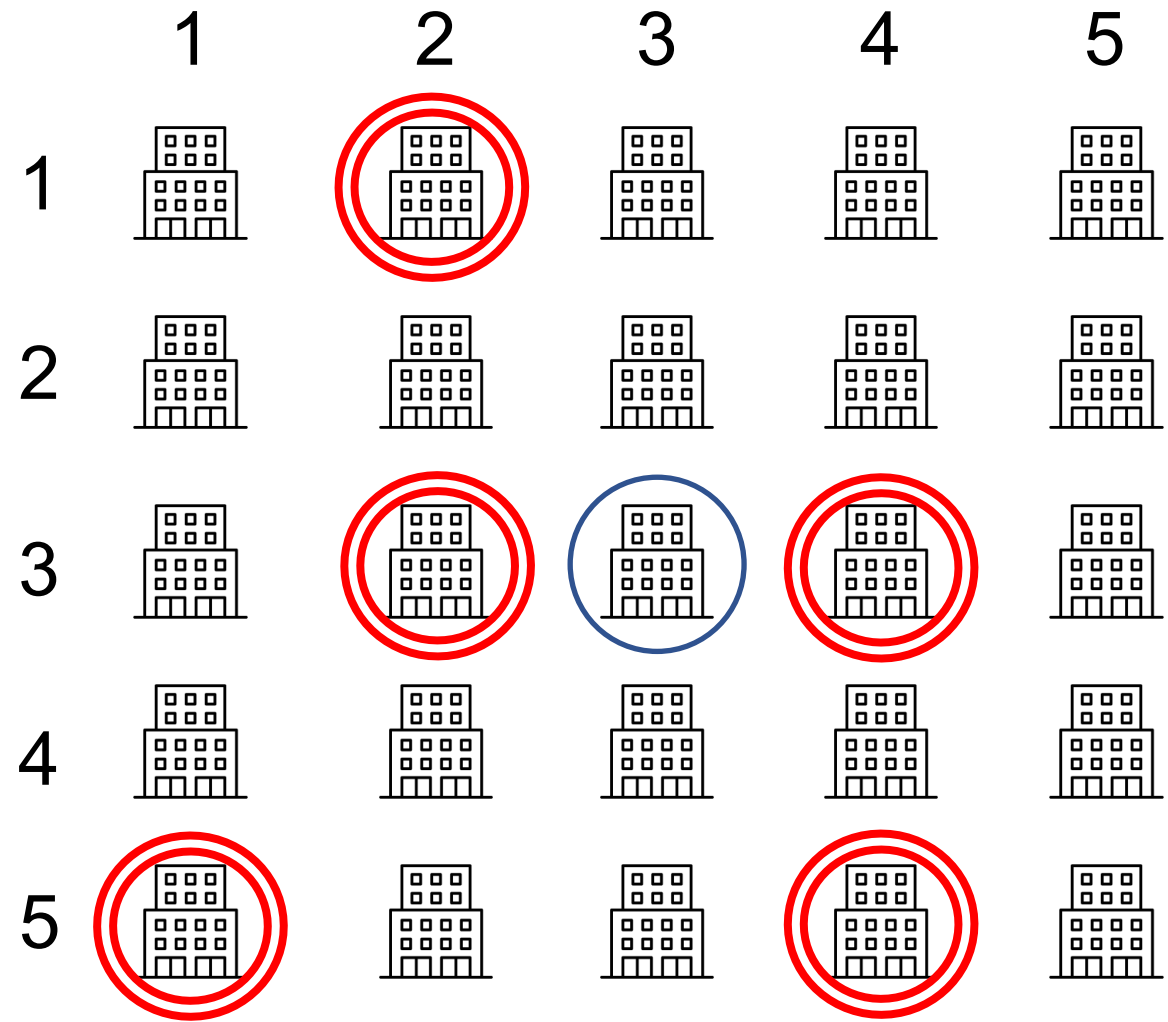
25 companies
each with a
4% probability

What once in 25-years (4%) really means



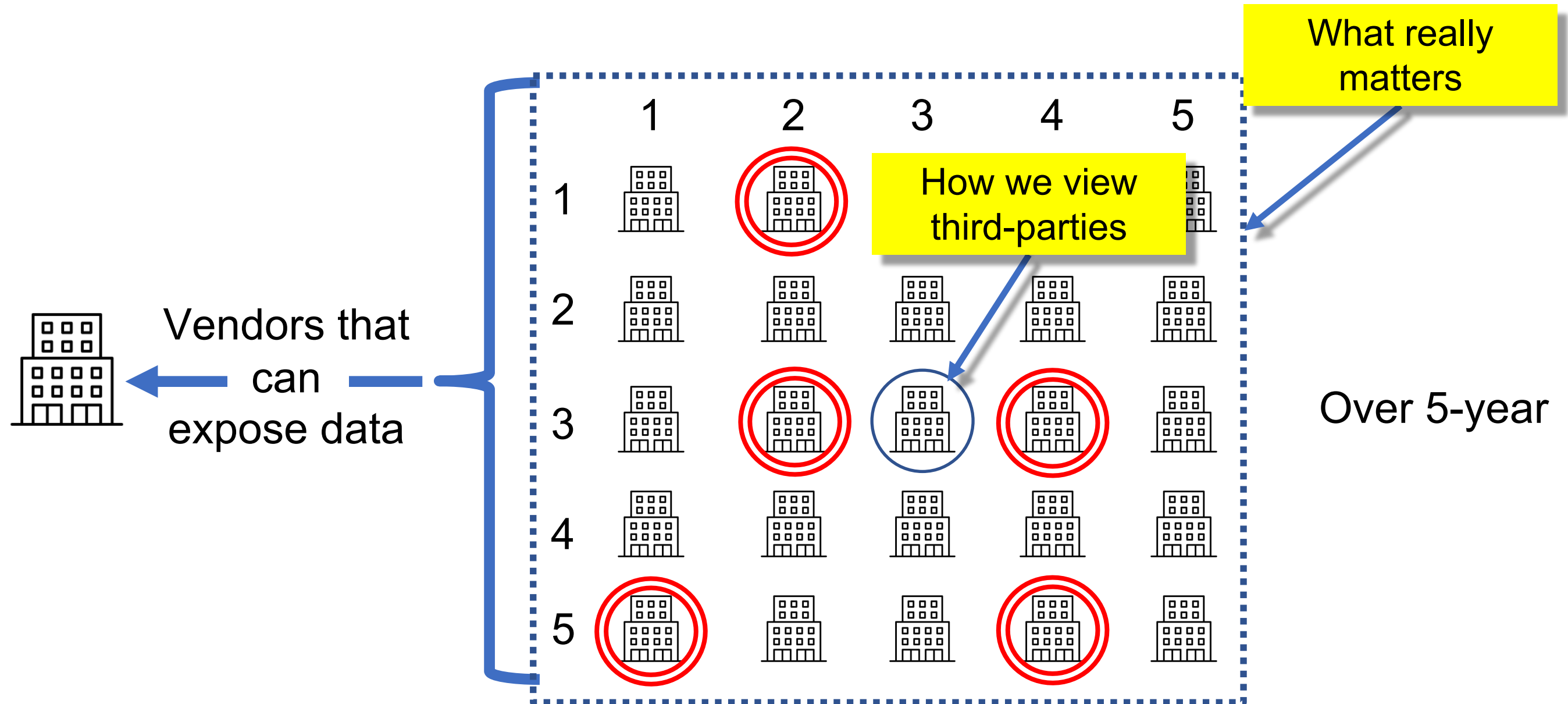
Over 1-year

What once in 25-years (4%) really means



Over 5-year

What once in 25-years (4%) really means



How to Develop a Model

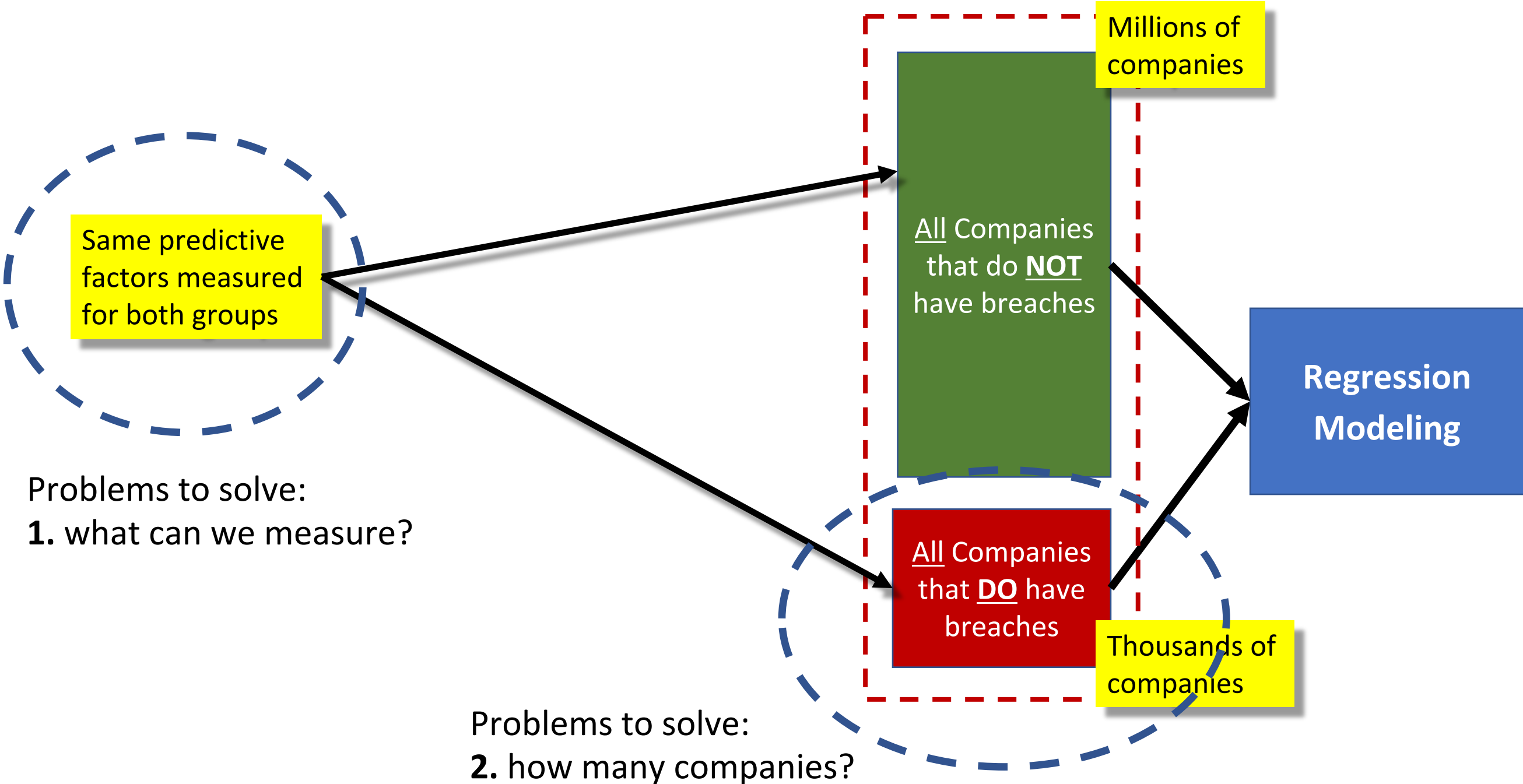
Finding the differences between all companies that did and did not experience data breach

Types of PII Data Breaches

Incident	Description	Examples
Malicious Outsider	Any attack by someone <u>unknown</u> to the company, that exposes PII data	<ul style="list-style-type: none">• Phishing attack• Entry through an unpatched vulnerability in DMZ• Malware
Malicious Insider	Any PII theft by someone <u>known</u> to the company, including employees, ex-employees and vendors	<ul style="list-style-type: none">• Unauthorized data access• PII theft by employee exiting company• PII theft from a call center
Accident	Any kind of accident within the company, or by company vendors, that exposes PII data	<ul style="list-style-type: none">• Email/mail PII data to wrong person or company• Placing SSN on an envelope• Deploy new software that allows unauthorized access• Failing to erase the disk of a discarded computer
Lost or Stolen	Any kind of lost or stolen device that exposes PII data	<ul style="list-style-type: none">• Laptop or thumb drive stolen from a car, house or offices• Magnetic tape that is lost in transit• Computer or backup drive lost track of in an office move• A misplaced thumb drive with PII data

How to predict data breach

Model all breach and all no-breach companies



How we solved problem 2

How many data breaches are there?

Sources of Data Breaches

Company	State	Maryland	Incident	Data	3rd Party
Ransom Memorial Hospital	KS	2	MO	PHI	No
HealthEquity, Inc. (attached)	UT	46	MO	PHI	Yes
Aflac	GA	78	MO	PHI	Yes
Jemison Internal Medicine,	AL	1	MO	PHI	No
Ohio Living	OH	1	MO	PHI	No
Flexible Benefit Service Co	IL	21	MO	PHI	No
Serene Sedation, LLC	MD	4327	MO	PHI	Yes
DecisionHR Holdings, Inc.	FL	2	MO	PHI	Yes
Michael Gruber, DMD, PA	NJ	14	MO	PHI	No
Arkansas Children's Hospital	AR				
Ebbs, Roberts, Head & Davidson	NM				
Cambridge Dental Consultants	NV				
Ruben U. Carvajal, MD	NY				
RISE Wisconsin, Inc	WI	2	MO	PHI	No
Dr. Robert Carpenter	TX	3	MO	PHI	No
SureFire, LLC	CA	39	MO	CHD	Yes
Capital Integration Systems	NY	181	MO	PFI	No
The Coca-Cola Company	GA	14	MO	PFI	No
Stein Eriksen Lodge Hotel	UT	5	MO	PII	No
Authentic Recovery, LLC	CA	3	MO	PHI	No
Securus Technologies, Inc.	TX	25	MO	PII	No
The Childrens Mercy Hospital	MO	6	MO	PHI	No

Companies with headquarters Across all 50 states

Maryland Data Breaches can be Accurately Predicted

State	Breaches Reported to Maryland		GDP	Maryland		State	Breaches Reported to Maryland		GDP	Maryland
	Observed	Predicted					Observed	Predicted		
AK	5	-3	51,479	4,273		MS	2	3	109,375	1,018
AL	10	8	211,196	780		MT	3	-1	47,079	1,947
AR	3	4	122,704	1,050		NC	22	21	540,497	437
AZ	7	9	326,446	2,277		NE	3	3	119,588	1,153
CA	99	97	2,797,601	2,625		NJ	27	26	602,069	179
CO	15	10	345,233	1,663		NM	1	1	94,211	1,861
CT	8	13	264,510	253		NV	1	2	158,302	2,408
DC	17	12	72,461	38		NY	58	62	1,606,601	188
DE	4	13	135,768	74		OH	28	25	645,747	413
FL	32	35	976,386	744		OK	3	5	188,632	1,319
GA	23	21	563,608	664		OR	10	4	227,155	2,808
IA	5	6	183,930	1,020		PA	40	34	756,269	101
ID	4	-1	72,294	2,379		RI	4	5	59,306	372
IL	45	30	822,540	702		SC	7	9	221,690	570
IN	15	14	352,273	579		TN	15	13	349,569	703
KS	9	4	159,108	1,254		TX	55	57	1,645,136	1,416
KY	5	8	202,175	610		UT	9	3	164,917	2,086
LA	2	8	235,960	1,126		VA	30	24	510,586	147
MA	21	22	542,979	412		VT	2	3	32,545	478
MD	67	65	399,538	0		WA	7	15	524,323	2,764
MI	11	19	508,905	628		WI	13	12	321,373	789
MN	13	12	350,179	1,109		WV	2	5	74,047	365
MO	9	10	303,763	1,060						



Total US Data Breaches can now be Accurately Predicted

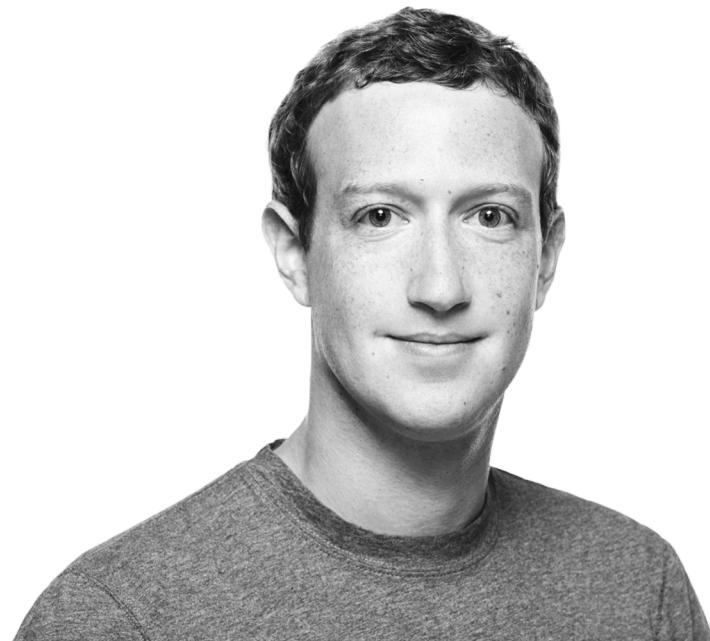
State	Breaches Reported to Maryland		GDP	Distance to Maryland		State	Breaches Reported to Maryland		GDP	Distance to Maryland
	Observed	Predicted					Observed	Predicted		
AK	5	52	51,479	0		MS	2	55	109,375	0
AL	10	58	211,196	0		MT	2	55	47,079	0
AR	3	55	122,704	0		NC	22	70	540,497	0
AZ	7	62	326,446	0		NE	3	55	119,588	0
CA	99	151	2,797,601	0		NJ	27	72	602,069	0
CO	15	63	345,233	0		NM	1	54	94,211	0
CT	8	60	264,510	0		NV	1	56	158,302	0
DC	17	53	72,461	0		NY	58	108	1,606,601	0
DE	4	55	135,768	0		OH	28	74	645,747	0
FL	32	86	976,386	0		OK	3	57	188,632	0
GA	23	71	563,608	0		OR	10	59	227,155	0
IA	5	57	183,930	0		PA	40	78	756,269	0
ID	4	53	72,294	0		RI	4	53	59,306	0
IL	45	80	822,540	0		SC	7	59	221,690	0
IN	15	63	352,273	0		TN	15	63	349,569	0
KS	9	56	159,108	0		TX	55	110	1,645,136	0
KY	5	58	202,175	0		UT	9	57	164,917	0
LA	2	59	235,966	0		VT	2	53	510,586	0
MA	21	70	542,977	0		WA	2	53	32,545	0
MD	67	65	399,533	0		WV	2	53	524,323	0
MI	11	69	508,900	0						
MN	13	63	350,179	0						
MO	9	61	303,763	0						

Set distance to zero

Forecast total breaches for state

Total breaches: 2957
Reported to Maryland: 771





Poll 1

If Russia required reporting of PII data breaches, would you expect:

- A. A similar number of data breaches as California, because Russia develops a lot of high technology
- B. A similar number of data breaches as Texas because both Russia and Texas have oil economies
- C. A similar number of data breaches as Texas because Texas GDP is \$1.6T and Russia GDP was \$1.7T

The Answer is C: Russia and Texas have similar GDP

State	Breaches Reported to Maryland		GDP	Distance to Maryland		State	Breaches Reported to Maryland		GDP	Distance to Maryland
	Observed	Predicted					Observed	Predicted		
AK	5	-3	51,479	4,273		MS	2	3	109,375	1,018
AL	10	8	211,196	780		MT	3	-1	47,079	1,947
AR	3	4	122,704	1,050		NC	22	21	540,497	437
AZ	7	9	326,446	2,277		NE	3	3	119,588	1,153
CA	99	97	2,797,601	2,625		NJ	27	26	602,069	179
CO	15	10	345,233	1,663		NM	1	1	94,211	1,861
CT	8	13	264,510	253		NV	1	2	158,302	2,408
DC	17	12	72,461	38		NY	58	62	1,606,601	188
DE	4	13	135,768	74		OH	28	25	645,747	413
FL	32	35	976,386	744		OK	3	5	188,632	1,319
GA	23	21	563,608	664		OR	10	4	227,155	2,808
IA	5	6	183,930	1,020		PA	40	34	756,269	101
ID	4	-1	72,294	2,379		RI	4	5	59,306	372
IL	45	30	822,540	702		SC	7	9	221,690	570
IN	15	14	352,273	579		TN	15	13	349,569	703
KS	9	4	159,108	1,254		TX	55	57	1,645,136	1,416
KY	5	8	202,175	610		UT	9	3	164,917	2,086
LA	2	8	235,960	1,126		VA	30	24	510,586	147
MA	21	22	542,979	412		VT	2	3	32,545	478
MD	67	65	399,538	0		WA	7	15	524,323	2,764
MI	11	19	508,905	628		WI	13	12	321,373	789
MN	13	12	350,179	1,109		WV	2	5	74,047	365
MO	9	10	303,763	1,060						

Technology

Agriculture

Oil



The Model

Headcounts can predict data breach

Predictors

Headcount	Description
CISA/IT	Audit (3rd-line) Certified Information Systems Auditor
CISSP/IT	Technical (2nd-line) Certified Information Systems Security Professional
MCSA/IT	Vendor (1st-line) Microsoft Certified Solutions Associate
Employees	Total employees



AUDITOR

People side of cybersecurity





Technical side of cybersecurity



Vendor side of cybersecurity
(consider also AWS, CISCO, etc.)

Predictors

Headcount	Description	Effect		
			Small Breach	Large Breach
	CISA/IT Audit (3rd-line) Certified Information Systems Auditor	Decrease Probability (decreased risk)	Strong but Saturates	Very strong, does NOT saturate
	CISSP/IT Technical (2nd-line) Certified Information Systems Security Professional		Strong but Saturates	Very strong, does NOT saturate
	MCSA/IT Vendor (1st-line) Microsoft Certified Solutions Associate		Modest and saturates	
	Employees	Increase Probability (increased risk)	Moderate	Moderate

Predictors

Headcount	Description	Effect		
			Small Breach	Large Breach
CISA/IT	Audit (3rd-line) Certified Information Systems Auditor	Decrease Probability (decreased risk)	Strong but, diminishing return	Very strong, no diminishing return
CISSP/IT	Technical (2nd-line) Certified Information Systems Security Professional		Strong but, diminishing return	Very strong, no diminishing return
MCSA/IT	Vendor (1st-line) Microsoft Certified Solutions Associate		Modest and diminishing return	Weak
Employees	Total employees	Increase Probability (increased risk)	Moderate	Moderate

Equal Effectiveness

Predictors

Other observations

- RHCE (Red Hat Certified Engineer) a Linux certification increases probability for data breach and is also in the model.
- Many other certifications were tried and found to be predictive by themselves, but did not increase the accuracy when combined with CISSP.
- Counting employees with certifications was better than simply a count of people in cybersecurity.

An analogy: predicting family size

Family Size ←



We might find the amount of milk is a predictor

This is like the CISSP

Perhaps this is measuring older children



We might find cereal is also.

This is like other certs – important but don't improve accuracy

But adding cereal with milk does not improve accuracy



We might find diapers is also a good predictor.

This is like CISA, measuring something different from CISSP

Perhaps this is capturing children not yet drinking milk



Adding diapers with milk is more accurate

An analogy: predicting family size

Best (most accurate) model

Note: there are lots of other things in the cart too! These other things are needed for a healthy family.

But all we need to measure is the milk and diapers.

Family Size



An analogy: predicting family size

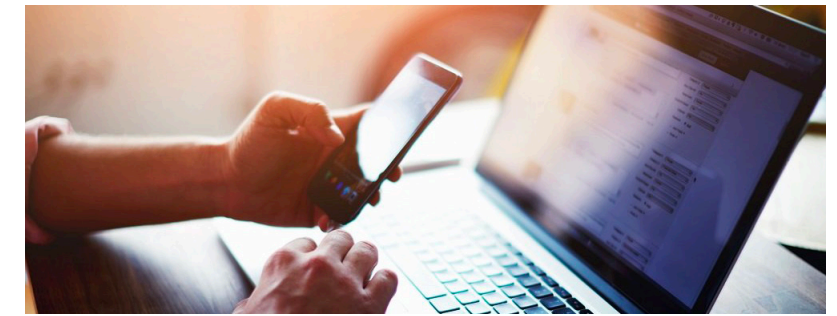
Best (most accurate) model



Certification-Handicapping

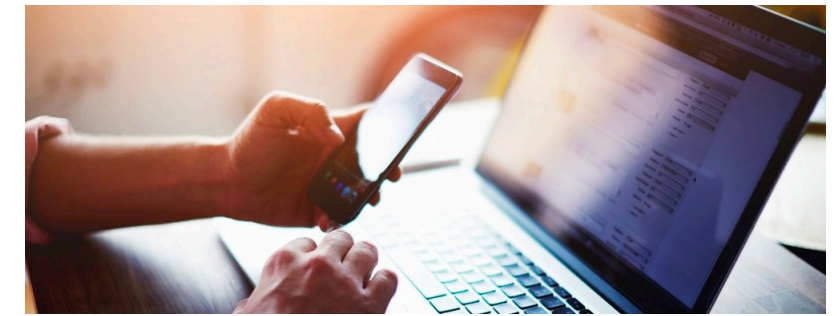
Predicting which company is best, without the direct use of the model

Take a picture of this with your phone



Measure	Decrease Probability (Decrease Risk)	
	Small Breach	Large Breach
CISA/IT	Strongly	Very Strongly
CISSP/IT	Strongly	Very Strongly
MCSA/IT	Moderate	Weak

Take a picture of this with your phone



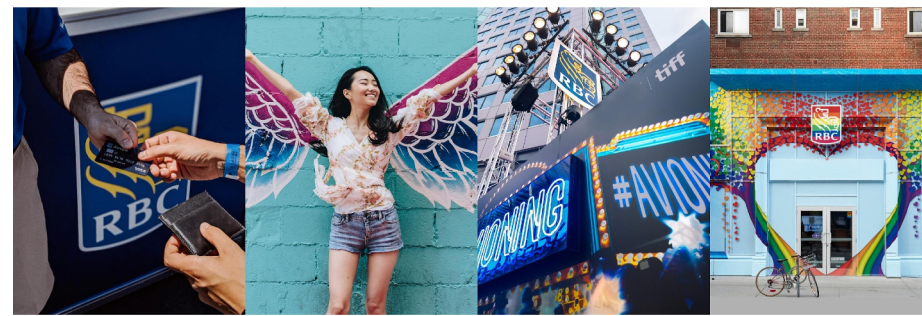
Measure	Increase Probability (Increase Risk)	
	Small Breach	Large Breach
Employees	Moderate	Moderate
RHCE/IT	Modest	Weak

Poll 2: According to the model, which has the lowest probability for a data breach?

Note: all of these banks are very good!

A

Headcount	Measure
82K	Employees
22	RHCE
204	CISSP
181	CISA
90	MCSA
7.5K	IT



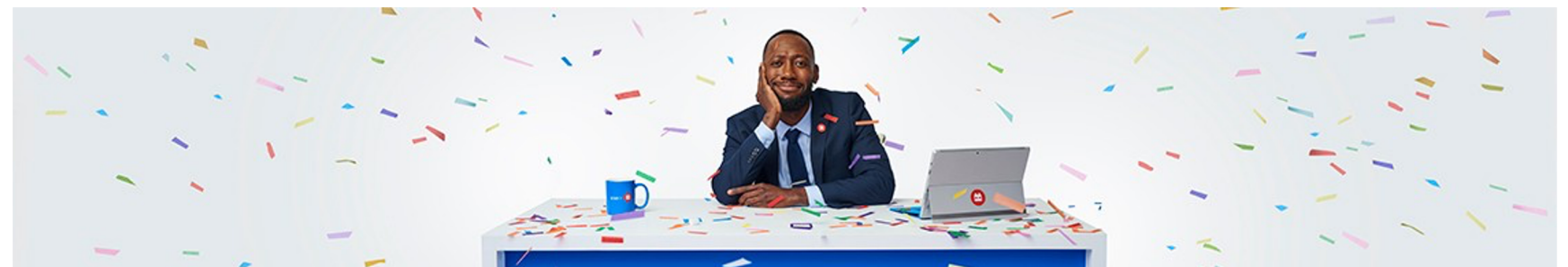
B

Headcount	Measure
89K	Employees
20	RHCE
333	CISSP
275	CISA
95	MCSA
7.5K	IT



C

Headcount	Measure
49K	Employees
21	RHCE
181	CISSP
157	CISA
74	MCSA
4.5K	IT



The answer is: BMO (C)

A

Headcount	Measure
82K	Employees
22	RHCE
204	CISSP
181	CISA
90	MCSA
7.5K	IT



	1+	10+	100+	1K+	10K+	100K+	1M+	10M+	100M+
Probability	26%	26%	25%	20%	8.2%	3.0%	0.84%	0.075%	0.011%
Years	3.8	3.8	4.0	5.0	12	33	119	1333	9091

Breach sizes

Annual probabilities

Years between breaches

B

Headcount	Measure
89K	Employees
20	RHCE
333	CISSP
275	CISA
95	MCSA
7.5K	IT



	1+	10+	100+	1K+	10K+	100K+	1M+	10M+	100M+
Probability	22%	22%	20%	15%	4.5%	1.4%	0.30%	0.019%	0.0022%
Years	4.5	4.5	5.0	6.7	22	71	333	5263	45455

C

Headcount	Measure
49K	Employees
21	RHCE
181	CISSP
157	CISA
74	MCSA
4.5K	IT



	1+	10+	100+	1K+	10K+	100K+	1M+	10M+	100M+
Probability	12%	12%	12%	8.3%	2.7%	0.83%	0.19%	0.012%	0.0015%
Years	8.3	8.3	8.3	12	37	120	526	8333	66667

The answer is: BMO (C)

A

Headcount	Measure
82K	Employees
22	RHCE
204	CISSP
181	CISA
90	MCSA
7.5K	IT



	1+	10+	100+	1K+	10K+	100K+	1M+	10M+	100M+
Probability	26%	26%	25%	20%	8.2%	3.0%	0.84%	0.075%	0.011%
Years	3.8	3.8	4.0	5.0	12	33	119	1333	9091

Large companies have small data breaches

B

Headcount	Measure
89K	Employees
20	RHCE
333	CISSP
275	CISA
95	MCSA
7.5K	IT



	1+	10+	100+	1K+	10K+	100K+	1M+	10M+	100M+
Probability	22%	22%	20%	15%	4.5%	1.4%	0.30%	0.019%	0.0022%
Years	4.5	4.5	5.0	6.7	22	71	333	5263	45455

C

Headcount	Measure
49K	Employees
21	RHCE
181	CISSP
157	CISA
74	MCSA
4.5K	IT



	1+	10+	100+	1K+	10K+	100K+	1M+	10M+	100M+
Probability	12%	12%	12%	8.3%	2.7%	0.83%	0.19%	0.012%	0.0015%
Years	8.3	8.3	8.3	12	37	120	526	8333	66667

Smaller companies can more easily control small breaches

The answer is: BMO (C)

Probability for large data breach is very sensitive to CISSP and CISA and there is no diminishing return.

A

Headcount	Measure
82K	Employees
22	RHCE
204	CISSP
181	CISA
90	MCSA
7.5K	IT



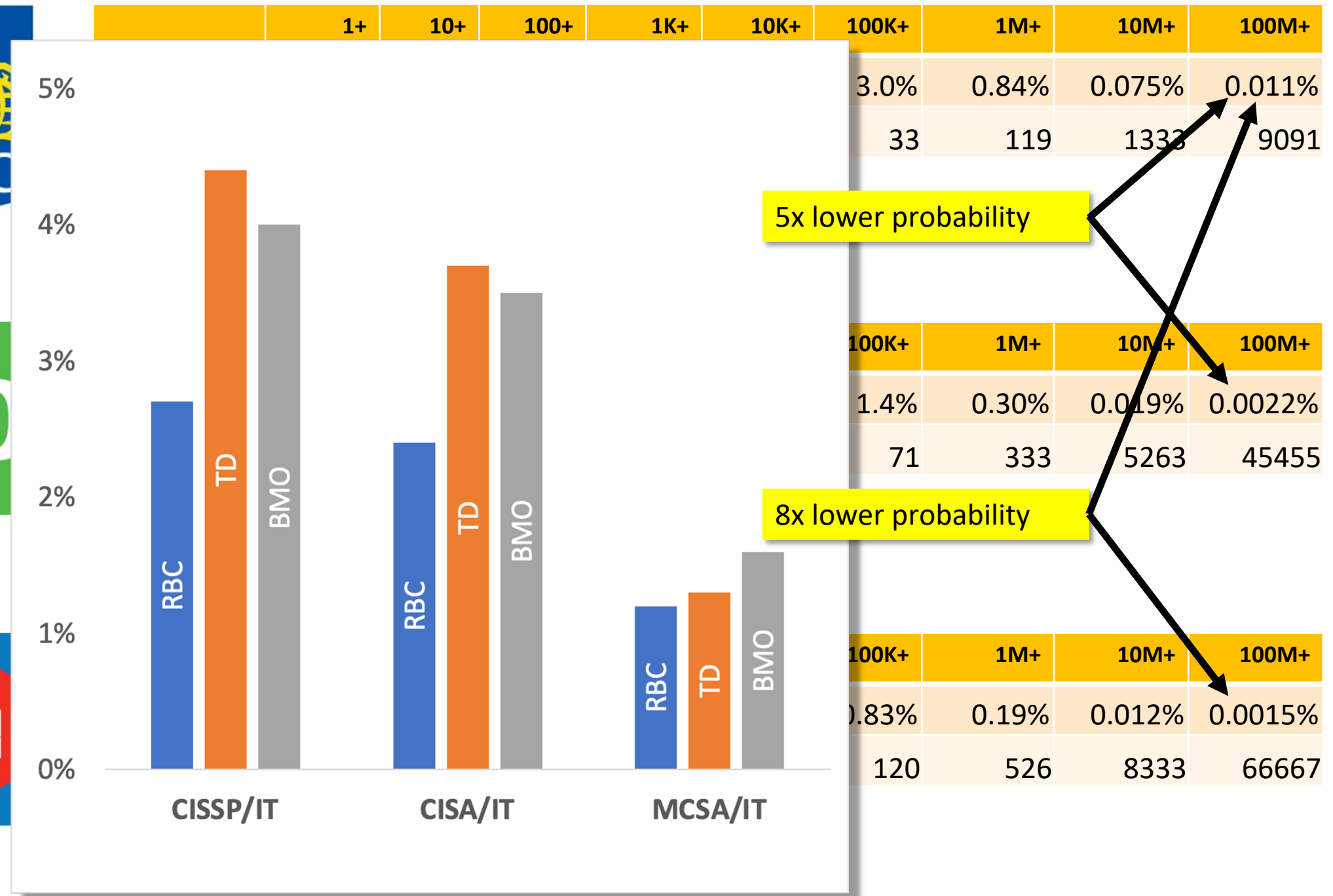
B

Headcount	Measure
89K	Employees
20	RHCE
333	CISSP
275	CISA
95	MCSA
7.5K	IT



C

Headcount	Measure
49K	Employees
21	RHCE
181	CISSP
157	CISA
74	MCSA
4.5K	IT



5x lower probability

8x lower probability

Poll 3: According to the model, which has the lowest probability for a data breach?

A

Headcount	Measure
50,148	Employees
3	RHCE
17	CISSP
10	CISA
6	MCSA
817	IT



B

Headcount	Measure
34,701	Employees
0	RHCE
6	CISSP
1	CISA
3	MCSA
784	IT



The answer is: AutoZone (A)

Large companies have small data breaches.
There is diminishing return with CISSP & CISA

A

Headcount	Measure
50,148	Employees
3	RHCE
17	CISSP
10	CISA
6	MCSA
817	IT



	1+	10+	100+	1K+	10K+	100K+	1M+	10M+	100M+
Probability	23%	23%	23%	20%	9.3%	3.9%	1.2%	0.13%	0.023%
Years	4	4	4	5	11	26	81	742	4350

B

Headcount	Measure
34,701	Employees
0	RHCE
6	CISSP
1	CISA
3	MCSA
784	IT



	1+	10+	100+	1K+	10K+	100K+	1M+	10M+	100M+
Probability	33%	33%	33%	30%	17%	8.2%	3.1%	0.44%	0.089%
Years	3	3	3	3	6	12	33	229	1127

The answer is: AutoZone (A)

Probability for large data breach is very sensitive to CISSP/IT and CISA/IT and there is no diminishing return.

A

Headcount	Measure
50,148	Employees
3	RHCE
17	CISSP
10	CISA
6	MCSA
817	IT



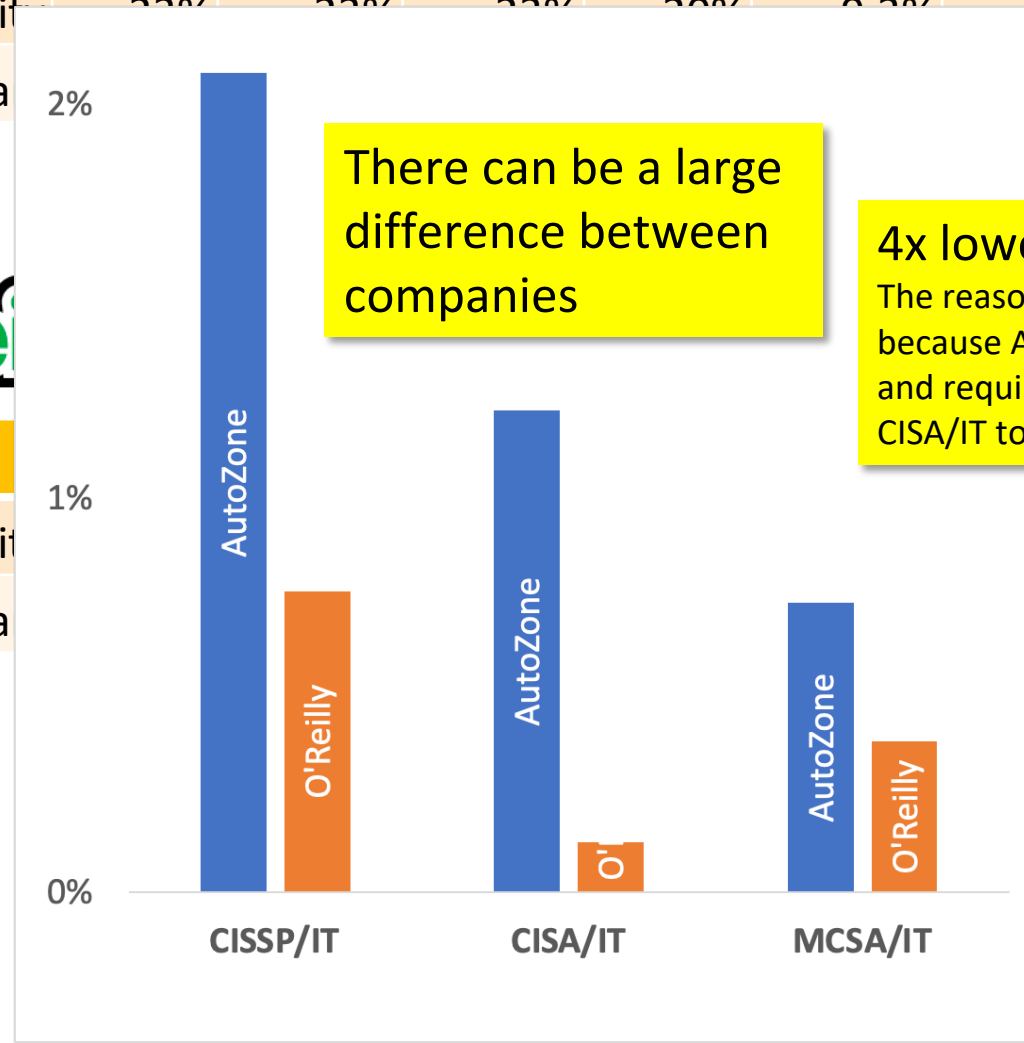
	1+	10+	100+	1K+	10K+	100K+	1M+	10M+	100M+
Probability	22%	22%	22%	22%	0.22%	3.9%	1.2%	0.13%	0.023%
Year	26	81	742	4350					

B

Headcount	Measure
34,701	Employees
0	RHCE
6	CISSP
1	CISA
3	MCSA
784	IT



	10M+	100M+
Probability	8.2%	0.089%
Year	12	1127



MODEL RISK

The risk from using a model to make business decisions

There are standards



Office of the Superintendent of
Financial Institutions Canada

Bureau du surintendant des
institutions financières Canada

Guideline

Subject: Enterprise-Wide Model Risk Management for Deposit-Taking Institutions

Category: Sound Business and Financial Practices

No: E-23

Date: September 2017

This Guideline outlines OSFI's expectations around institutions' establishment of sound policies and practices for an enterprise-wide model risk management framework. It applies to banks, bank holding companies, federally regulated trust and loan companies and cooperative retail associations, and collectively referred to as 'institutions'.



BOARD OF GOVERNORS
OF THE
FEDERAL RESERVE SYSTEM

WASHINGTON, D.C. 20551

DIVISION OF BANKING
SUPERVISION AND REGULATION

SR 11-7

April 4, 2011

**TO THE OFFICER IN CHARGE OF SUPERVISION
AT EACH FEDERAL RESERVE BANK**

SUBJECT: Guidance on Model Risk Management

The Federal Reserve and Office of the Comptroller of the Currency (OCC) are issuing the attached *Supervisory Guidance on Model Risk Management*, which is intended for use by banking organizations and supervisors as they assess organizations' management of model risk. This guidance should be applied as appropriate to all banking organizations supervised by the Federal Reserve, taking into account each organization's size, nature, and complexity, as well as the extent and sophistication of its use of models (as defined and discussed below).

Two ways to test

During model development

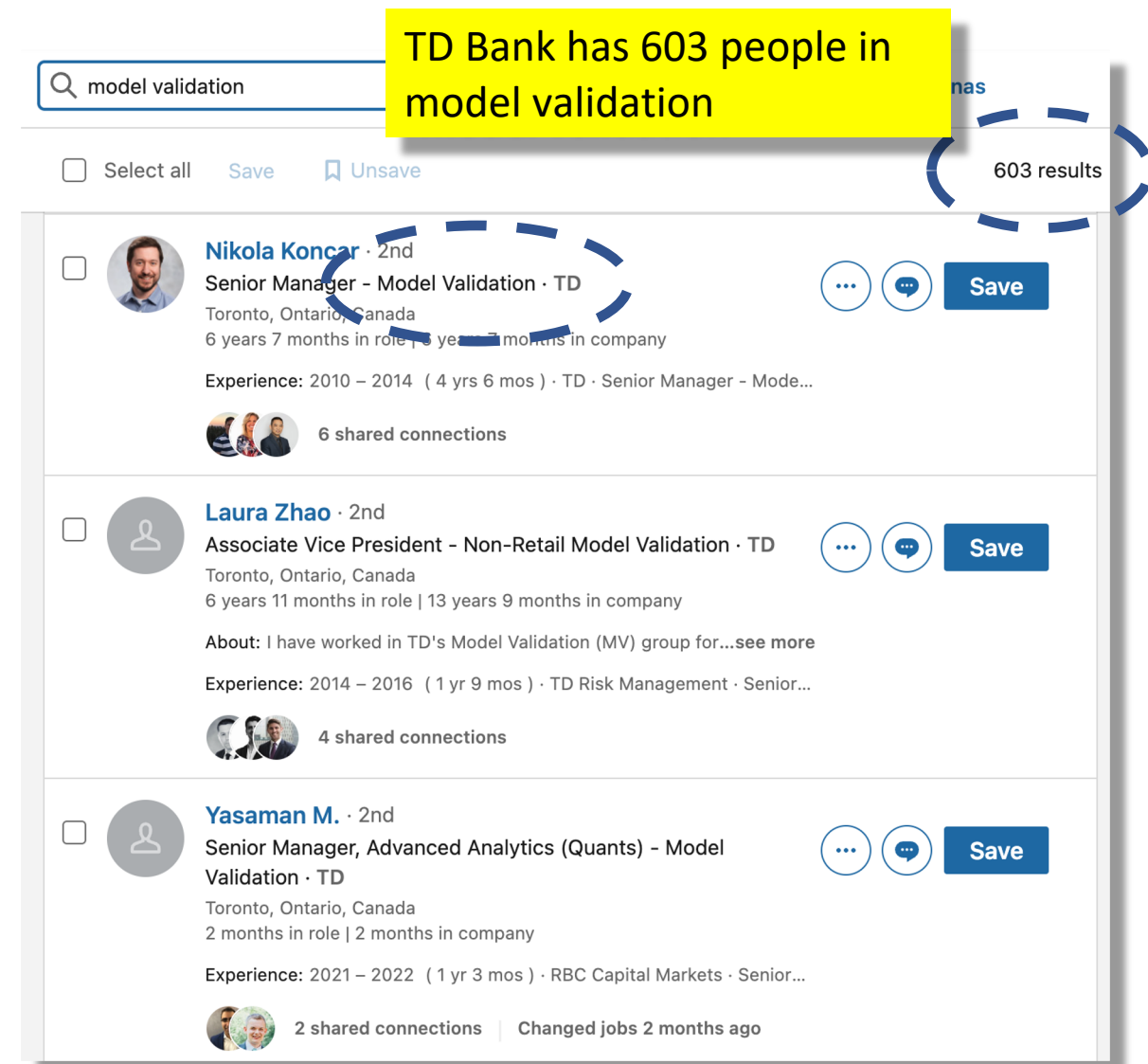
- Jackknife
- Benchmarking
- lift chart

A larger bank will have a model validation team to evaluate this before a model can be used.

After development

- Back testing: apply to vendors and compare forecast with data breach history

Note: that data breaches are rare events. A forecasting model cannot be test on a single company.



Back Testing

Cumulative forecast matches the past

About 20 data breaches in 10 years:
2 per year

~50 vendors

Company and Threats		Annual Probability by Breach Size									Breaches
Company	Sum	1	10	100	1,000	10,000	100,000	1,000,000	10,000,000	100,000,000	Year(breaches)
Anthem Blue Cross and Blue Shield	14.032%	13.937%	12.955%	9.633%	5.205%	2.014%	0.52341%	0.09409%	0.01471%	2017(2) 2015(2) 2014(1) 2013(4)	
Amazon Web Services	21.997%	21.563%	18.326%	10.872%	4.306%	1.178%	0.20601%	0.02392%	0.00257%	2019(1)	
Microsoft	32.687%	32.364%	29.356%	20.404%	9.997%	3.457%	0.78969%	0.12313%	0.01702%	2015(1) 2014(1) 2013(1)	
Linkedin Corporation	12.569%	12.481%	11.584%	8.575%	4.602%	1.767%	0.45509%	0.08100%	0.01255%	2016(1) 2013(1) 2012(1)	
Best Buy	28.057%	27.911%	26.294%	20.349%	11.674%	4.856%	1.37165%	0.27023%	0.04573%	2018(1) 2011(1)	
Comcast	42.554%	42.303%	39.607%	30.087%	16.788%	6.744%	1.83129%	0.34543%	0.05629%	2015(1) 2013(1) 2012(1) 2009(1)	
	2	2	2	1.5	0.79	0.30	0.079	0.014	0.0023		

Forecasts 2 per year

CISA is the low hanging fruit

Hiring more CISA is the most effective way to reduce data breaches

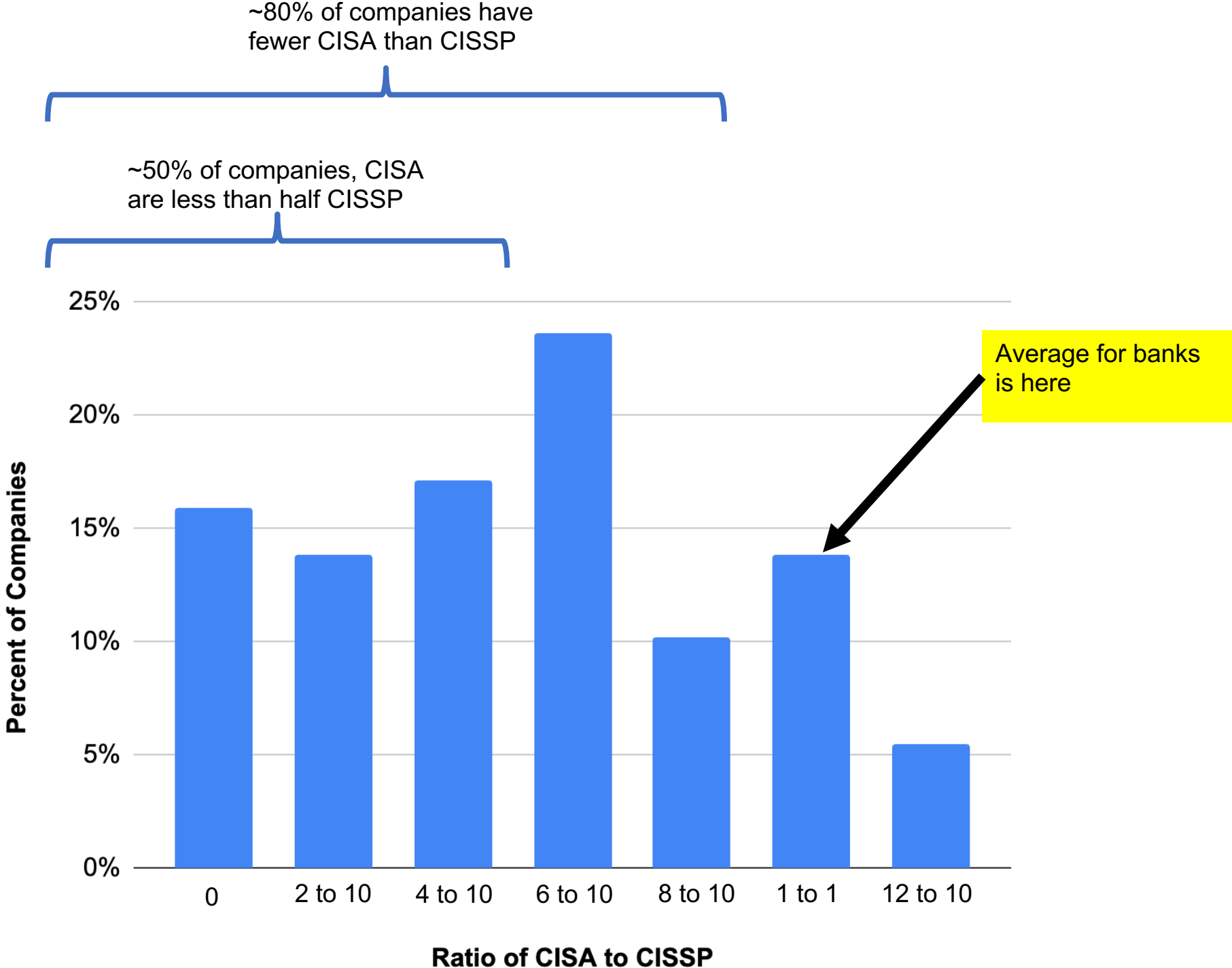
Predictors

Headcount	Description	Effect	Effect	
			Small Breach	Large Breach
		2 Diminishing return		
CISA/IT	Audit (3rd-line) Certified Information Systems Auditor	Decrease Probability (decreased risk)	Strong but, diminishing return	Very strong, no diminishing return
CISSP/IT	Technical (2nd-line) Certified Information Systems Security Professional		Strong but, diminishing return	Very strong, no diminishing return
MCSA/IT	Vendor (1st-line) Microsoft Certified Solutions Associate		Modest and diminishing return	Weak
Employees	Total employees	Increase Probability (increased risk)	Moderate	Moderate

1
Equal Effectiveness

Ratio of 3rd-line of defense (CISA) to 2nd-line of defense (CISSP)

Analysis of 1,500 companies



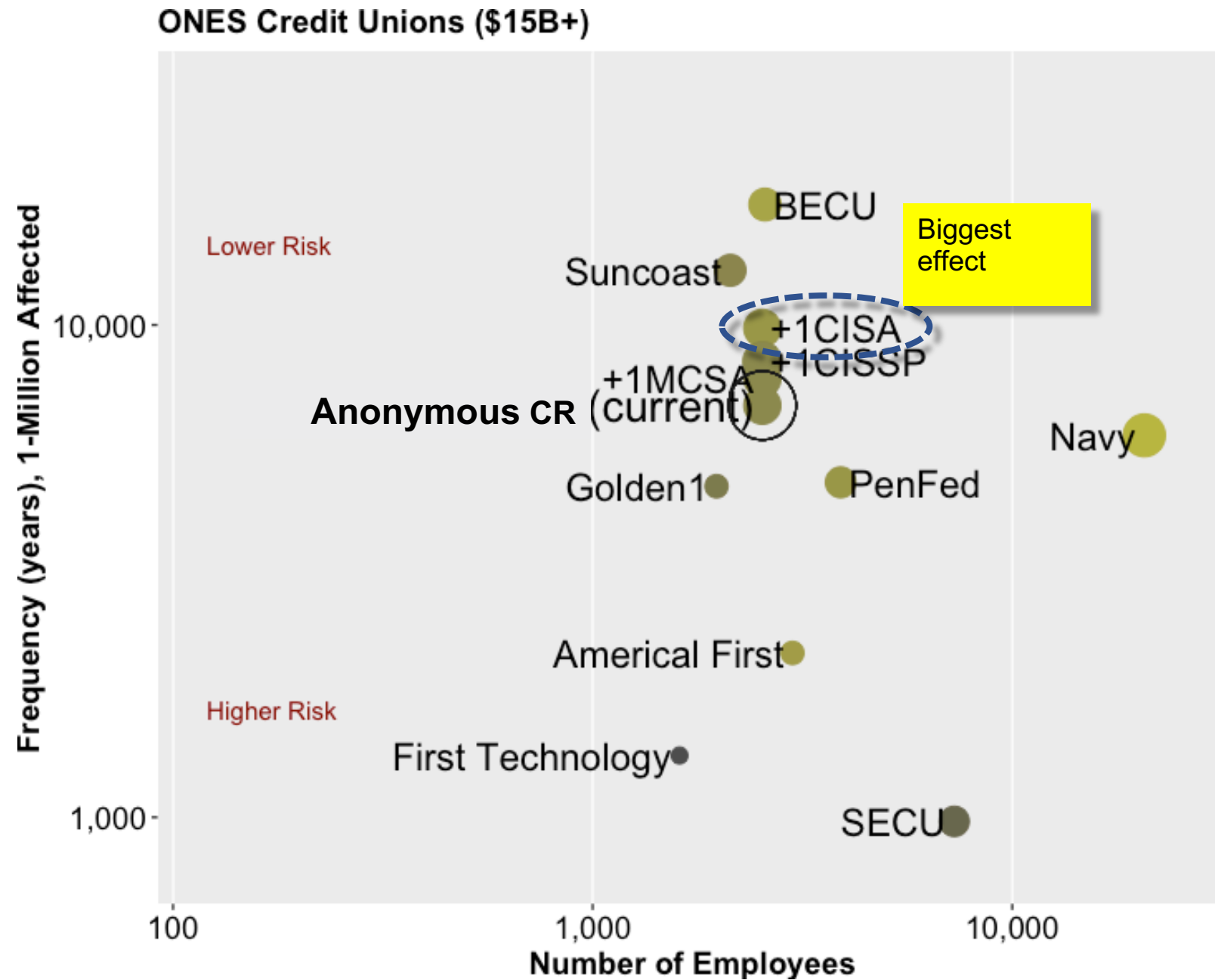
Sensitivity Analysis

Increasing CISA-headcount has the biggest effect

Anonymous CR Current headcount

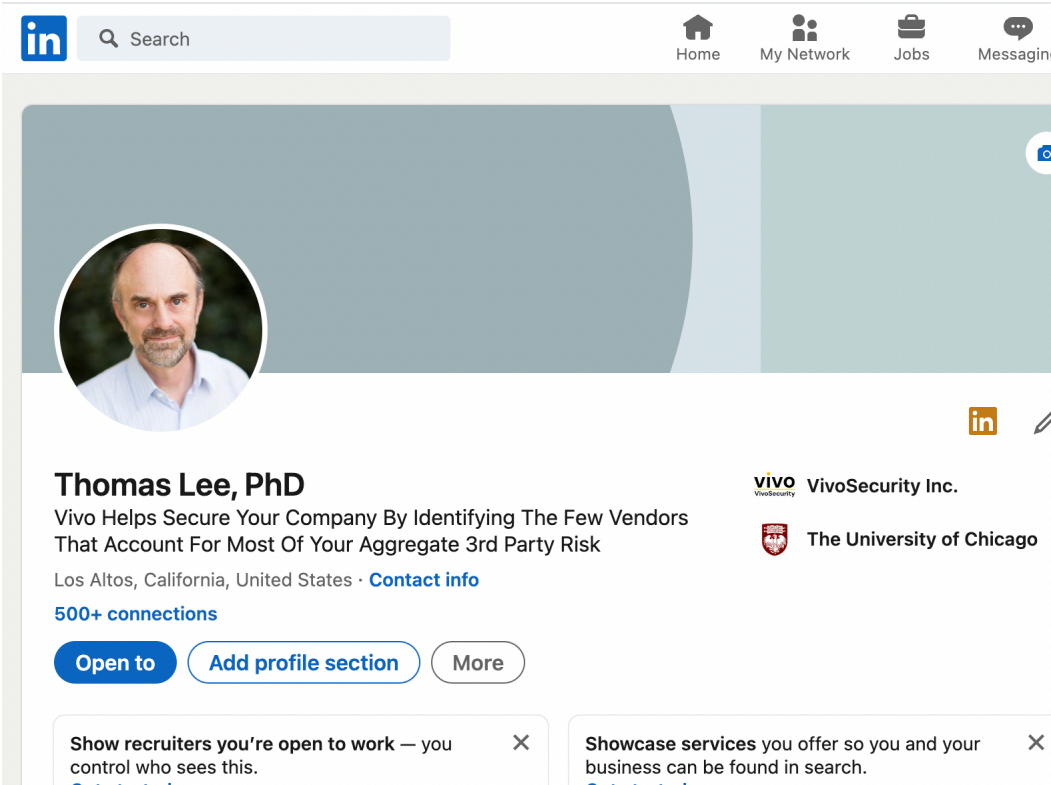
Cert	Count
CISSP	8
CISA	2
MCSA	2

Diminishing returns reduces value of adding more CISSP compared with CISA



Linkup with me

Tell me how you interpret results from modeling



Send me an email

Tell me how you interpret results from modeling

ThomasL@VivoSecurity.com

