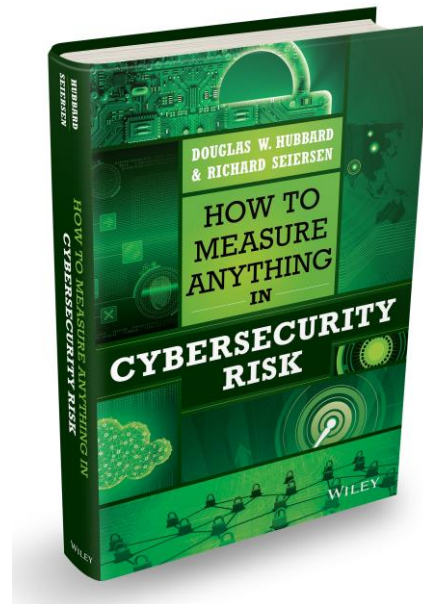




# How to Measure Anything in Cybersecurity Risk

## *An Introduction*



Hubbard Decision Research  
2 South 410 Canterbury Ct  
Glen Ellyn, Illinois 60137  
[www.hubbardresearch.com](http://www.hubbardresearch.com)

# My Co-Author and I



## Richard Seiersen

Currently the General Manager of Cybersecurity and Privacy at GE Health Care. Data driven executive with ~20 years experience spanning subject matters in Cyber Security, Quantitative Risk Management, Predictive Analytics, Big Data and Data Science, Enterprise Integrations and Governance Risk and Compliance (GRC). Led large enterprise teams, provided leadership in multinational organizations and tier one venture capital backed start-ups.



## Douglas Hubbard

Mr. Hubbard is the inventor of the powerful Applied Information Economics (AIE) method. He is the author of the #1 bestseller in Amazon's math for business category for his book titled ***How to Measure Anything: Finding the Value of Intangibles in Business*** (Wiley, 2007; 3<sup>rd</sup> edition 2014). His other two books are titled ***The Failure of Risk Management: Why It's Broken and How to Fix It*** (Wiley, 2009) and ***Pulse: The New Science of Harnessing Internet Buzz to Track Threats and Opportunities*** (Wiley, 2011).

# Uses of Applied Information Economics

AIE was applied initially to IT business cases. But over the last 20 years it has also been applied to other decision analysis problems in all areas of Business Cases, Performance Metrics, Risk Analysis, and Portfolio Prioritization.

## IT

- Prioritizing IT portfolios
- Risk of software development
- Value of better information
- Value of better security
- Risk of obsolescence and optimal technology upgrades
- Value of infrastructure
- Performance metrics for the business value of applications

## Business

- Movie / film project selection
- New product development
- Pharmaceuticals
- Medical devices
- Publishing
- Real estate

## Engineering

- Risks of major engineering projects
- Risk of mine flooding

## Government & Non Profit

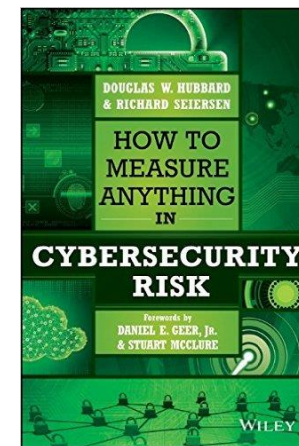
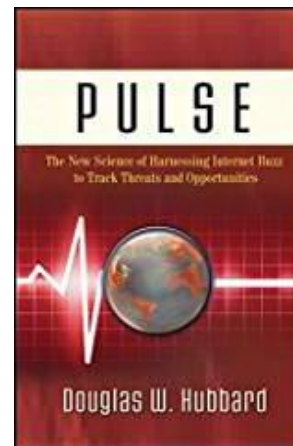
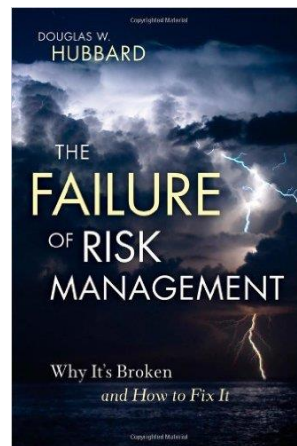
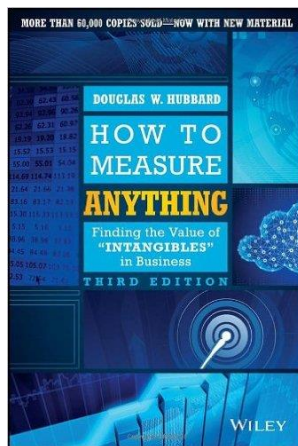
- Environmental policy
- Sustainable agriculture
- Procurement methods
- Grants management

## Military

- Forecasting battlefield fuel consumption
- Effectiveness of combat training to reduce roadside bomb / IED casualties
- R&D portfolios

# What I Write

- Four books, over 100,000 copies sold in 8 languages.
  - How to Measure Anything: Finding the Value of Intangibles in Business
  - The Failure of Risk Management: Why It's Broken and How to Fix It
  - Pulse: The New Science of Harnessing Internet Buzz to Track Threats and Opportunities
  - How to Measure Anything in Cybersecurity Risk



- First two are required reading in Society of Actuaries Exam Prep and used in courses at 20+ universities.

# The Biggest Cybersecurity Risk

**Question: What is your single biggest risk in cybersecurity?**

**Answer: How you measure cybersecurity risk.**

**(This also applies to risk in general.)**

# Current Solution

Here are some risks plotted on a “typical heat map”.

Suppose mitigation costs were:

- Risk 1: \$725K - **High**
- Risk 2: \$95K - **Low**
- Risk 3: \$2.5M - **Critical**
- Risk 4: \$375K - **Moderate**

Impact					Likelihood
Low	Medium	High	Extreme		
Moderate	High	Critical	Critical	Extreme	
Low	Moderate	High	Critical	High	
Low	Moderate	High	High	Medium	
Low	Low	Moderate	Moderate	Low	
Low	Low	Low	Moderate	Negligible	
4	2	1	3		

What mitigations should be funded and what is the priority among those?

# Summarizing Research on Risk Matrices

- Tony Cox “What’s wrong with Risk Matrices” investigates various mathematical consequences of ordinal scales on a matrix.
  - “...they can be “worse than useless,” leading to worse-than-random decisions.”
- Bickel et al. “The Risk of Using Risk Matrices”, *Society of Petroleum Engineers*, 2014
  - “The burden of proof is squarely on the shoulders of those who would recommend the use of such methods to prove that these obvious inconsistencies do not impair decision making, much less improve it, as is often claimed.’

# Is Risk Analysis Actually Supporting Decisions?

- If risks and mitigation strategies were quantified in a meaningful way, decisions could be supported.
- In order to compute an ROI on mitigation decisions, we need to quantify likelihood, monetary impact, cost, and effectiveness.

	Expected Loss/Yr	Cost of Control	Control Effectiveness	Return on Control	Action
DB Access	\$24.7M	\$800K	95%	2,832%	Mitigate
Physical Access	\$2.5M	\$300K	99%	727%	Mitigate
Data in Transit	\$2.3M	\$600K	95%	267%	Mitigate
Network Access Control	\$2.3M	\$400K	30%	74%	Mitigate
File Access	\$969K	\$600K	90%	45%	Monitor
Web Vulnerabilities	\$409K	\$800K	95%	-51%	Track
System Configuration	\$113K	\$500K	100%	-77%	Track



# How to Build a Method That Works

Start with components that work – that is, the improvement has been measured in controlled tests.

- We can still rely on expert judgement, but it must be calibrated to account for various known errors in expert judgement.
- We have to do the math right and do it with real probabilities including how to update probabilities with empirical data.

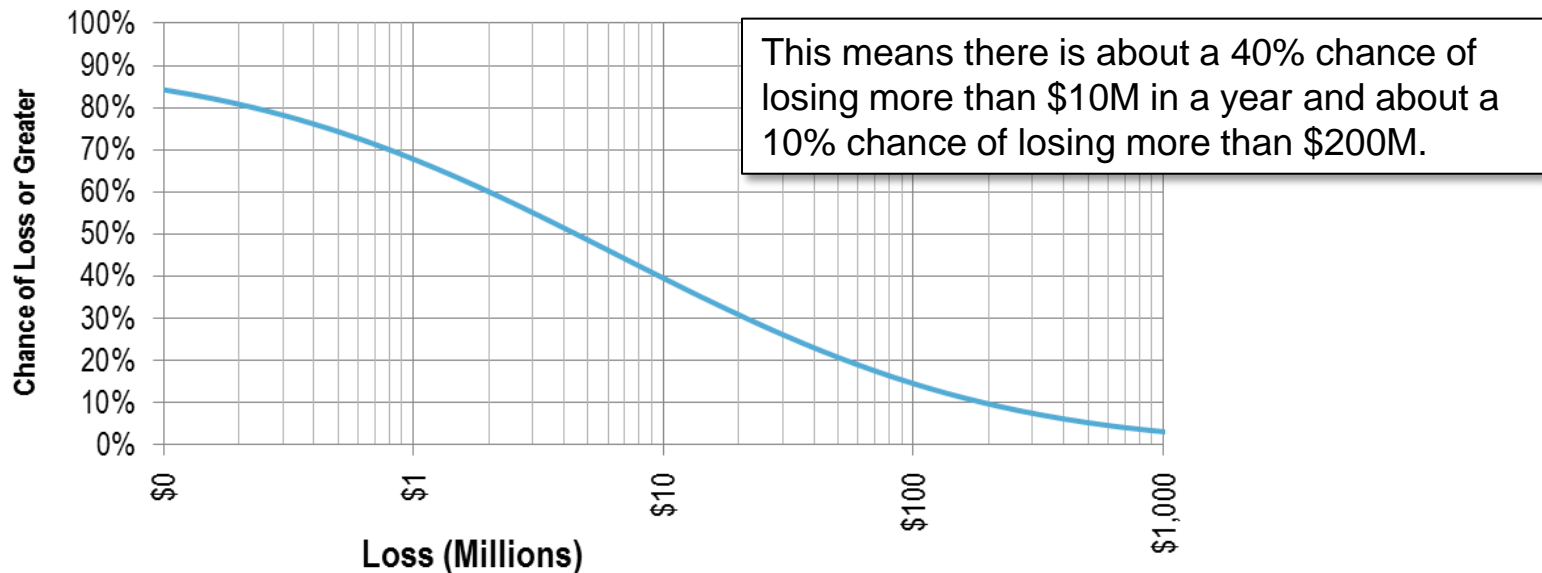
If you can't answer "What is the probability of losing more than X in the next 12 months due to event Y?" then you aren't doing risk analysis.

# What Measuring Risk Looks Like

What if we could measure risk more like an actuary – “The probability of losing more than \$10 million due to security incidents in 2016 is 16%”

What if we could prioritize security investments based on a “Return on Mitigation”?

	Expected Loss/Yr	Cost of Control	Control Effectiveness	Return on Control	Action
DB Access	\$24.7M	\$800K	95%	2,832%	Mitigate
Physical Access	\$2.5M	\$300K	99%	727%	Mitigate
Data in Transit	\$2.3M	\$600K	95%	267%	Mitigate
Network Access Control	\$2.3M	\$400K	30%	74%	Mitigate
File Access	\$969K	\$600K	90%	45%	Monitor
Web Vulnerabilities	\$409K	\$800K	95%	-51%	Track
System Configuration	\$113K	\$500K	100%	-77%	Track



# A Simple “One-For-One Substitution”

Each of these examples can be found on

**[www.howtomeasureanything.com/cybersecurity](http://www.howtomeasureanything.com/cybersecurity)**

Event	Event Probability (per Year)	Impact (90% Confidence Interval)		Random Result (zero when the event did not occur)
		Lower Bound	Upper Bound	
AA	.1	\$50,000	\$500,000	0
AB	.05	\$100,000	\$10,000,000	\$8,456,193
AC	.01	\$200,000	\$25,000,000	0
AD	.03	\$100,000	\$15,000,000	0
AE	.05	\$250,000	\$30,000,000	0
AF	.1	\$200,000	\$2,000,000	0
AG	.07	\$1,000,000	\$10,000,000	\$2,110,284
AH	.02	\$100,000	\$15,000,000	0
↓ ↓ ↓ ↓ ↓				
ZM	.05	\$250,000	\$30,000,000	0
ZN	.01	\$1,500,000	\$40,000,000	0
Total:				\$23,345,193

Each “Dot” on a risk matrix can be better represented as a row on a table like this

The output can then be

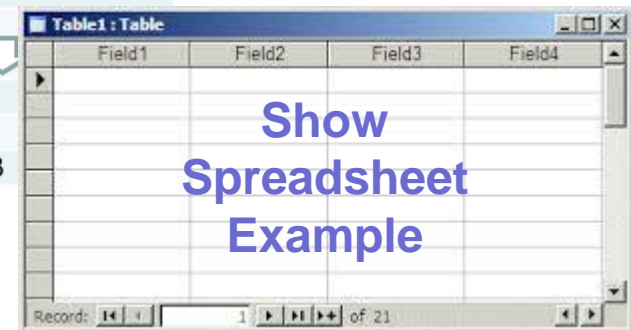


Table1 : Table

Field1	Field2	Field3	Field4

Record: 1 of 21

Show Spreadsheet Example

# Quantifying Your Current Uncertainty

- Decades of studies show that most managers are statistically “overconfident” when assessing their own uncertainty.
- Studies also show that measuring *your own* uncertainty about a quantity is a general skill that can be taught with a **measurable** improvement.
- Training can “calibrate” people so that of all the times they say they are 90% confident, they will be right 90% of the time.
- HDR has calibrated over 1,000 people in the last 20 years – 85% of participants reach calibration within a half-day of training

“Overconfident professionals sincerely believe they have expertise, act as experts and look like experts. You will have to struggle to remind yourself that they may be in the grip of an illusion.”

Daniel Kahneman, Psychologist, Economics Nobel



# Calibration Exercise: Ranges

- For the following questions, provide a range (an upper and lower bound) that you are 90% certain contains the correct answer:

	Lower Bound	Upper Bound
Napoleon Bonaparte was born what year?		
What is the average weight of an adult male African elephant (tons)?		
The Coliseum in Rome held how many spectators?		
How many countries are in NATO?		
In what year did Newton publish the Laws of Gravitation?		

# Calibration Exercise: True/False

- For each statement below, answer whether you believe it is true or false and provide a percentage confidence that your answer is correct. Confidence is any value between 50% (“no idea”) to 100% (certainty).

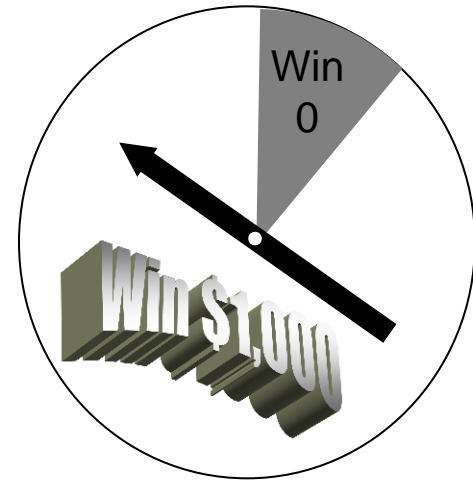
	True or False?	% Confidence
Brazil has a larger population than Spain.		
A hockey puck will fit in a golf hole.		
The Yangtze River is the longest river in Asia.		
Mars is always further away from Earth than Venus is from Earth.		
The movie <i>Titanic</i> still holds the record for box office receipts in the first six weeks.		

# Calibration Aid: “The Equivalent Bet”

One method used in calibrating subject matter experts is the “equivalent bet.” Research shows that merely pretending to bet money on an estimate can improve estimates.

- For 90% Confidence Interval questions, which would you rather have?
  - **A:** Win \$1,000 if your interval contains the correct answer
  - **B:** Spin a dial with a 90% chance to win \$1,000
- For the Binary Confidence questions, which would you rather have?
  - **A:** Win \$1,000 if your answer is correct
  - **B:** Spin a dial with a chance to win \$1,000 equal to your stated confidence

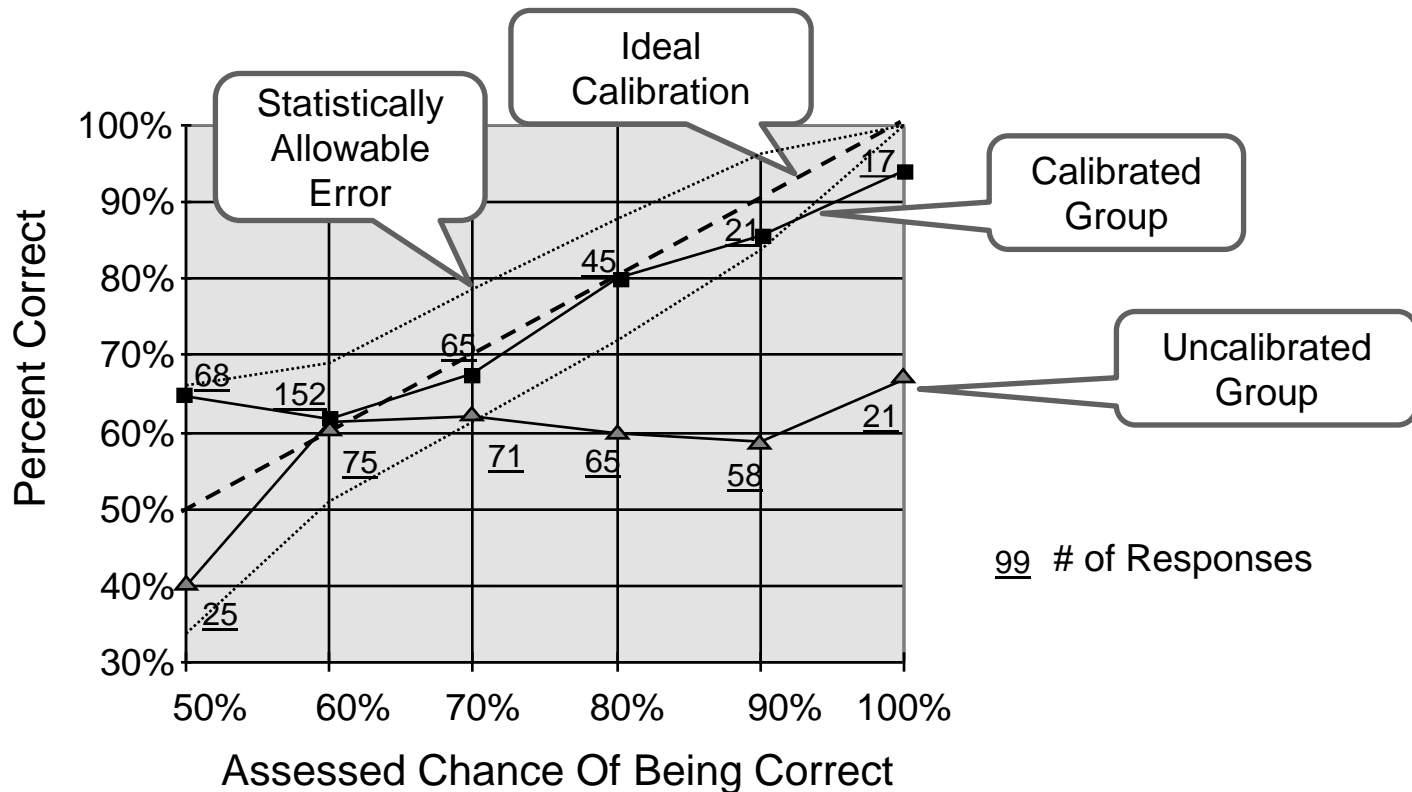
## Option B:



## Spin the Dial!

# Training Experts to Give Calibrated Probabilities

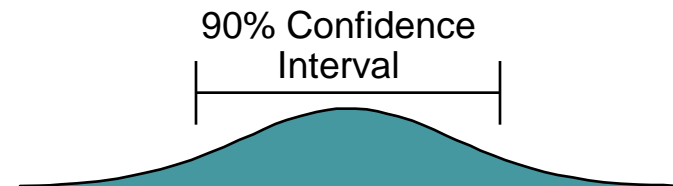
- Training can “calibrate” people so that of all the times they say they are 90% confident, they will be right 90% of the time.





# Overconfidence in Ranges

- The same training methods apply to the assessment of uncertain ranges for quantities like the duration of a future outage, the records compromised in a future breach, etc.



Group	Subject	% Correct (target 90%)
Harvard MBAs	General Trivia	40%
Chemical Co. Employees	General Industry	50%
Chemical Co. Employees	Company-Specific	48%
Computer Co. Managers	General Business	17%
Computer Co. Managers	Company-Specific	36%
AIE Seminar (before training)	General Trivia & IT	35%-50%
AIE Seminar (after training)	General Trivia & IT	~90%

# Calibration Answers

	Lower Bound
Napoleon Bonaparte was born what year?	1769
What is the average weight of an adult male African elephant (tons)?	3.5 tons
The Coliseum in Rome held how many spectators?	50,000
How many countries are in NATO?	28
In what year did Newton publish the Laws of Gravitation?	1687

	True or False?
Brazil has a larger population than Spain.	True
A hockey puck will fit in a golf hole.	True
The Yangtze River is the longest river in Asia.	True
Mars is always further away from Earth than Venus is from Earth.	False
The movie Titanic still holds the record for box office receipts in the first six weeks.	False

# Inconsistency vs. Discrimination

- Discrimination is how much your estimates vary when given different information.
- Inconsistency is the amount of your discrimination that is due to random differences in estimates - this may be in addition to differences in interpreting verbal scales, so let's assume we are using explicit probabilities.
- Experts are routinely influenced by irrelevant, external factors - *anchoring*, for example, is the tendency for an estimator to be influenced by recent exposure to an another unrelated number (Kahneman).



# More Aspirational: Lens Method

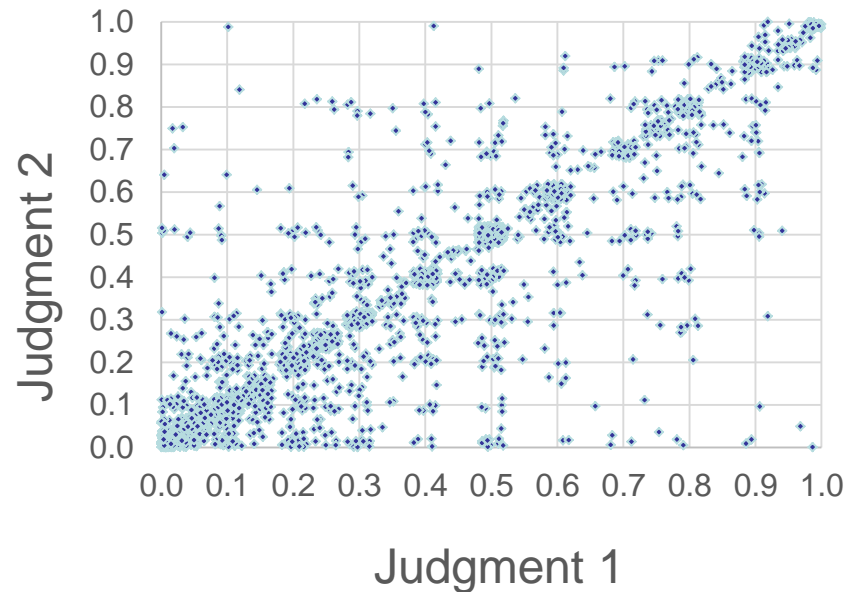
Example #	MS OS	Patch Current	Host/ Service Location	Users	Known Exploit	Chance of Breach (Team Avg.)
1	0	1	Data Center	<20	1	0.775
2	0	1	Vendor 1	>=20,<1K	0	0.675
3	1	0	Vendor 1	>10K	0	0.7425
4	1	1	Vendor 2	>1K,<10K	0	0.5875
5	0	1	Data Center	>=20,<1K	1	0.805
6	1	1	Vendor 1	>10K	01	0.8025
7	0	1	Data Center	>1K,<10K	0	0.7025
8	0	1	Vendor 2	>=20,<1K	0	0.6875
9	0	1	Vendor 2	>10K	1	0.6925
10	1	1	Data Center	>1K,<10K	1	0.7825
11	1	0	Vendor 1	>=20,<1K	0	0.8
12	1	1	Data Center	>1K,<10K	0	0.81
13	1	1	Vendor 2	>=20,<1K	0	0.765
14	0	0	Data Center	>=20,<1K	0	0.74
15	0	1	Data Center	>10K	0	0.6825
16	1	0	Vendor 3	>1K,<10K	1	0.65
17	1	1	Vendor 2	>=20,<1K	1	0.6975
18	0	1	Vendor 1	>1K,<10K	1	0.81
19	1	0	Vendor 1	>1K,<10K	0	0.78
20	0	1	Data Center	>=20,<1K	0	0.855
21	0	1	Vendor 2	>10K	1	0.8425
22	0	1	Data Center	>1K,<10K	1	0.92

- The Lens Method is a regression model of expert judgement.
- Experts are given a list of hypothetical scenarios to judge – effectively, this is sample of an NPT.
- Experts are averaged to create a group average for each scenario.
- A regression model (using standard Excel tools) is created to model their judgement.

# Calibrating Expert Consistency

- We have gathered estimates of probabilities of various security events from:
  - 48 experts from 4 different industries.
  - Each expert was given descriptive data for over 100 systems.
  - For each system each expert estimated probabilities of six or more different types of security events.
- Total: Over 30,000 individual estimates of probabilities
- These estimates included over 2,000 duplicate scenarios pairs.

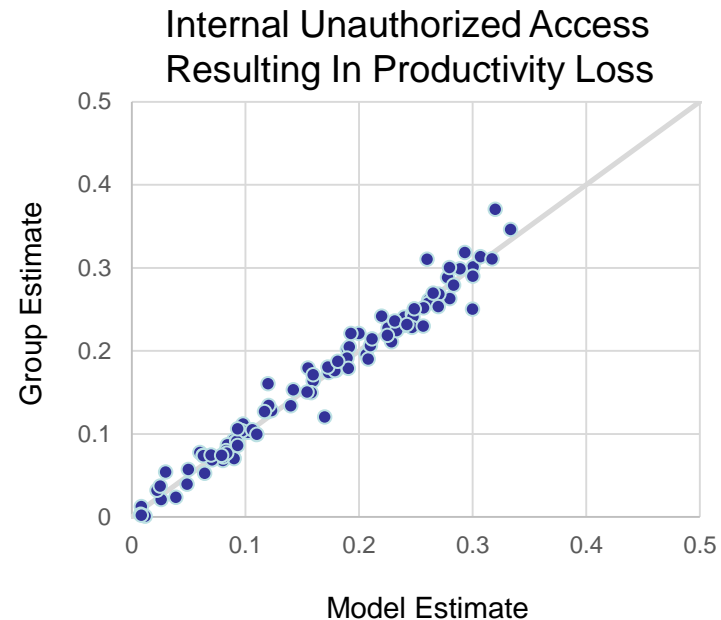
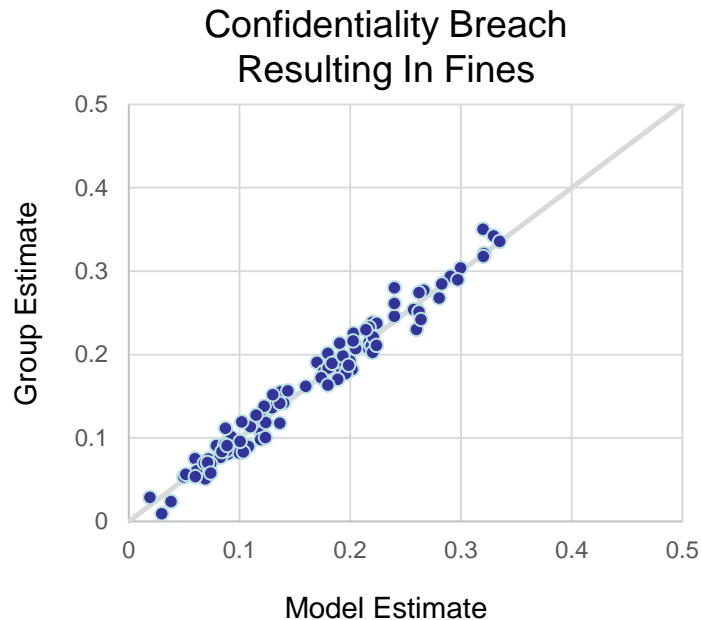
Comparison of 1<sup>st</sup> to 2<sup>nd</sup> Estimates of Cyber risk judgements by same SME



**21% of variation in expert responses are explained by *inconsistency*.**  
(79% are explained by the actual information they were given)

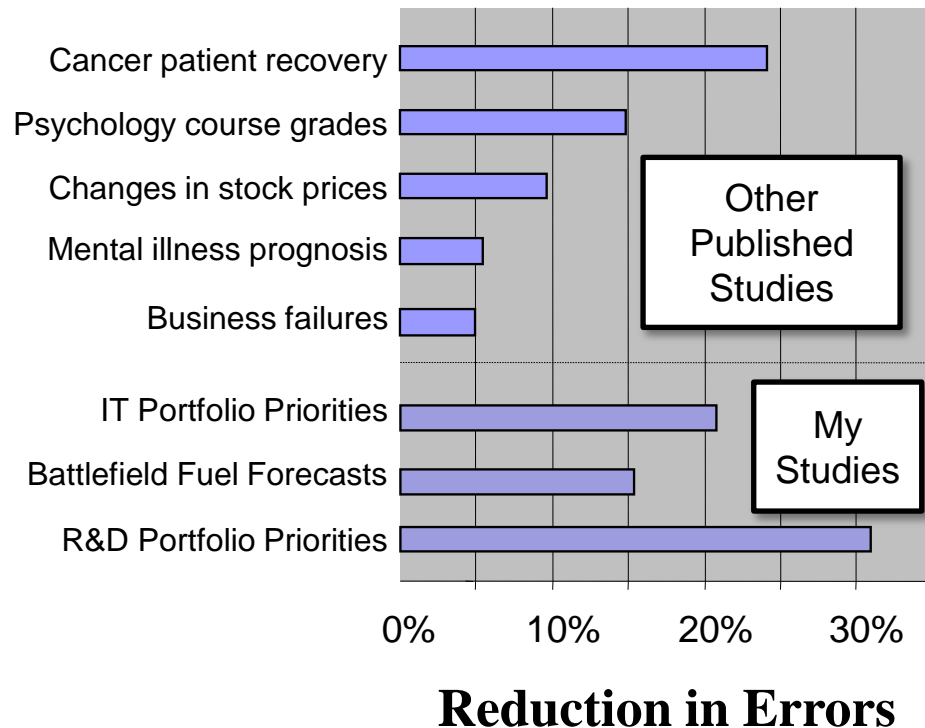
# Modeling Group Estimates of IT Security Event Likelihood

Examples of Models vs. Group Averages: Probabilities of different security events happening in the next 12 months for various systems prior to applying particular controls.



- The models created produce results which closely match the group's average.
- A large portion of the model error is due to judge inconsistency.
- This nearly eliminates the inconsistency error.

# Effects of Removing Inconsistency Alone

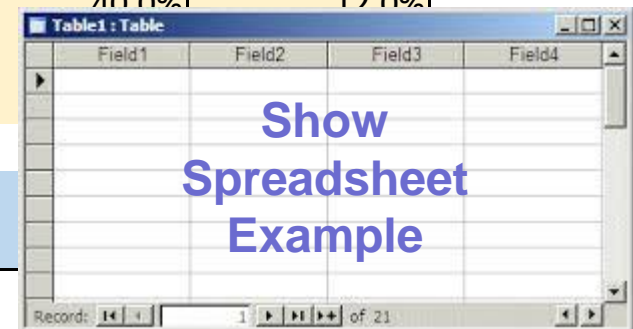


- A method of improving expert estimates of various quantities was developed in the 1950's by Egon Brunswik.
- He called it the "Lens Method"
- It has been applied to several types of problems, including expert systems, with consistently beneficial results.

# Logodds Model (Aspirational?)

- A Logodds Model is a relatively simple approximation to “add up” a number of parameters that modify a probability when NPTs would be large.
- Logodds of  $X = LO(X) = \ln(P(X)/(1-P(X)))$
- Adjustment due to condition  $Y = A(Y) = LO(P(X|Y)) - LO(P(X))$
- $P(X|A,B,...) = A(\text{Sum of } (LO(A), LO(B), ...) + LO(P(X)))$
- The more independent the parameter are, the better the Rasch approximation.

Initial Prob: P(E)	10%			
Baseline Logodds	-2.197			
	Conditions			
	A	B	C	D
P(E   X)	34.0%	15.0%	10.0%	12.0%
P(E   ~X)	5.5%	9.0%		
P(X)	16.0%	20.0%		
Test P( E )	10.1%	10.2%		
Logodds change   X	1.5339	0.4626		
Logodds change   ~X	-0.6466	-2.3136		





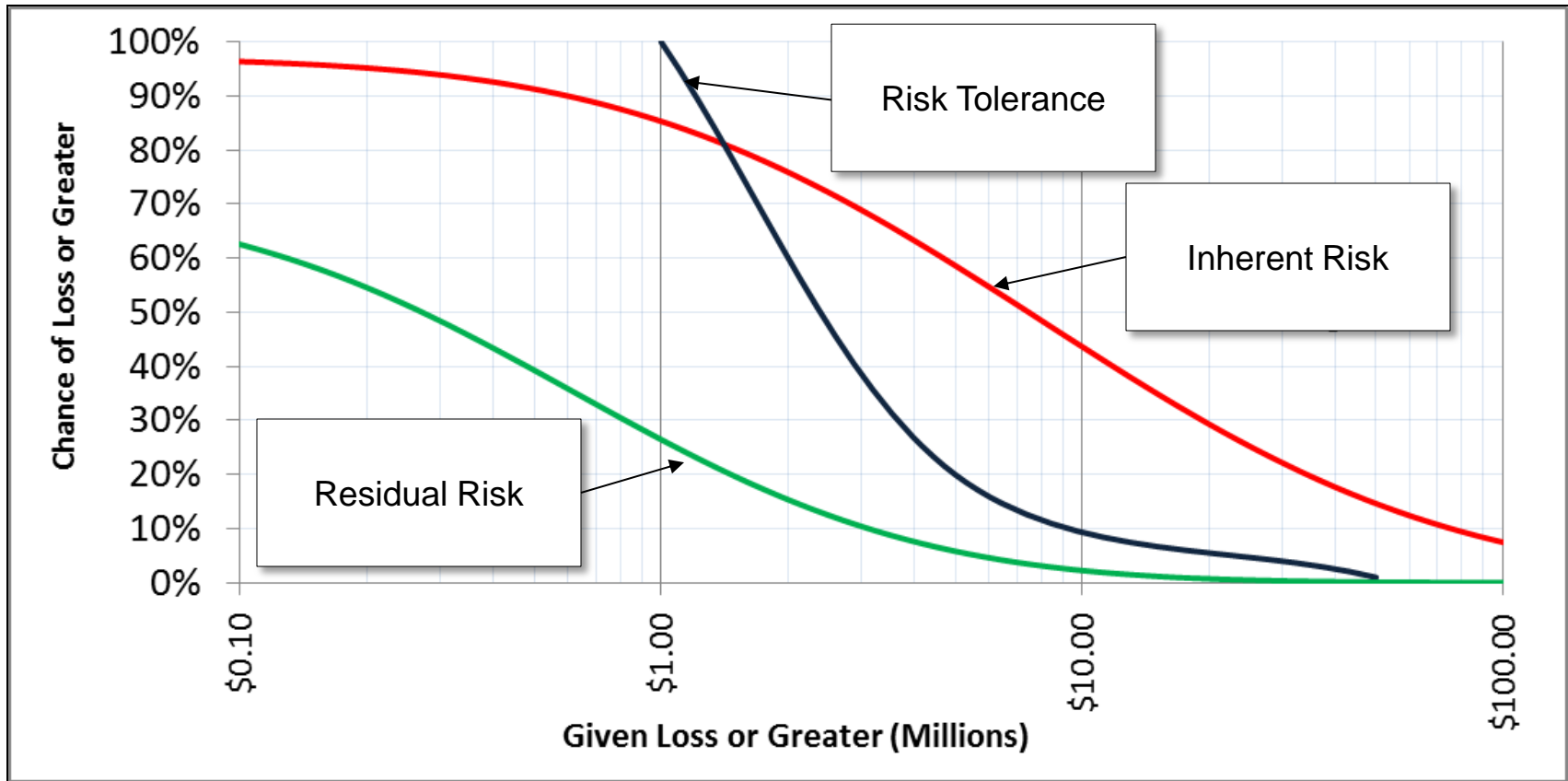
# Calibrating Risk Tolerance

- Studies have shown risk aversion changes due to what should be irrelevant external factors including:

Factor	Risk Aversion
Being around smiling people	↓
Recalling an event causing fear	↑
Recalling an event causing anger	↓
A recent win in an unrelated decision	↓
A recent loss in an unrelated decision	↑

# Loss Exceedance Curves: Before and After

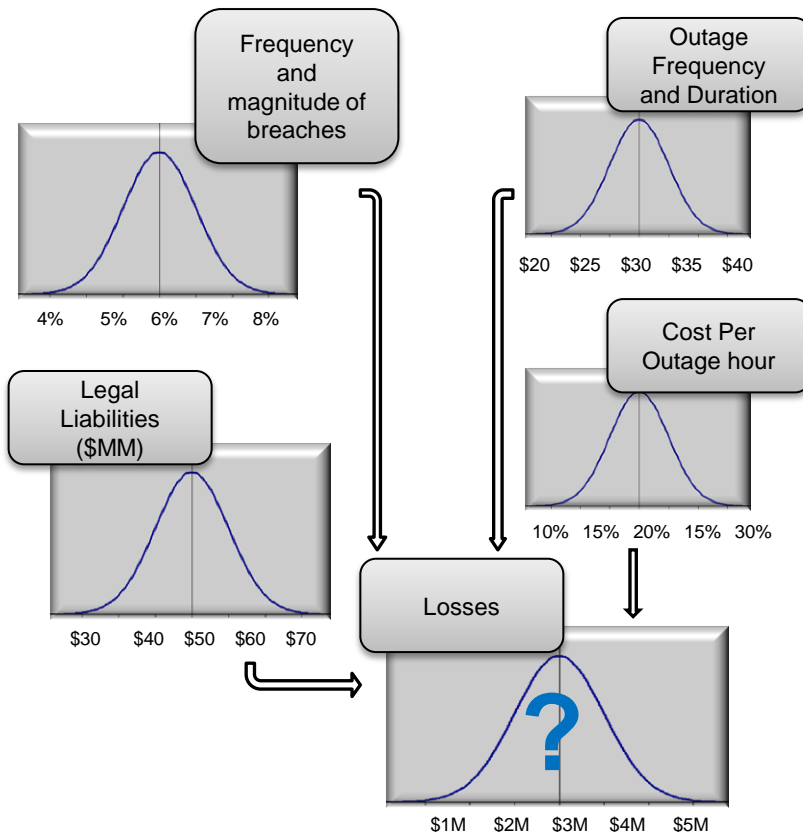
How do we show the risk exposure after applying available mitigations?



# Doing the Math with Probabilities

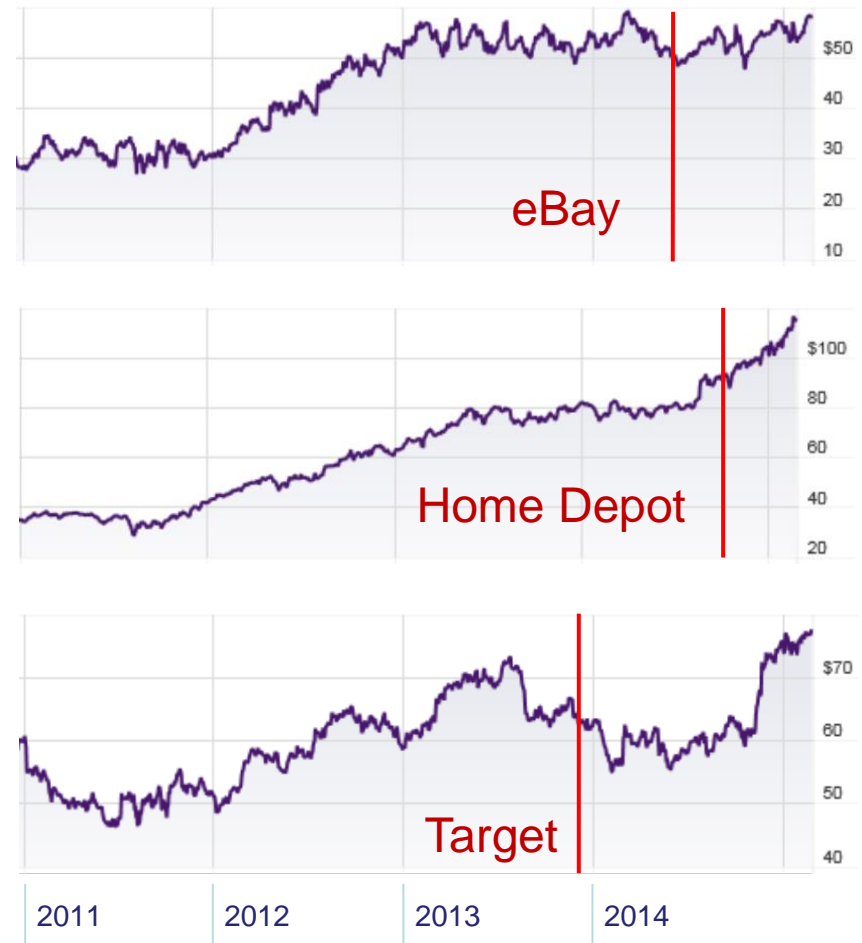
## What Published Research Says (See sources slide for details)

- Psychologists showed that simple decomposition greatly reduces estimation error for estimating the most uncertain variables.
- In the oil industry there is a correlation between the use of quantitative risk analysis methods and financial performance
- Data at NASA from over 100 space missions showed that Monte Carlo simulations and historical data beat softer methods for estimating cost and schedule risks.



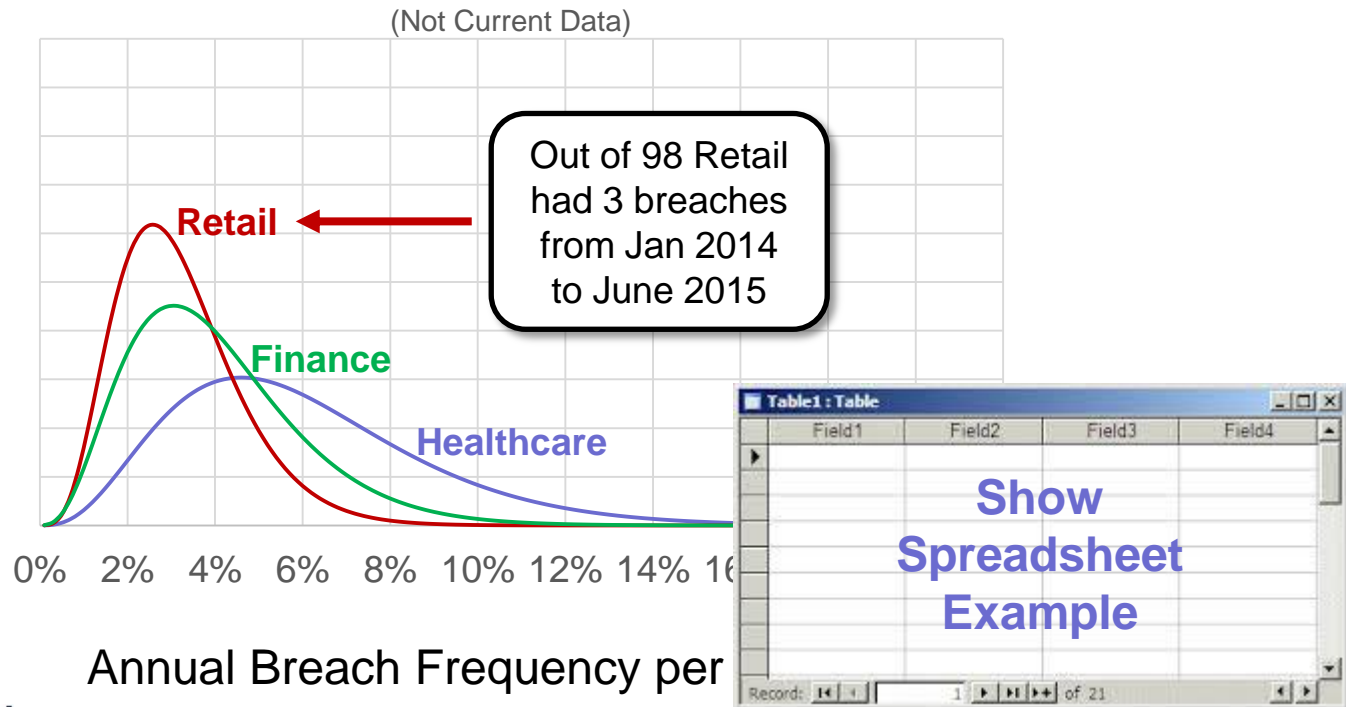
# Measurement Challenge: Reputation Damage

- One of the perceived most difficult measurements in cybersecurity is damage to reputation.
- Trick: *There is no such thing as a “secret” damage to reputation!*
- How about comparing stock prices after incidents? (That’s all public!)
- So what is the *REAL* damage?
  - Legal liabilities,
  - Customer outreach
  - “Penance” projects (security overkill)
- The upshot, damage to reputation actually has available information and easily observable measured costs incurred to *avoid* the bigger damages!



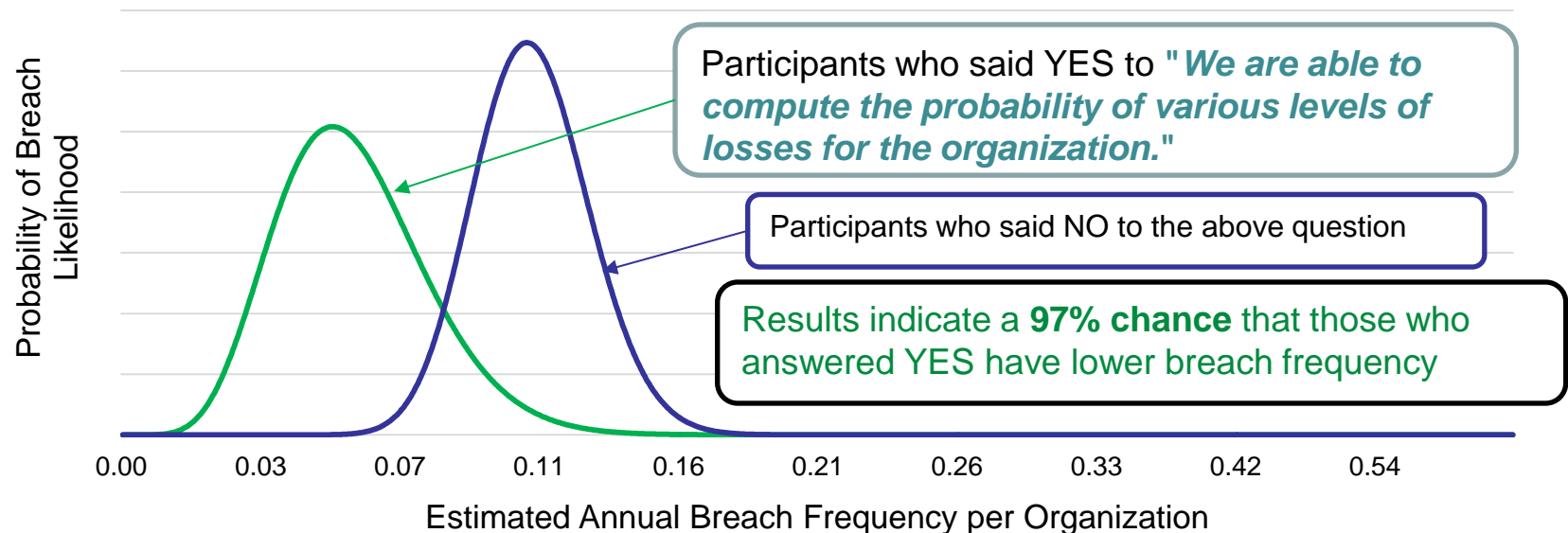
# Statistics Needs Less Data Than You Think

- You have relatively few examples of major, reported breaches in each industry.
- There is a statistical method for estimating the frequency of breaches based on small samples. This is the “beta” distribution and it is provided in Excel as “=betadist(proportion, hits, misses)”
- Spreadsheet for this at [www.howtomeasureanything.com/cybersecurity](http://www.howtomeasureanything.com/cybersecurity)



# What Reduces Data Breach Risk?

- The survey reveals another interesting result.
- Those who said they computed the probability of losses reported fewer breaches than those who did not.
- I would not treat this observation alone as sufficient – but it agrees with other independent evidence.



# What to Do Next

## Things you can do now:

- Stop using ordinal scales and risk matrices for evaluating risk – they only create the illusion of analysis
- Replace risk matrix activities with the One-for-One Substitution model
- Experiment with simple additional decompositions as shown in the Chapter 6 download (both spreadsheets available at [www.hubbardresearch.com/cybersecurity](http://www.hubbardresearch.com/cybersecurity))

## Things to strive toward (the effort is easily justified for Cybersecurity):

- Get calibrated so you can quantify your uncertainty
- Learn more advanced decompositions including Log Odds and the Lens Method
- Update the initial model with empirical data using slightly more advanced statistical methods

Measure what matters. Make better decisions.

# Questions?

---

## Contact:

Doug Hubbard

Hubbard Decision Research

[dwhubbard@hubbardresearch.com](mailto:dwhubbard@hubbardresearch.com)

[www.hubbardresearch.com](http://www.hubbardresearch.com)

630 858 2788



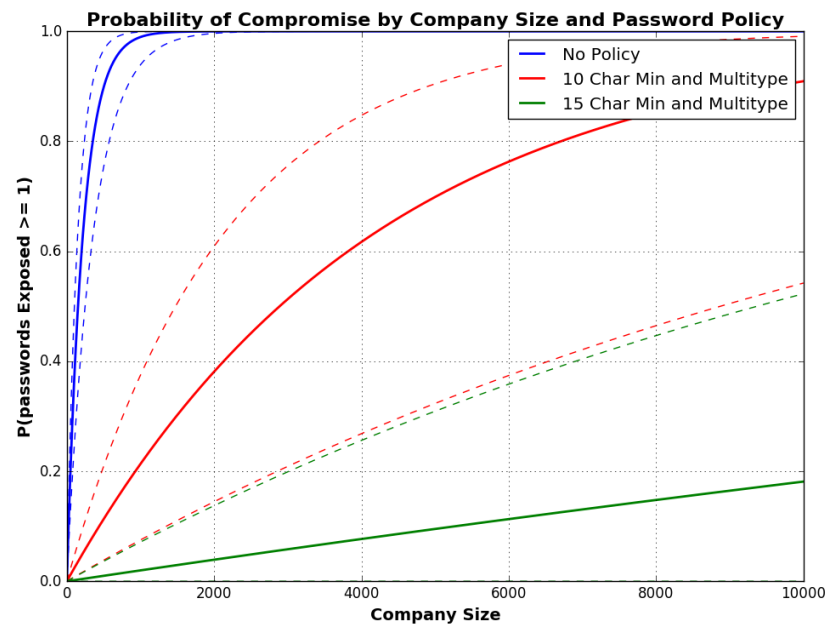


# Supplementary Material

Hubbard Decision Research  
2 South 410 Canterbury Ct  
Glen Ellyn, Illinois 60137  
[www.hubbardresearch.com](http://www.hubbardresearch.com)

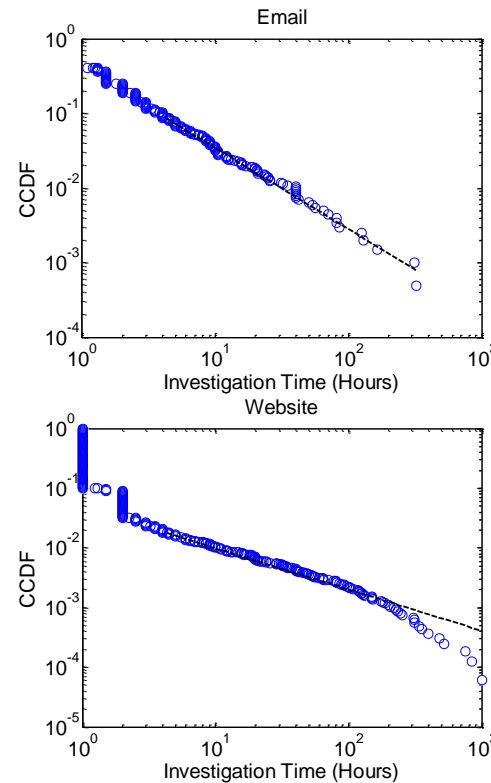
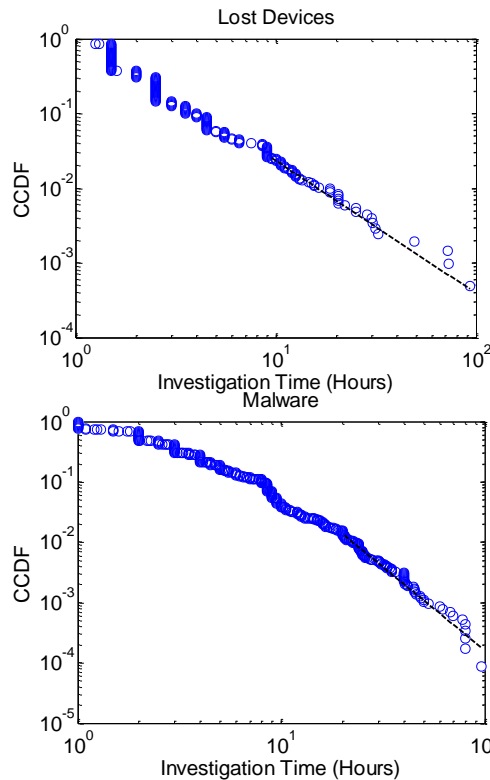
# Password Compromise Statistics

Source: Anton Mobley, GE Healthcare



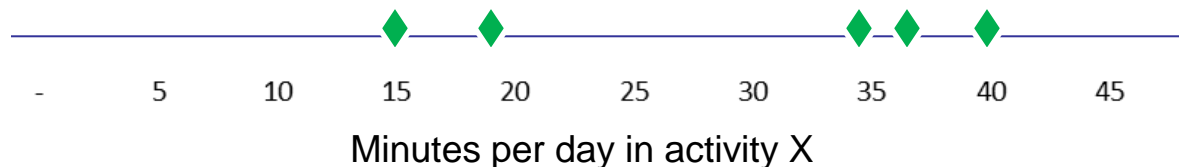
# Power Laws in Investigation Times

The investigation times of several types of events are shown to have “Power Law” distributions. (Source: Marshal Kuypers)



# Testing Our Measurement Intuition

- Suppose, as part of a cybersecurity measurement problem, you are interested in determining how much time is in some process.
- You are extremely uncertain about how much time per day is spent in this activity in a company of 10,000 people
- Imagine you randomly sample 5 people out of a company and they spend an amount of time in this activity as shown by the data points below
- Is this statistically significant?
- Is it possible to estimate the chance the median time spent per person per day is between 15 and 40 minutes?

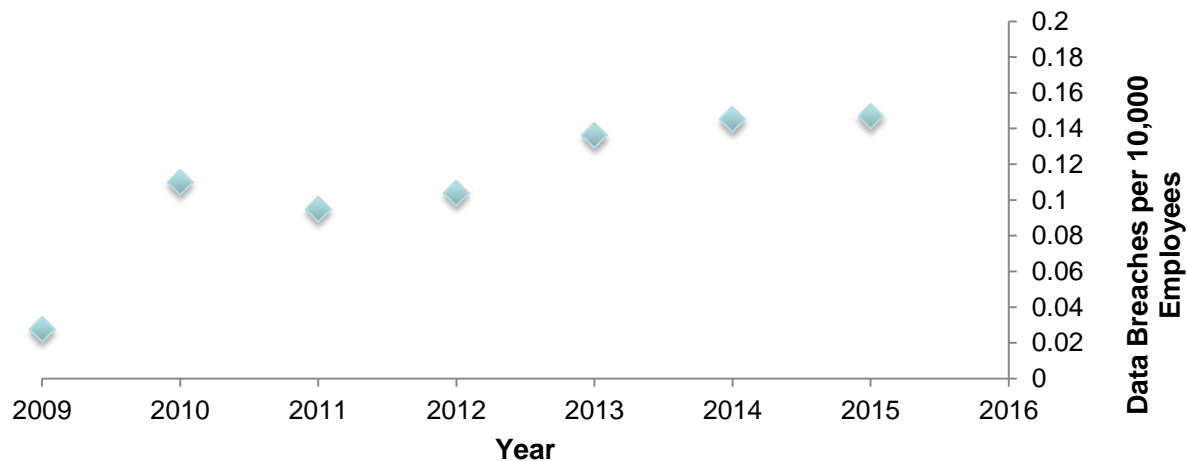


# Selected Sources

- Tsai C., Klayman J., Hastie R. “Effects of amount of information on judgment accuracy and confidence” *Org. Behavior and Human Decision Processes*, Vol. 107, No. 2, 2008, pp 97-105
- Heath C., Gonzalez R. “Interaction with Others Increases Decision Confidence but Not Decision Quality: Evidence against Information Collection Views of Interactive Decision Making” *Organizational Behavior and Human Decision Processes*, Vol. 61, No. 3, 1995, pp 305-326
- Andreassen, P.” Judgmental extrapolation and market overreaction: On the use and disuse of news” *Journal of Behavioral Decision Making*, vol. 3 iss. 3, pp 153-174, Jul/Sep 1990
- Williams M. Dennis A., Stam A., Aronson J. “The impact of DSS use and information load on errors and decision quality” *European Journal of Operational Research*, Vol. 176, No. 1, 2007, pp 468-81
- Knutson et. al. “Nucleus accumbens activation mediates the influence of reward cues on financial risk taking” *NeuroReport*, 26 March 2008 - Volume 19 - Issue 5 - pp 509-513
- A small study presented at Cognitive Neuroscience Society meeting in 2009 by a grad student at U. of Michigan showed that simply being briefly exposed to smiling faces makes people more risk tolerant in betting games.
- Risk preferences show a strong correlation to testosterone levels – which change daily (Sapienza, Zingales, Maestripieri, 2009).
- Recalling past events that involved fear and anger change the perception of risk (Lerner, Keltner, 2001).

# Data Breaches/Yr vs. Number of Employees

Data from the HHS “Wall of Shame” indicates that the rate of data breaches (more than 500 confidential records) is now consistently 14% per year per 1,000 employees.



Source: Vivosecurity

# Informative Decompositions

Informative decompositions use what you know or data you can get to improve estimates in models.

## Informative Decompositions:

- **Systems:** you have fairly detailed knowledge of your applications, what data they have and the hardware it runs on. Some of the parameters of these systems would change your estimate of a risk.
- **Types of Impacts:** You separate confidentiality, integrity and availability events. You have an idea of business volumes like sales and other processes. If a breach or outage occurred, you can describe something about the consequences.
- **Staff:** You have knowledge of the number of employees, device loss rates, and some knowledge of what data they may have.
- **Vendors & Customers:** You know who the parties you interact with and you have some knowledge about them.
- **Insurance:** Any cyber-insurance will have detailed language regarding limitations, exclusions, etc.