RSAConference2016

San Francisco | February 29 – March 4 | Moscone Center

Connect to Protect

SESSION ID: GRC-W05

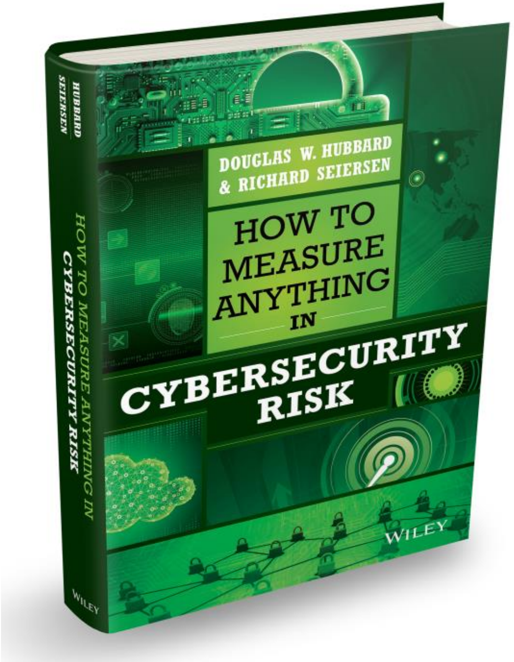
How To Measure Anything In Cybersecurity Risk

Doug Hubbard
CEO
Hubbard Decision Research

Richard Seiersen
GM Cyber Security & Privacy
GE Healthcare

#RSAC

Hubbard Decision Research



Hubbard Decision Research
2 South 410 Canterbury Ct
Glen Ellyn, Illinois 60137
www.hubbardresearch.com

The Coauthors

Richard Seiersen

Currently the General Manager of Cybersecurity and Privacy at GE Health Care. He is an analytics driven executive with ~20 years experience spanning subject matters in Cyber Security, Risk Management and Product Development. He is an active public speaker and tireless advocate for improving security through better, more quantitative, risk management. He has 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; 3rd 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).

The Biggest Cybersecurity Risk

Question: What is Your Single Biggest Risk in Cybersecurity?

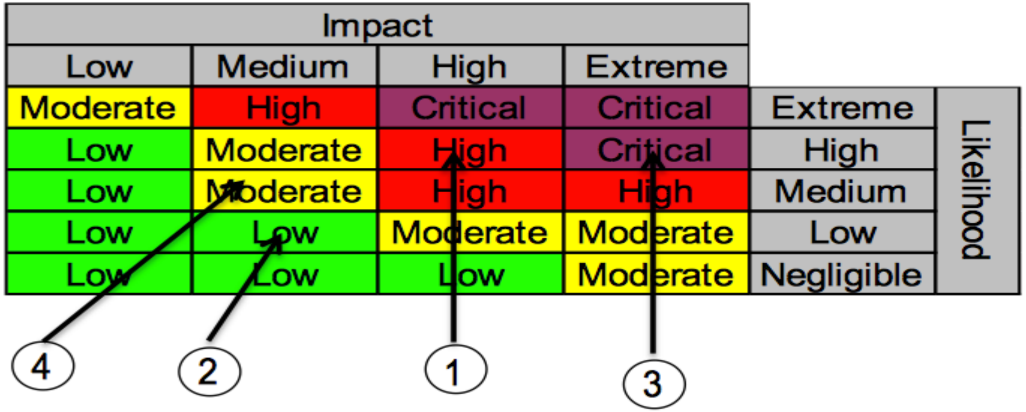
Answer: How You Measure Cybersecurity Risk

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**



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

Current Solutions

Most standards and certification tests promote risk analysis as a type of ordinal scoring method

The “**Risk Rating Methodology**” on **OWASP.org** states:

- “Once the tester has identified a potential risk and wants to figure out how serious it is, the first step is to estimate the "**likelihood**". At the highest level, this is a rough measure of how likely this particular vulnerability is to be uncovered and exploited by an attacker. It is not necessary to be over-precise in this estimate. *Generally, identifying whether the likelihood is low, medium, or high is sufficient* .”

Can Analysis Or Expertise Be A “Placebo”?

“The first principle is that you must not fool yourself, and you are the easiest person to fool.” — Richard P. Feynman

- Collecting more than a few data points on horses makes experts worse at estimating outcomes. (Tsai, Klayman, Hastie)
- Interaction with others only improves estimates up to a point, then they get worse. (Heath, Gonzalez)
- Collecting more data about investments makes people worse at investing. Collecting more data about students makes counselors worse at predicting student performance. (Andreassen)
- An experiment with a structured decision analysis method shows confidence increased whether decisions are improved or degraded. (Williams, Dennis, Stam, Aronson)

In short, we should *assume* increased confidence from analysis is a “placebo.” Real benefits have to be measured.

What The Research Says

- There is mounting evidence against (and none for) the effectiveness of “risk scores” and “risk matrices.”
- Fundamental misconceptions about statistical inference may keep some from adopting quantitative methods.
- Experts using even naïve statistical models outperform human experts who do not.

Note: Every improvement we are about to has already been adopted in several cybersecurity environments.

Summarizing Research on Ordinal Scales

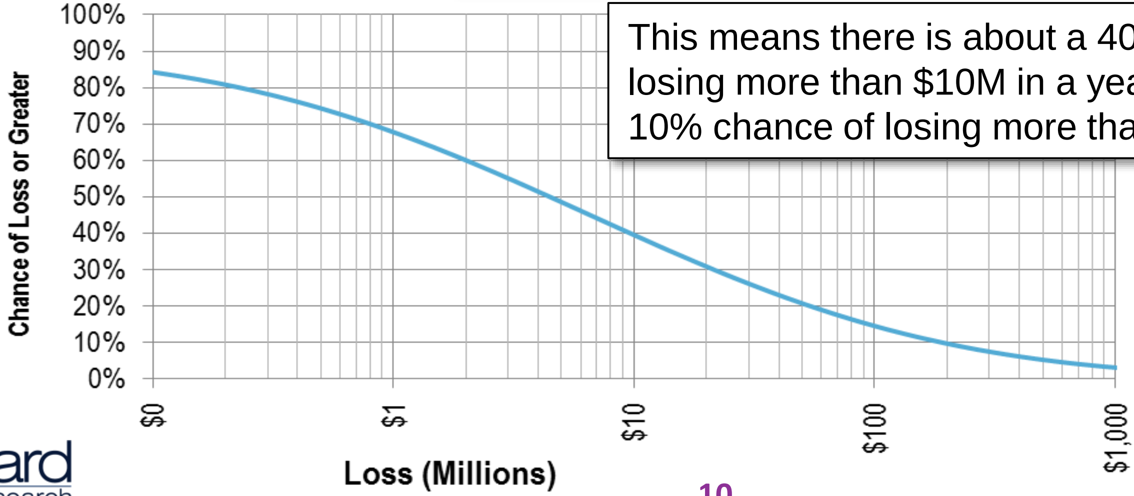
- Bickel et al. “The Risk of Using Risk Matrices”, *Society of Petroleum Engineers, 2014*
- They performed an extensive literature review to-date as well as a statistical analysis of RM used in Petroleum Engineering Risk (which are nearly identical to RM’s in Cyber) – including computing a “Lie Factor” of the degree of distortion of data.
  - “How can it be argued that a method that distorts the information underlying an engineering decision in nonuniform and uncontrolled ways is an industry best practice? <u>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.’</u>

What If We Could *Actually Measure Risk* in Cybersecurity?

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



This means there is about a 40% chance of losing more than $10M in a year and about a 10% chance of losing more than $200M.

Hubbard Decision Research

Why Not Better Methods?

- Cybersecurity is too complex or lacks sufficient data for quantitative analysis…

  …yet can be analyzed with unaided expert intuition or soft scales.

- Probabilities can't be used explicitly because _______ ….

  …yet we can *imply* probabilities with ambiguous labels.

Remember, softer methods never *alleviate* a lack of data, complexity, rapidly changing environments or unpredictable human actors…

…they can only *obscure* it.

A Major Fallacy Regarding Comparing Methods

- Don't make the classic "Beat the Bear" fallacy.

  Exsupero Ursus

- If you doubt the effectiveness of quantitative methods, remember, all you have to do is outperform the alternative:
- …unaided expertise or soft scoring methods.

Your Intuition About Sample Information Is Wrong!

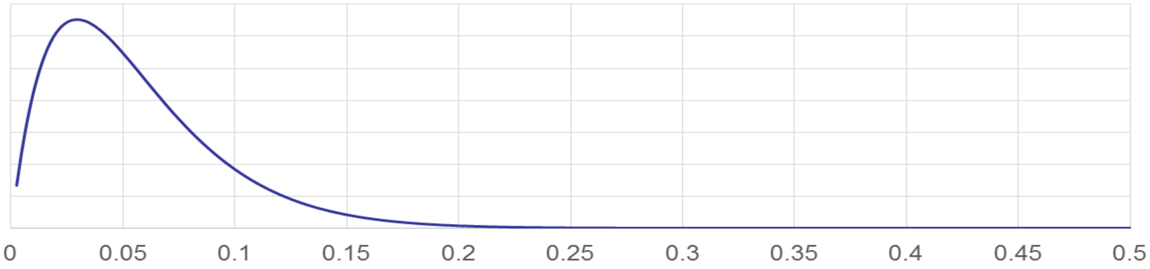
- Cybersecurity experts are not immune to widely held misconceptions about probabilities and statistics – especially if they vaguely remember some college stats.
- These misconceptions lead many experts to believe they lack data for assessing uncertainties or they need some ideal amount before anything can be inferred.

"Our thesis is that people have strong intuitions about random sampling…these intuitions are wrong in fundamental respects...[and] are shared by naive subjects and by trained scientists"

Amos Tversky and Daniel Kahneman,
Psychological Bulletin, 1971

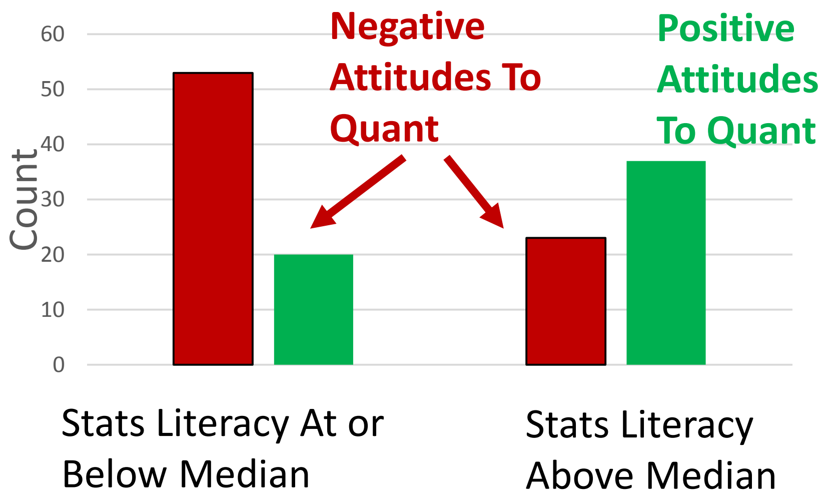
You Need Less Data Than You Think

- A beta distribution computes the probability of a frequency being below a given amount (e.g. chance that rate of occurrence is <2/100)
- In Excel it can be written as “=Betadist(frequency,alpha,beta)”
- A uniform prior can be made with alpha=1 and beta=1. This can be used as a starting point for maximum uncertainty.
- “Hits” and “Misses” can be simply added to the priors (=Betadist(frequency,hits+1,misses+1))

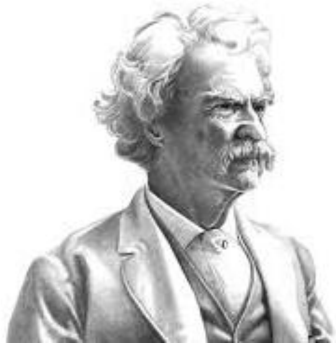


Survey Results: Stats Concepts Quiz

- We conducted a survey of 171 Cybersecurity professionals
- One Finding: Strong opinions against “quant” are associated with poor stats understanding.



“It’s not what you don’t know that will hurt you, it’s what you know that ain’t so.”
Mark Twain



Historical Models - Still Better Than Experts

When experts assess probabilities, many events ". . .are perceived as so unique that past history does not seem relevant to the evaluation of their likelihood." Tversky, Kahneman, *Cognitive Psychology* (1973)

Yet, Historical models routinely outperform experts in a variety of fields (even considering "Black Swans")

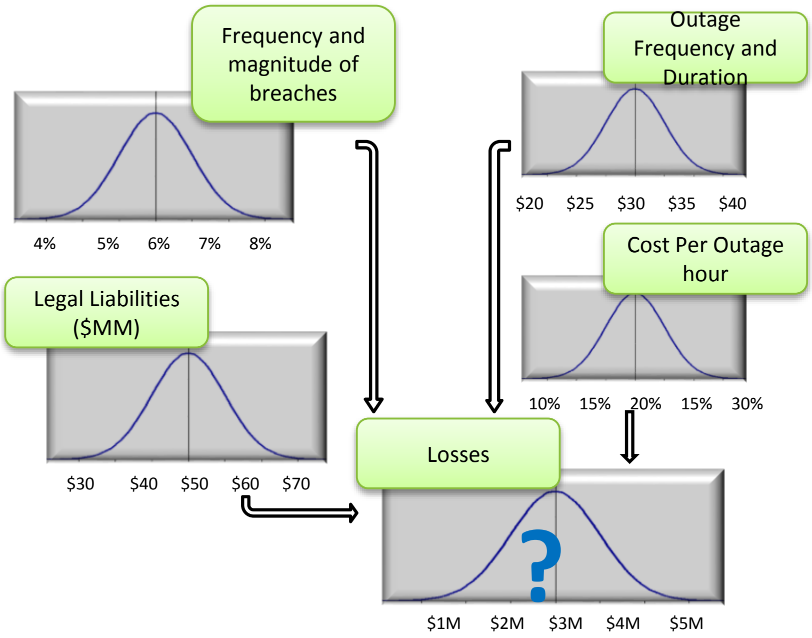
Paul Meehl assessed 150 studies comparing experts to statistical models in many fields (sports, prognosis of liver disease, etc.).

"There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one."

Philip Tetlock tracked a total of over 82,000 forecasts from 284 political experts in a 20 year study covering elections, policy effects, wars, the economy and more.

"It is impossible to find any domain in which humans clearly outperformed crude extrapolation algorithms, less still sophisticated statistical ones."

Monte Carlo: How to Model Uncertainty in Decisions



- Simple decomposition greatly reduces estimation error for estimating the most uncertain variables (MacGregor, Armstrong, 1994)
- As Kahneman, Tversky and others have shown, we have a hard time doing probability math in our heads
- In the oil industry there is a correlation between the use of quantitative risk analysis methods and financial performance – and the improvement started after using the quantitative methods. (F. Macmillan, 2000)
- Data at NASA from over 100 space missions showed that Monte Carlo simulations beat other methods for estimating cost, schedule and risks (I published this in *The Failure of Risk Management* and *OR/MS Today*).

A Simple One For One Substitution

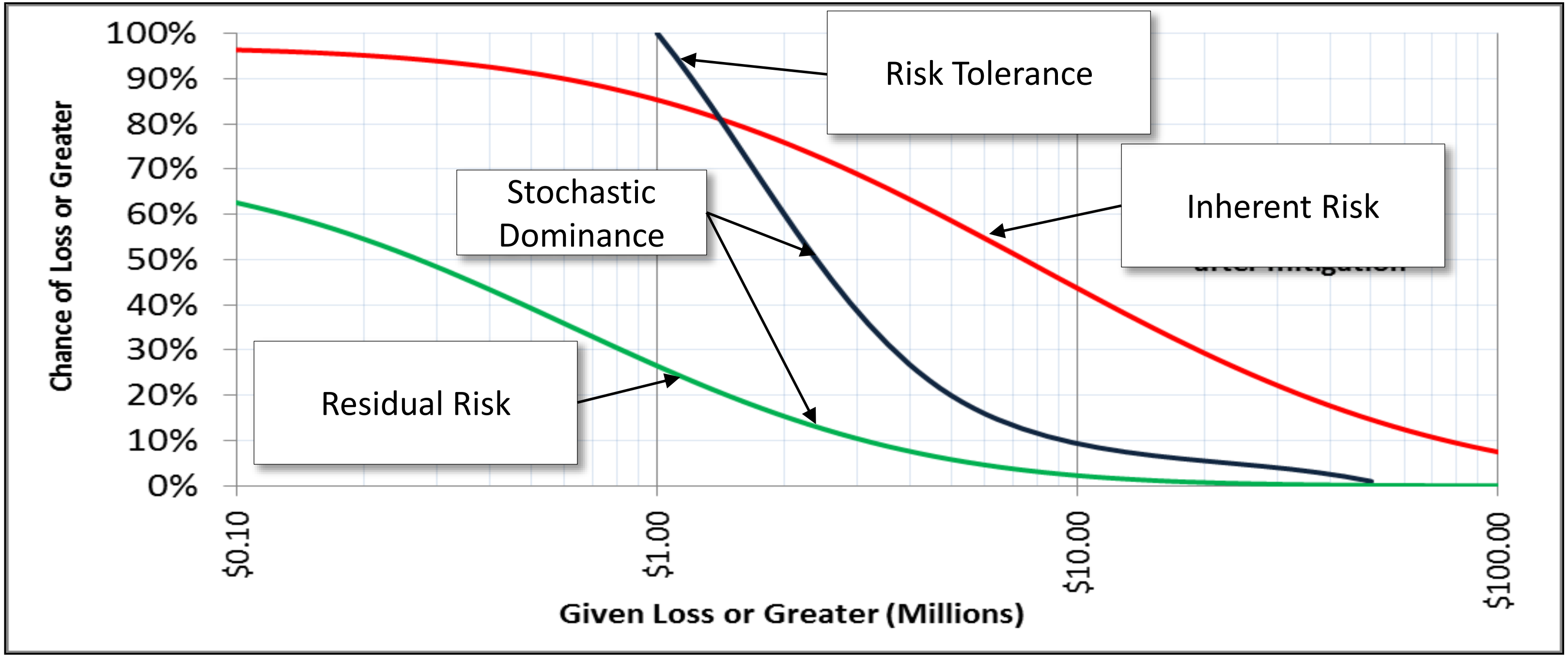
Event	Event Probability (per Year)	Impact (90% Confidence Interval) Lower Bound	Upper Bound	Random Result (zero when the event did not occur)
AA	.1	$50,000	$500,000	0
AB	.05	$100,000	$10,000,000	$8,456,193
AC	.01	$200,000	$25,000,000	0
AD	.03	$100,000	$15,000,000	0
AE	.05	$250,000	$30,000,000	0
AF	.1	$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 represented as a Loss Exceedance Curve.

Loss Exceedance Curves: Before and After

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



Overconfidence

"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

- 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 <u>can be taught</u> 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.

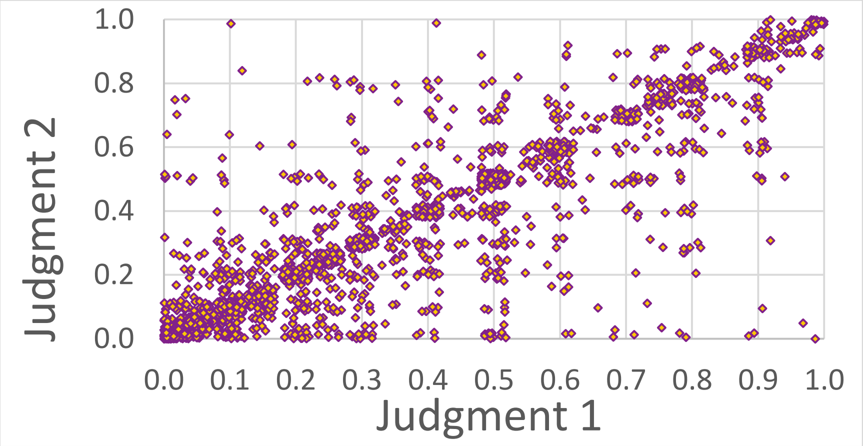
Inconsistency vs. Discrimination

- *<u>Discrimination</u>* is how much your estimates vary when given different information.
- *<u>Inconsistency</u>* 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).

Inconsistency Measurement Results

- 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 1st to 2nd Estimates of Cyber risk judgements by same SME

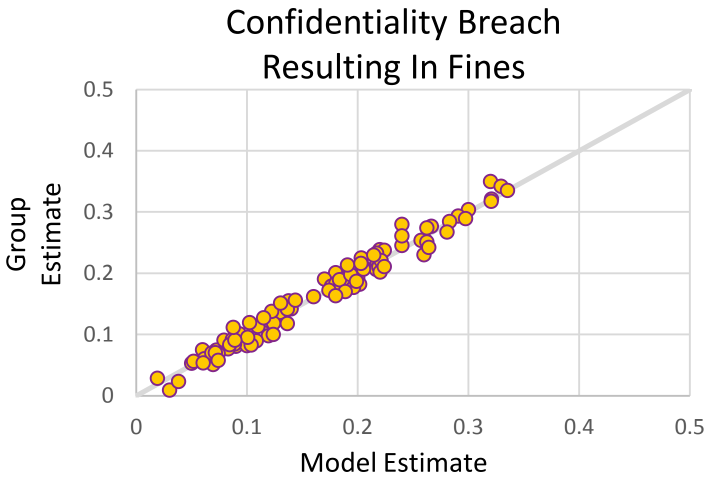


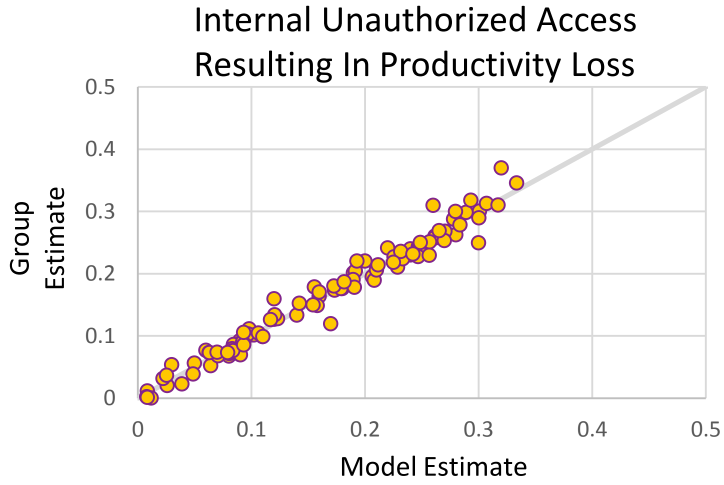
21% of variation in expert responses are explained by *inconsistency.*

79% are explained by actual information given

Modeling Group Estimates of IT Security Event Likelihood

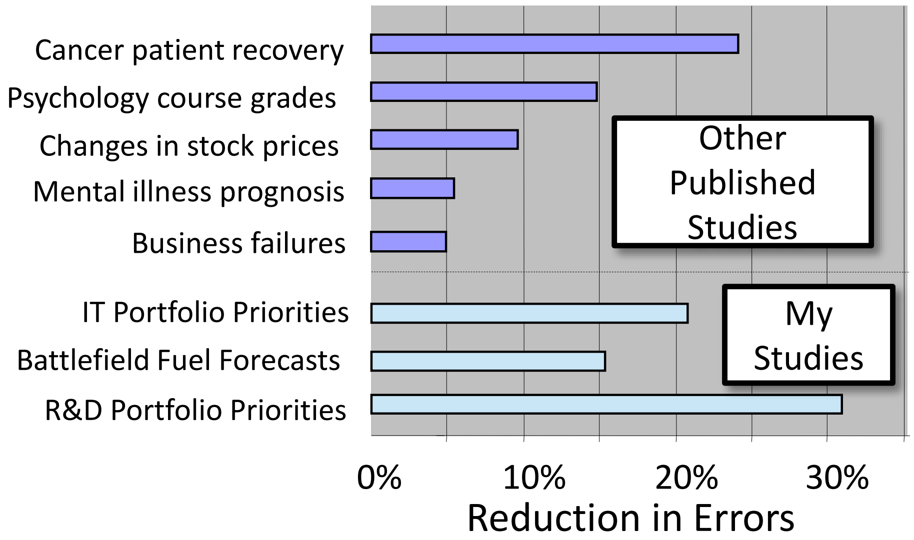
Examples of Models vs. Group Averages: Probabilities of different security events happening in the next 12 months for various systems prior to applying particular controls.





- 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.

Rasch (Logodds) Model

- A Rasch 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(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%	40.0%	12.0%
P(E|~X)	5.5%	9.0%	3.0%	8.0%
P(X)	16.0%	20.0%	19.0%	50.0%
Test P(E)	10.1%	10.2%	10.0%	10.0%
Logodds change|X	1.5339	0.4626	1.7918	0.2048
Logodds change|~X	-0.6466	-2.3136	-3.4761	-2.4423

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!

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

Risk	Likelihood / Yr	Impact / Yr	Mitigation Effectiveness	Mitigation Cost / Yr	Mitigation ROI	Action
Risk 1	37%	$2M to $40M	95%	$725K	725%	Mitigate
Risk 2	11%	$50K to $400K	100%	$95K	-80%	Track
Risk 3	34%	$5M to $80M	90%	$2.5M	329%	Monitor
Risk 4	29%	$500K to $20M	98%	$375K	437%	Mitigate

- The optimal solution would be to mitigate Risks1 & 4 first.
- If you have the resources, then mitigate Risk 3.
- Risk 2 is not worth fixing.

Call To Action For Cybersecurity!

- Organizations should stop using risk scores and risk matrixes and standards organizations should stop promoting them
- Adopt simple probabilistic methods now: They demonstrate a measurable improvement over unaided intuition and they have already been used. So there is no reason not to adopt them.
- Build on simple methods when you are ready – always based on what shows a measurable improvement.

Questions?

Hubbard Decision Research
www.hubbardresearch.com
info@hubbardresearch.com