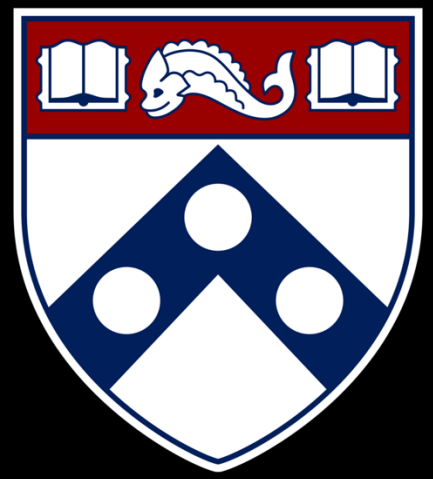
Empirical Security & Privacy, for Humans

UPenn CIS 7000-010

11/18/2025



Quantifying Cyber Risk

Reading

SoK: Quantifying Cyber Risk

Daniel W. Woods
University of Innsbruck
Innsbruck, Austria
daniel.woods@uibk.ac.at

Rainer Böhme
University of Innsbruck
Innsbruck, Austria
rainer.boehme@uibk.ac.at

***Abstract*—This paper introduces a causal model inspired by structural equation modeling that explains cyber risk outcomes in terms of latent factors measured using reflexive indicators. First, we use the model to classify empirical cyber harm studies. We discover cyber harms are not exceptional in terms of typical or extreme losses. The increasing frequency of data breaches is contested and stock market reactions to cyber incidents are becoming less damaging over time. Focusing on harms alone breeds fatalism; the causal model is most useful in evaluating the effectiveness of security interventions. We show how simple statistical relationships lead to spurious results in which more security spending or applying updates are associated with greater rates of compromise. When accounting for threat and exposure, indicators of security are shown to be important factors in explaining the variance in rates of compromise, especially when the studies use multiple indicators of the security level.**

***Index Terms*—cyber risk, security metrics, cyber harm, control effectiveness, science of security, causal model, structural equation modeling**

I. Introduction

Unsupported claims about the increasing risk of cyber attacks pervade introductions to security talks and papers. Organisations are expected to invest more in security even though research has inconsistently demonstrated how interventions reduce risk. This state of affairs leads to perceptions that cyber risk is more art than science.

With this in mind, our paper aims to systematise what is known about quantifying cyber risk. Risk estimates can justify additional resources for mitigation or be used to guide post-incident response. The term *cyber* will bristle with many in the security community. However, it is the concept of choice for policymakers and business leaders who make many of the decisions that security research should hope to influence. Such decisions are premised on foundational questions like:

RQ1 How much harm results from cyber incidents?
RQ2 Which security interventions effectively reduce harm?
RQ3 Have these answers changed over time?

Whereas security vendors scramble to provide self-interested answers with shaky methodologies [7, 82], this paper finds answers in empirical studies of real-world security outcomes. We systematise the literature using a causal model linking latent variables for security, exposure, and threat to security outcomes. The proposed model captures empirical cyber risk research ranging from machine learning models predicting web server compromise through to finance studies quantifying shareholder losses resulting from cyber incidents.

We focus on classifying studies quantifying cyber risk in organisations. The term *cyber risk* has two components, *risk* describes possible negative consequences (*harm*) weighted by the probability of occurrence. *Cyber* restricts our scope to incidents caused by logical (as opposed to physical) force [17]. Under this definition a fire (physical force) in a data centre (information harm) is not a cyber risk, whereas fire damage (physical harm) caused by compromised control systems (logical force) would be. *Incidents* within scope include denial of service attacks, machine and web-resource compromise, and organisational incidents. Associated harms range from lost shareholder value to ransomware payments to wasted time.

Our literature search first identified relevant works in top security conferences and the *Workshop on the Economics of Information Security*. We used backwards and forwards reference searches to identify additional relevant works until saturation was reached. Doing so captured relevant studies from disciplines including law, information systems, finance, and physics. We included studies that empirically measure real-world compromise or harm affecting organisations, which is a minority approach within the science of security [60, p. 12]. Studies providing promising ways of measuring security, exposure or threat were also included even when harm was not considered. We point readers to Anderson et al. [7] for aggregate estimates of cybercrime costs, and Dambra et al. [32] for cyber risk transfer research.

Section II introduces the causal model. Section III surveys harm studies speaking to **RQ1**. Section IV identifies mitigation studies that address **RQ2**. Temporal trends are identified throughout (**RQ3**). Section V discusses progress towards **RQ1–3**, model limitations, and future work.

II. A causal model of cyber risk

Risk is unobservable but we can indirectly measure its realisation as losses. Figure 1 uses artificial data to illustrate the stochastic relationship; the highest observed loss has multiple twins with similar security levels but much smaller losses.

We thank Stefan Laube and the anonymous reviewers for their thoughtful comments and the cited authors who responded to our request for feedback, especially Ben Stock, Camelia Simoiu, Kanta Matsuura, Martin Loeb, and Stefan Savage. The causal model grew out of Dagstuhl Seminar 16461, in particular the breakout group chaired by the second author. It was refined at the Empirical Cybersecurity Research Winter School in Obergurgl, Austria, 2019. The first author thanks Tom Verstraten for extolling the value of related work. The project is funded by the European Commission's call H2020-MSCA-IF-2019 under grant number 894700.

The problem

- Unsupported claims about increasing cyber risk are common
- Research inconsistently demonstrates how interventions reduce risk
- "Cyber risk is more art than science"
- Need: Systematize what we know about quantifying cyber risk

Three research questions

- RQ1: How much harm results from cyber incidents?
- RQ2: Which security interventions effectively reduce harm?
- RQ3: Have these answers changed over time?

A naïve model relating loss to security level

- Simple regression (blue line): more security implies more losses?!
- Problem: Confounding variables (especially threat level)

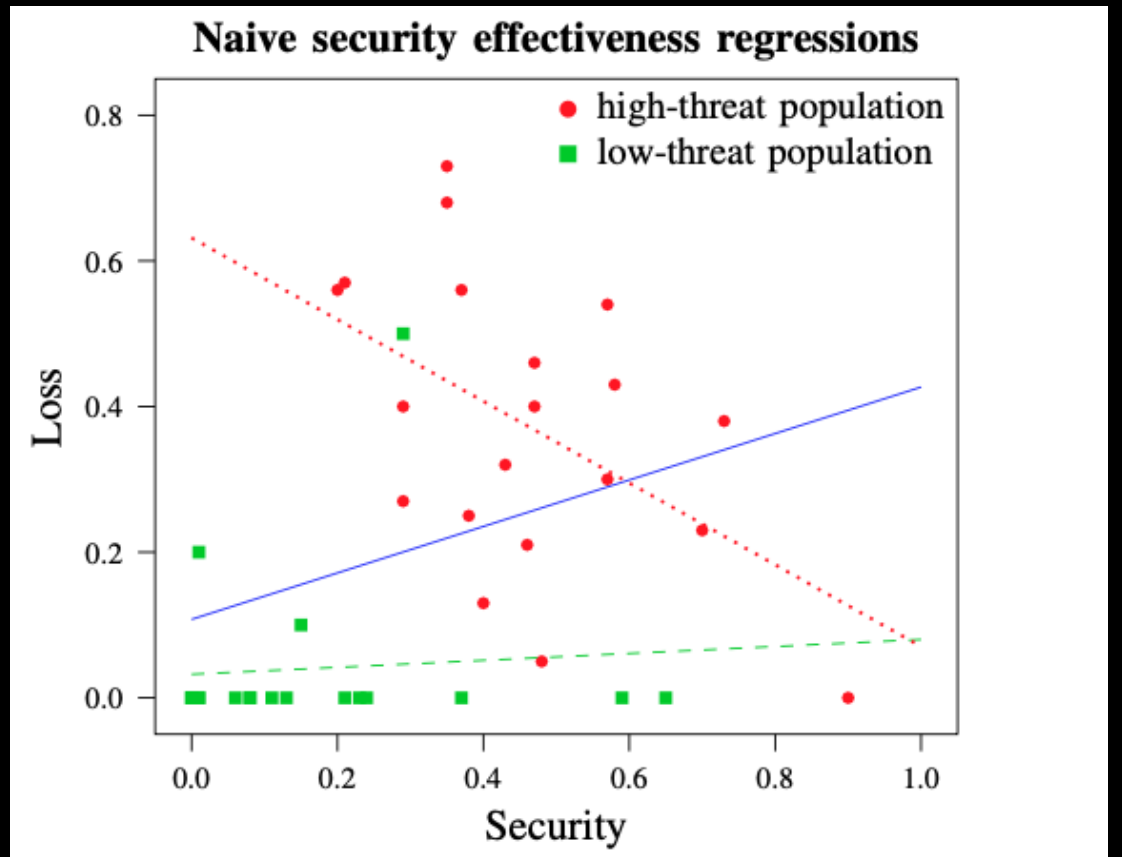


Fig. 1. The solid blue line fails to account for threat level, which may lead the high-threat population to under estimate the effectiveness of security.

A simple causal model of cyber risk

- **Threat**: The motivation, capability and activity of adversaries
- **Harm**: Negative consequences resulting from compromise
- **Exposure** *amplifies* E(+) the relationship between Threat and Harm
- **Security** *moderates* S(−) the relationship between Threat and Harm
  - Security has **reflexive indicators** I_1, I_2, … I_k, which have corresponding **measurements** m_1, m_2, … m_k

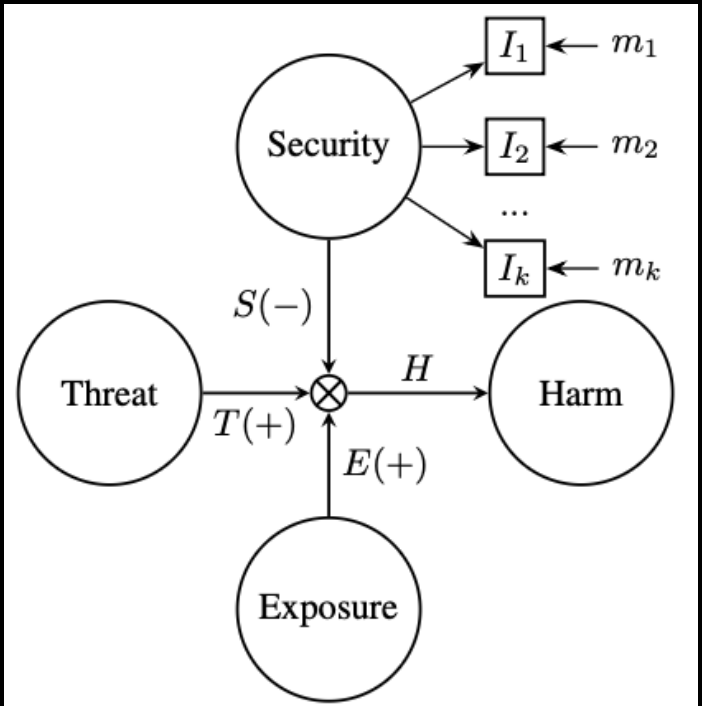
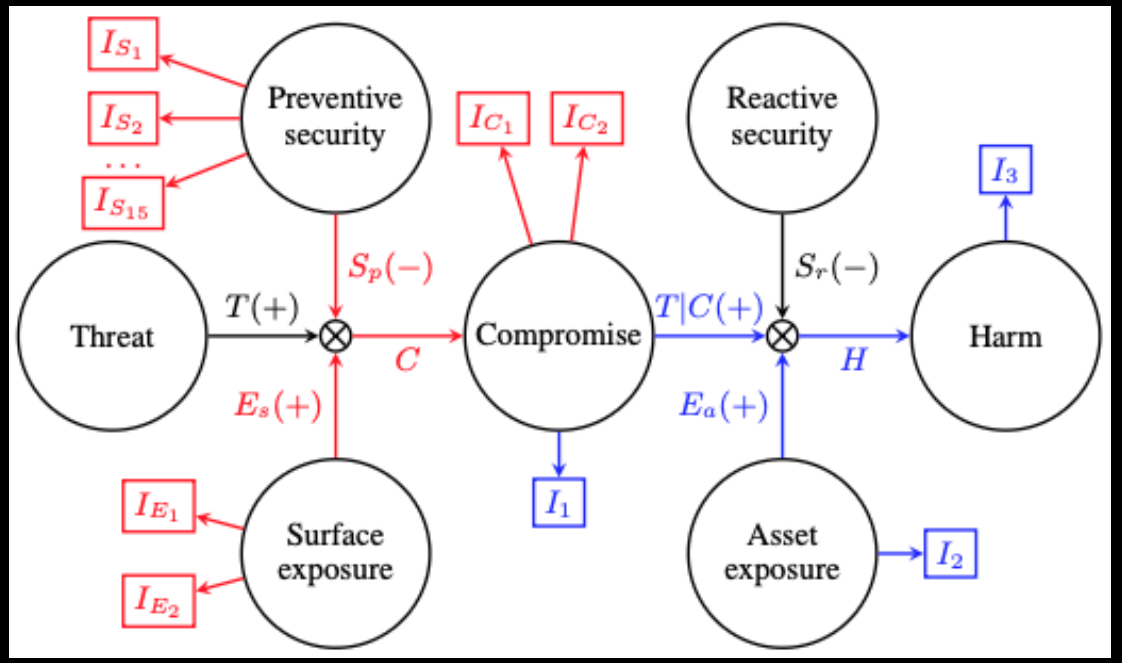


Fig. 2. A high-level causal model of cyber risk.

A more sophisticated model of cyber risk

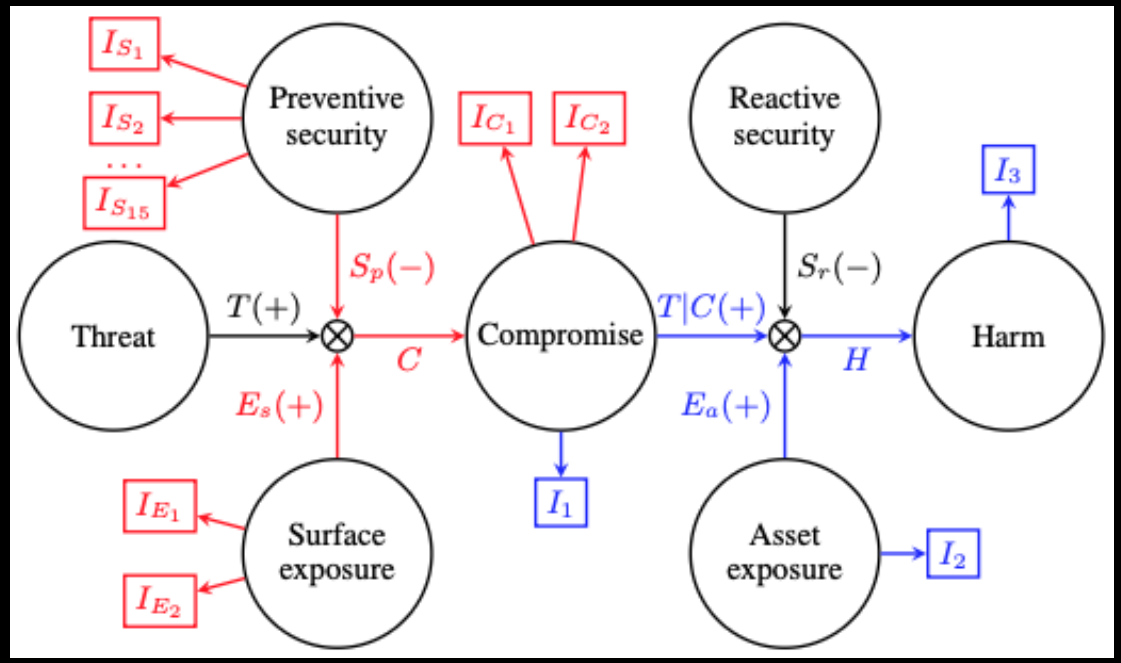
- Introduces **Compromise**: Violation of victim security goal
- Subdivides **Security**
  - **Preventive**: Interventions reducing the ease of compromise
  - **Reactive**: Interventions reducing the impact of compromise
- Subdivides **Exposure**
  - **Surface**: Factors increasing potential vectors of compromise
  - **Asset**: Factors increasing the value of what can be compromised
- Adds more reflexive indicators



A more sophisticated model of cyber risk

TABLE VI
TECHNICAL INDICATORS [116] CORRESPONDING TO I_x IN FIGURE 3

	Technical indicator
I_{C_1}	# domains in phishing blacklist
I_{C_2}	# domains in malware blacklist
I_{E_1}	# IPs on shared hosting
I_{E_2}	# domains on shared hosting
I_{S_1}	HTTP server version
I_{S_2}	SSL version
I_{S_3}	Admin panel version
I_{S_4}	PHP version
I_{S_5}	OpenSSH version
I_{S_6}	CMS version
I_{S_7}	HttpOnlyCookie
I_{S_8}	X-Frame-Options
I_{S_9}	X-Content-Type-Options
$I_{S_{10}}$	Mixed-content inclusions
$I_{S_{11}}$	Secure cookie
$I_{S_{12}}$	Content-Security-Policy
$I_{S_{13}}$	HTTP Strict-Transport-Security
$I_{S_{14}}$	SSL-stripping vulnerable form
$I_{S_{15}}$	Browser XSS protection



Parts in red modeled in a prior study.

S. Tajalizadehkhoob, T. Van Goethem, M. Korczyinski, A. Noroozian, R. Boehme, T. Moore, W. Joosen, and M. van Eeten. Herding vulnerable cats: A statistical approach to disentangle joint responsibility for web security in shared hosting. In CCS, 2017.

[RQ1: Harm] Data breach studies

TABLE II
THE OFTEN CONTRADICTORY FINDINGS FROM DATA BREACH STUDIES.

				Breach frequency		Breach size		∞
Reference	**Type of data**	n	**Years**	**Distribution**	**Trend**	**Distribution**	**Trend**	**Moment**
Curtin et al. (2008) [30]	N + M (USA)	899	2005–07	?	↗	?	?	?
Maillart et al. (2010) [83]	N + M (USA)	956	2000–08	?	↗	Power law	→	Yes
Edwards et al. (2016) [36]	N + M (USA)	2253	2005–15	Negative binomial	→	Lognormal (M)	→	No
Wheatley et al. (2016) [127]	M (World)	5365	2007–15	Poisson gen LM	→ (USA)	DT power law (t)	↗	Yes*
Eling et al. (2017) [40]	N + M (USA)	2266	2005–15	Negative binomial	↘	Skew-normal	→	No
Xu et al. (2018) [131]	M (USA)	600	2005–17	ARMA/GARCH	↗	Gen Pareto (t)	→	No
Wheatley et al. (2019) [128]	N + M (USA)	1713	2005–17	Negative binomial	→	Pareto	→, ↗ (M)	?
Carfora et al. (2019) [23]	N + M (USA)	5724	2005–17	Negative binomial	↗ (M)	Skew/Lognormal	?	No
Farkas et al. (2020) [43]	N + M (USA)	6160	2005–19	Binomial	?	Lognormal (t)	?	Yes

N/M = Negligent/malicious breach, (t) = distribution of the tail, LM = linear model, DT = double truncated, * = without maximum, ? = not reported.

- 10 years of studies, same data, contradictory results
  - Frequency trends: decreasing, stable, or increasing?
  - Size trends: stable or increasing?
- Heavy-tailed distributions, but unclear mapping to financial cost

[RQ1: Harm] Stock market reactions

- Effect decreasing over time (from -7.9% to -0.05%)
  - Firms learned to manipulate announcements
- Evidence of strategic timing and insider trading

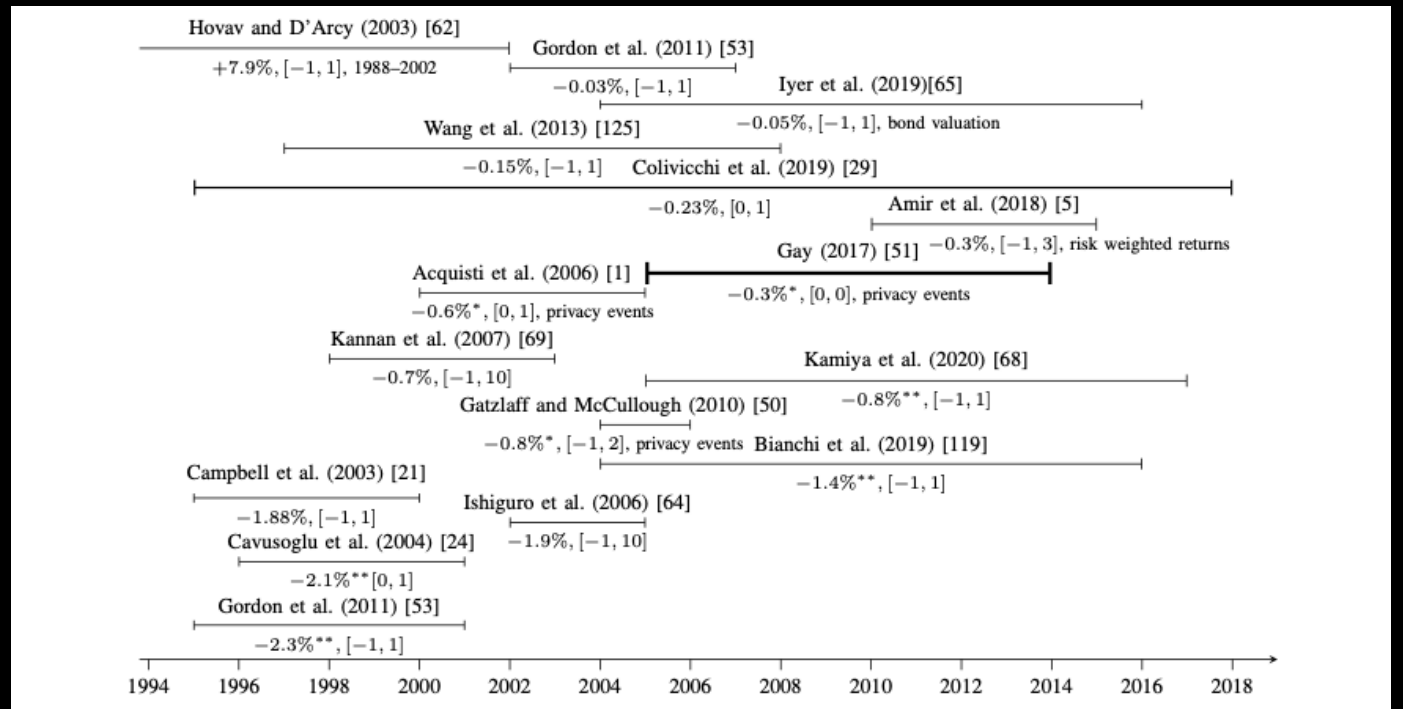


Fig. 4. Impact of security incident disclosure on firm stock market value. The effect is reported as cumulative abnormal returns (CAR), bar placement and thickness describe sample period and size, event window reported as [Days before, days after], statistical significance levels * $p \leq 0.05$, ** $p \leq 0.01$.

19 studies from 1988-2019

[RQ1: Harm] Are cyber harms exceptional?

- No.
- Mean cyber loss: $4.1M-$43M (varies by sample)
- Cyber losses smaller and less heavy-tailed than non-cyber operational losses
  - Compare to fraud, theft, bad debt
- Typical breaches less extreme than media reports suggest

TABLE I
OVERVIEW OF DIFFERENT APPROACHES TO QUANTIFYING CYBER HARM.

Unit of analysis	# of studies	Econ loss	Sample size	Earliest study	Earliest sample
Public reports (Section III-A)					
Data breach	9	✗	600–6160	2008	2000
Operational loss	3	✓	341–1579	2015	$<$2003
Cyber incident	1	✓	2216	2016	2005
Private reports (Section III-B)					
Internal incident	2	✗	1800–23000	2010	1996
Insurance claim	1	✗	70	2019	2015
Crime reports	1	✓	7925	2020	2017
Firm survey response	3	✓	664–4209	2012	2012
Individual survey response	5	✓	1500–64287	2014	2010s
Externally observed (Section III-C)					
Legal case	2	✗	19–230	2011	1999
Legal case	1	✓	118	2017	2010
Bitcoin transaction	3	✓	10m	2014	2009
Criminal forum post	2	✓	13m	2007	2006
Insurance prices	1	✓	6828	2019	2007
Stock market reaction	19	✓	43–542	2003	1988
System-wide harm (Section III-D)					
Multi-party incident	1	✓	800	2019	2008

[RQ2: Effect] Measuring security is hard

- Single indicators fail:
  - Certifications: 86% of PCI-DSS certified sites violated requirements
  - Security $$ budgets: positively correlated with breaches
- Need: Multiple technical indicators
  - Self-reported indicators (SeBIS scale) for individuals predict behavior but not linked to harm outcomes. Costly to collect at organization-level

[RQ2: Effect] Measuring threat: 3 approaches

- Time-based: Track malicious activity over time
- Target-based: Study who gets attacked
  - E.g., larger banks targeted more
- Researcher intervention: Honeypots, controlled experiments
  - Challenge: Rational attackers use undetectable malware

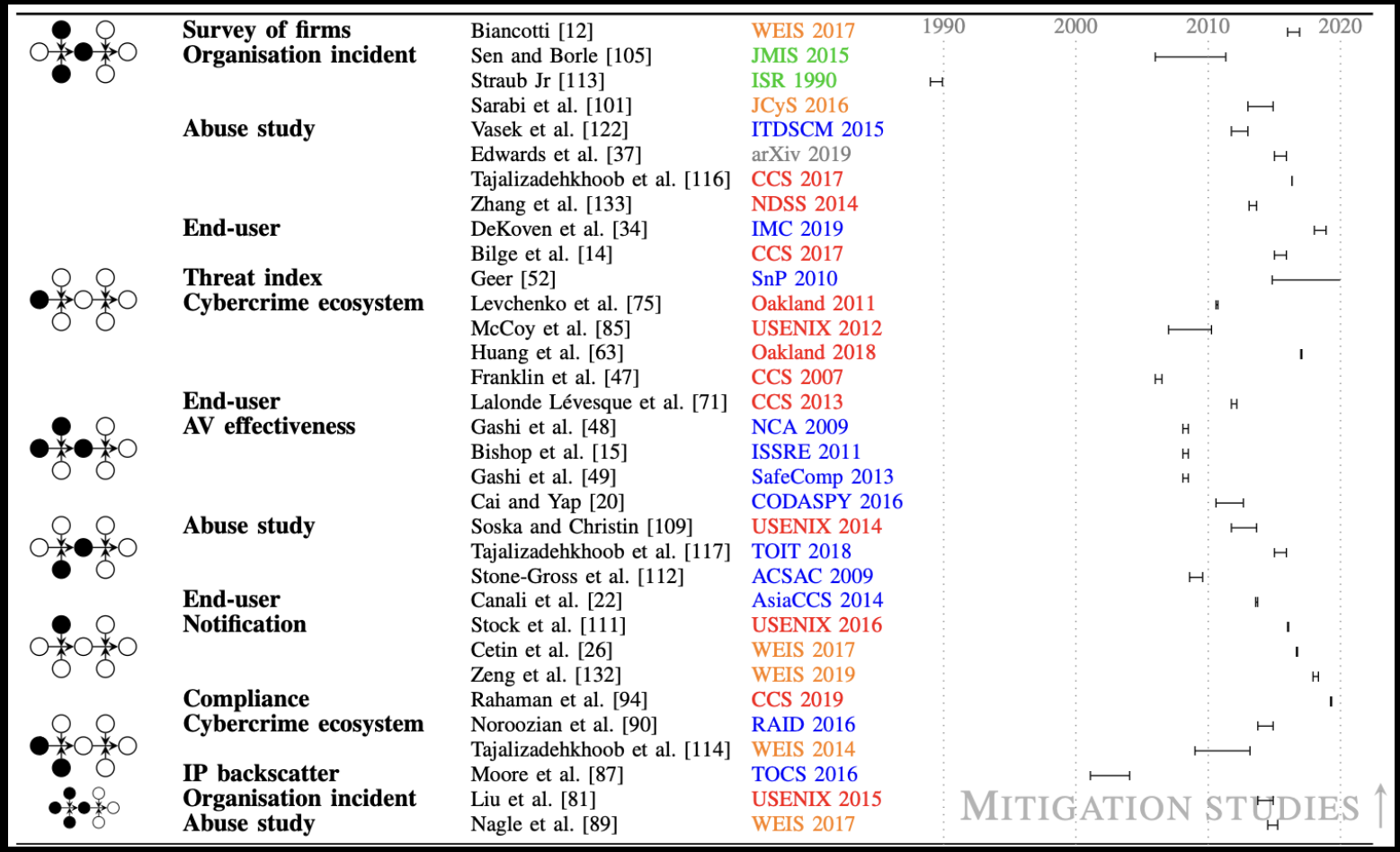
[RQ2: Effect] Measuring exposure

- Unit of analysis matters (AS vs. hosting provider)
- Exposure can be highly influential in predicting harm. Example:
  - 1 variable of hosting provider explains 20% of phishing abuse, but 4 variables explain 84% of abuse
  - Can train a classifier to identifier compromised website with 66%/17% true/false positive rates

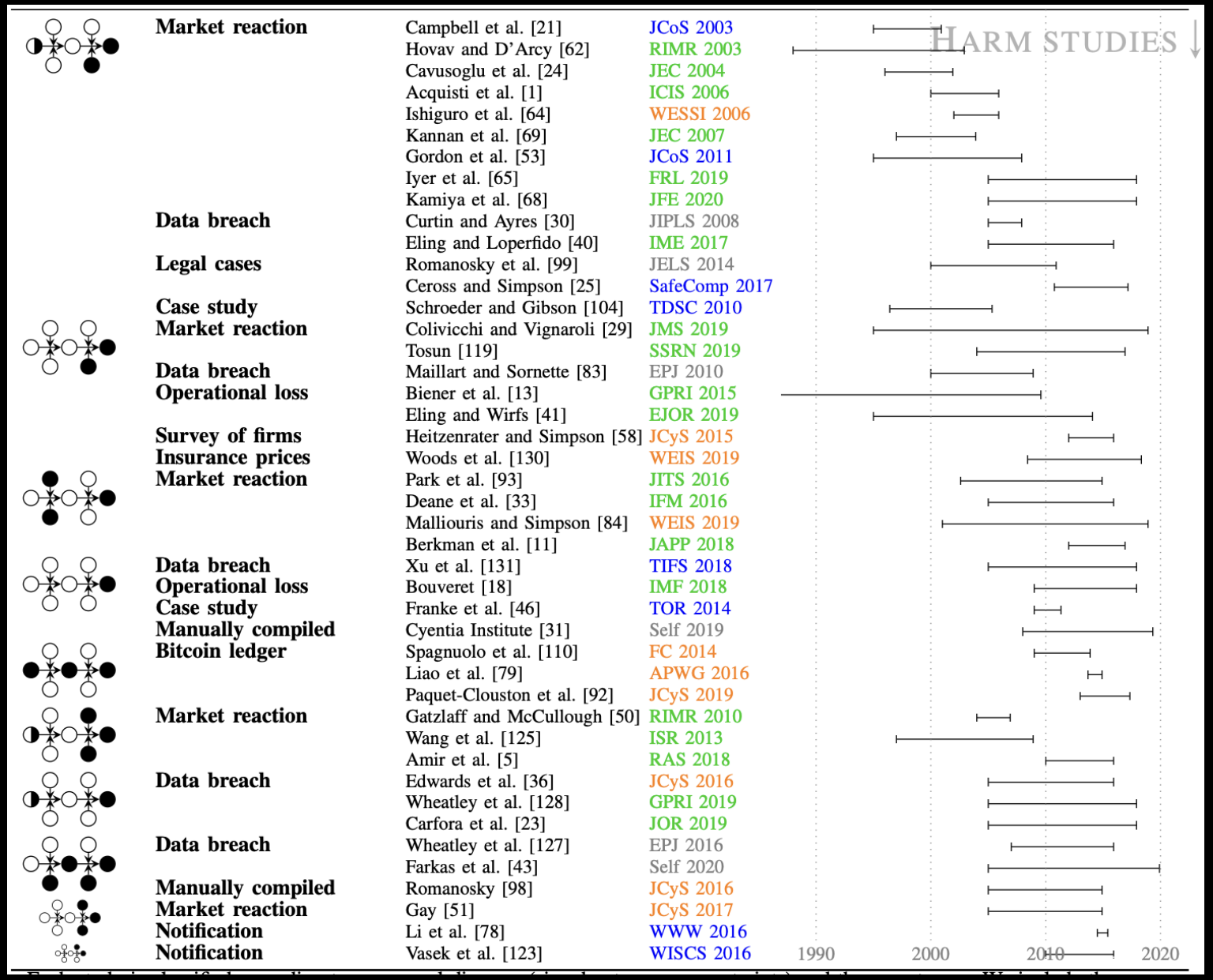
[RQ2: Effect] Structural relationships

- Between-subject designs can be misleading
  - Example: Updated software associated with *more* compromise
    - Until you control for threat level
    - Then: 22.6% of updated sites re-compromised vs. 33.5% never-updated
- Within-subject designs help control confounders
- The relative infrequency of compromise undermines statistical power

[RQ2] Evidence for security effectiveness



[RQ2] Evidence for security effectiveness



RQ3: Temporal trends

- Harm studies: longer time windows (up to 20 years)
- Mitigation studies: brief windows (often <3 years)
- Stock market reactions: decreasing over time
- Cyber insurance prices: trending downward 2008-2018
- Data breach frequency: stable overall, increasing for malicious breaches

Key findings

- RQ1: Cyber harms not exceptional; typical losses smaller than claimed
- RQ2: Security effective *only* when controlling for threat and exposure
- RQ3: Some evidence of decreasing stock market impact; otherwise limited temporal data
- Single indicators misleading; multiple indicators essential
- But lots we don’t know
  - Systemic cyber risk: insufficient observations
  - Causal effects of specific interventions
  - How to prioritize security investments
  - Long-term trends in mitigation effectiveness

Implications

- For practice
  - Don't underestimate exposure (very predictive)
  - Avoid single-indicator solutions from vendors
  - Security teams need resources for diverse tasks
  - Be skeptical of exceptional harm claims
- For research
  - Use the causal model framework
  - Include threat *and* exposure controls
  - Use multiple indicators of security
  - Longer time windows needed for mitigation studies
  - Consider randomized controlled trials (notification studies)

Future directions

- No data breach studies link security to compromise/harm
- Need better methods for systemic risk
- Institutional data collection and sharing
- Move beyond prediction to causal understanding