Survival of the safest:
Examining organizational risk factors for cybersecurity incidents

Rachel Whitman, Ana Kriletic, Thomas Wilmore, Kate Conkey, and Dr. Daniel Svyantek

INTRODUCTION

Given that employees pose a large threat to organizational cybersecurity, much research attention has been directed to identifying individual risk factors for cybersecurity noncompliance and misbehavior. However, no study to date has formally examined how the risk of organizational cybersecurity incident changes over time, or how organizational characteristics affect this risk. Examining changes in risk across time has become a necessity due to the abundant evidence that cybersecurity incidents are increasing in both frequency and severity. Previously-employed methods such as odds ratios fail to account for the time-based component needed for properly analyzing the continuously-changing threat of cyberattacks. Therefore, proposed study aims to answer the question*: what organizational factors are associated with cyber breaches?*

HYPOTHESES

H1: Industry type will significantly influence the hazard ratio of cyberbreaches.

H2: Organizational revenue will be associated with an increased hazard ratio.

H3: Handling sensitive information will be associated with an increased hazard ratio.

PROPOSED METHODOLOGY

Sample:

Organizations listed in the top Fortune 1000 from 2005-2019 will be used as the population at risk. Business records will be collected from publicly-available sources. Event data regarding the occurrence of a cybersecurity incident will be collected from the dataset maintained by Privacy Rights Clearinghouse for those years, which currently contains information on over 9,000 security incidents.

ANALYSES

The proposed study aims to conduct a survival analysis (SA) of cybersecurity events across the past decade, examining broad factors that impact the changing probability of cyberincidents. In particular, the proposed study will examine associations between cyberbreaches and industry type, annual revenue, and the sensitivity of information handled in the organization. To analyze the impact of organizational factors on the risk of cyberincident, the current study will record security breaches (or lack thereof) using publically-available data recorded by Privacy Rights Clearinghouse.

Sample Data Set:

2004 Rank	Business Name	Revenues ($mil	Profits ($million	Org Type	Data Breach (0=	Industry	Censor
1	Wal-Mart Stores	258,681.00	9,054.00	retail	1	4	2015
2	Exxon Mobil	213,199.00	21,510.00	energy	0	1	2019
3	General Motors	195,645.20	3,822.00	automotive	0	4	2019
4	Ford Motor	164,496.00	495	automotive	0	4	2019
5	General Electric	134,187.00	15,002.00	energy	0	1	2019
6	ChevronTexaco	112,937.00	7,230.00	energy	0	1	2019
7	ConocoPhillips	99,468.00	4,735.00	energy	0	1	2019
8	Citigroup	94,713.00	17,853.00	financial	1	6	2005
9	Intl. Business Machines	89,131.00	7,583.00	healthcare	1	5	2011
10	American Intl. Group	81,300.00	9,274.20	financial	0	6	2019
11	Hewlett-Packard	73,061.00	2,539.00	tech, retail	1	4	2006
12	Verizon Communication	67,752.00	3,077.00	telecoms	1	2	2016
13	Home Depot	64,816.00	4,304.00	retail	1	4	2014
14	Berkshire Hathaway	63,859.00	8,151.00	everything	0	2	2019
15	Altria Group	60,704.00	9,204.00	tobacco	0	4	2019
16	McKesson	57,129.20	555.4	healthcare	0	5	2019
143	AdvancePCS	14,110.90	168.4	healthcare	0	5	2003
144	Emerson Electric	13,999.00	1,089.00	energy	0	1	2019
145	UAL	13,724.00	-2,808.00	transportation	0	3	2010

Table 1: *Preliminary Covariate Results.*

Variable	Coef	Exp(coef)	SE(coef)	Z	p
Industry	0.5965	1.8158	0.2404	2.482	0.0131*
Information	-0.8601	0.4231	0.6456	-1.332	0.1828
Profit	0.0183	1.0185	0.0063	2.915	0.0036*
Revenue	0.0005	1.0005	0.0005	0.911	0.3614

Note: For Information, 0=non-sensitive, 1=sensitive; Profit and Revenue in billions of dollars per year at initial time of collection. Exp(coef) indicates covariate influence on baseline hazard, such that numbers <1 associate with reduced risk of event, and >1 associate with increased risk of event.

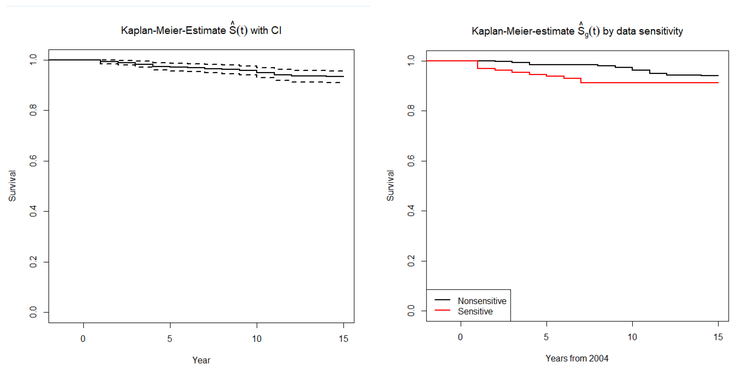

Figure 1: *Preliminary Kaplan-Meier Curves for Survival from 2004-2019 with 95% CI [Left], and for survival between organizations dealing with sensitive vs non-sensitive data [Right].*

Table 2: *Preliminary Survival Results*

Time	*n.risk*	*n.event*	*Survival*	*SE*	*95% CI*	
1	497	4	0.992	0.0040	[0.984,	1.000]
2	486	2	0.988	0.0049	[0.978,	0.998]
3	473	3	0.982	0.0061	[0.970,	0.994]
4	460	4	0.973	0.0074	[0.959,	0.988]
5	449	1	0.971	0.0077	[0.956,	0.986]
6	445	1	0.969	0.0080	[0.953,	0.984]
7	440	2	0.964	0.0085	[0.948,	0.981]
8	433	1	0.962	0.0088	[0.945,	0.979]
9	430	2	0.958	0.0093	[0.940,	0.976]
10	422	4	0.949	0.0103	[0.929,	0.969]
11	418	4	0.939	0.0111	[0.918,	0.961]
12	411	2	0.935	0.0115	[0.913,	0.958]
14	406	1	0.933	0.0117	[0.910,	0.956]

PRELIMINARY RESULTS AND DISCUSSION

Preliminary results of 31 events across the 500 companies included in initial analyses indicate an overall increase in risk of cyberincident from 2004-2019. Industry type significantly associated with increased risks of breach, with financial and healthcare organizations suffering more incidents than energy or communication companies. The sensitivity of information processed was not significantly related to the base risk of breach, though higher organizational profit was positively related.

Discussion

Initial results support the first hypothesis regarding associations between industry type and experience of cyberevent. The "industry" variable itself is nominal, so the change in hazard ratio cannot be interpreted beyond the observation that financial and healthcare institutions experienced a higher hazard ratio than did communications and energy corporations. The second hypothesis—which proposed that higher-profit organizations would be the recipient of more cyberattacks than lower-profit organizations—was supported, such that organizational profit was positively associated with an increased risk of cyberbreach. As the proposed study measures only successful cyberattacks, it cannot determine whether these organizations are targeted more in general, or simply invest less into cybersecurity efforts. However, the latter is less likely to be the case, given the immense amount of spending funneled into cybersecurity each year.

Preliminary results did not support Hypothesis 3—though insignificant, the direction of the effect ran contrary to expectations. One potential explanation is the current study's operationalization of *cybersecurity incidents* as breaches reported as successful. It may be that organizations dealing with sensitive information *do* experience more cyberattacks, but that they successfully deter the vast majority of them. Further investigations into the differences in successful vs non-successful attacks are warranted.

CONTRIBUTIONS AND LIMITATIONS

Contributions:

The initial results—albeit gleaned from a condensed dataset—highlight the insight to be provided by the proposed study. Longitudinally identifying characteristics that increase or decrease the risk of cyberincidents will emphasize to organizations the particular importance of considering their vulnerability to security-related issues—issues that are predicted to only grow in importance for organizations, their customers, and their employees.

Limitations:

Currently, limited results are available. Full conclusions cannot be drawn regarding the hypothesized effects until the complete set of data is collected and analyzed. Additionally, the difficult nature of establishing a *population at risk* that is both comprehensive and manageable has resulted in a slightly biased sample. There is limited evidence to suggest that top-performing companies are more targeted by cyberattackers than poorer-performing companies, and relying on the freely-available *Fortune* list inevitably leads to an overinclusion of top companies. Results drawn from these analyses must emphasize the context of the organizations included. However, the issue of cybersecurity remains relevant to high-performing companies.

REFERENCES

Evans, M., Maglaras, L. A., He, Y., & Janicke, H. (2016). Human behaviour as an aspect of cybersecurity assurance. *Security and Communication Networks*, *9*(17), 4667-4679.

Kassambara, A., & Kosinski, M. (2018). survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.3. https://CRAN.R-project.org/package=survminer

Keiley, M. K., & Martin, N. C. (2005). Survival Analysis in Family Research. *Journal of Family Psychology, 19*(1), 142–156.

Klein, M., with modifications by Jun Yan (2012). KMsurv: R package version 0.1-5. https://CRAN.R-project.org/package=KMsurv

Petersen, T. (2013). Analyzing Event History Data. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 2, pp. 486–516). New York, NY: Oxford University Press.

Therneau, T. (2015). _A Package for Survival Analysis in S_. version 2.38, <URL: https://CRAN.R-project.org/package=survival>.

Thonnard, O., Bilge, L., Kashyap, A., & Lee, M. (2015). Are you at risk? Profiling organizations and individuals subject to targeted attacks. In *International Conference on Financial Cryptography and Data Security* (pp. 13-31). Springer, Berlin, Heidelberg.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H., Romain, F., Henry, L., & Müller, K. (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.0.1. https://CRAN.R-project.org/package=dplyr