Hezah! Ransomware: Externalities, Cost Internalization, and Security Investment Intentionality

Eric Jardine, Eireann Leverett, and Dan Geer

Abstract (227): Cybersecurity is often subject to under consumption due to externalities, where a significant portion of the cost of a cyber attack is borne by others in the network environment. Government regulation is a common solution to externalities as a type of market failure. In this article, we propose a novel framework to illustrate that cyber-attack modalities vary in terms of their organic cost internalization quotient. Some attacks, such as data breaches, are largely cost externalizing, meaning exposure should create little incentive for subsequent security investment following exposure to this attack modality. Other attacks, such as ransomware, are largely cost internalizing and should lead to a greater investment intentionality following exposure. Using nationally representative survey data of the United States population (n=2,228) analyzed with optimal full statistical matching, we show that self-reported past exposure to a high *cost internalizing* cyber attack (i.e. ransomware) significantly increases declared security investment intentionality by almost half a unit of standard deviation (ATE = 0. 474; 95% CI = 0.292 – 0.655; p <= 0.001). Similar exposure to a high *cost externalizing* attack type (i.e. a data breach) does not result in any statistically meaningful changes to declared investment intentionality among respondents (ATE = 0.056; 95% CI = -0.035 – 0.147; p<= 0.264). The results suggest that the market for cybersecurity fails, but perhaps only in the case of some attack types and not others.

Introduction

Ransomware attacks are on the rise, with high profile intrusions during 2021 into the systems of Colonial Pipeline and JBS threatening the energy and food supply infrastructure of the United States. The increase in ransomware is partially due to changing attack modalities among malicious actors, with known data breaches within the United States declining as ransomware attack rise [1]. Many take the increase in volume and sophistication of ransomware attacks as a negative indicator of the trajectory of cybersecurity, threatening the operation of critical infrastructure such as hospitals, the bottom line of firms, the functionality of government, and the well-being of individual users.

The move toward a greater volume of ransomware attacks speaks more broadly to a persistent problem of cybersecurity under-provision. A number of variables contribute to this state of affairs [2-8], but one oft-cited reason is that security is subject to negative consumption externalities [9-14]. Individuals who are affected, for example, by a data breach often do not bear the full costs of that event, with the organization that experienced the breach absorbing the majority share of the costs. Individuals whose devices are unwittingly dragooned into a botnet lose some processing power, but can also contribute inadvertently to downtime of a major financial institution's website, potentially costing millions. In both of these examples, individuals will invest less in security than the socially optimal amount because large portions of the costs that result from that under investment are located elsewhere in the highly interdependent system of cyberspace [12]. But some attack types plausibly internalize costs to a greater degree than others. Many ransomware variants would be one such example, as the owner of an afflicted device is either forced to pay the ransom or the cost of remediation. Attack types selected by malicious actors that result in a higher proportion of internalized costs are likely to result in the opposite investment tendency to those with high cost externalization, that is, greater internalized costs should lead to more expenditure as a proportion of expected losses rather than less [14].

In this sense, we propose that the variable attack types that malicious actors can select are unequal in their effects, not just in terms of differing intrusion vectors and potential damages, but also in terms of how they will inadvertently tool the resulting incentive structure of targets through greater or lesser cost internalization effects. By extension, the growth in ransomware, while probably negative in the short term, might actually result in medium-to-long run improvements in cybersecurity outcomes because the attack vector internalizes costs to the affected parties better than many alternative attack modalities.

We test the intuitive basis of this claim using original representative survey data of the US population (n=2,228). After implementing optimal full matching so that self-reported exposure in the past year to either ransomware or a data breach is more akin to a randomized treatment in an experimental setting, we estimate the average treatment effect (ATE) of each type of security incident on a standardized (z-score) measure of omnibus cybersecurity investment intentionality (α=0.8599). Consistent with the simple argument that ransomware has a high cost-internalization quotient compared to other attack modalities, we observe that self-reported data breach exposure in the past year has no association with security investment intentions (ATE = 0.056; 95% CI = -0.035 – 0.147; p<= 0.264), while self-reported ransomware exposure in the past year

significantly increase such behavioral intentions (ATE = 0. 474; 95% CI = 0.292 – 0.655; p <= 0.001).

In the case of ransomware, density plots of weighted potential outcome means show that ransomware exposure leads to morphological changes to the distribution and not just greater mean-level cybersecurity investment intentionality. Similar to other settings where human traits predict cybersecurity behaviors [15], the gender seems to associate with morphological differences that follow from ransomware exposure. However, the changes seem to be the result of a base-rate effect and not a clear heterogenous treatment effect. In particular, men in the sample tend to report both higher baseline security investment intentionality than women absent ransomware exposure (weighted $\mu = 0.105$ vs weighted $\mu = -0.259$, respectively) and report greater mean security investment intentionality in cases of ransomware exposure (Male weighted $\mu = 0.613$ vs. Female weighted $\mu = 0.138$). Both groups increase investment intentionality following ransomware, however, by statistically similar amounts (0.508 units of standard deviation for men and 0.397 units of SD for the female subset), so the treatment is having a common effect across genders, but the differential base-rate investment levels are leading to modestly observable morphological differences post exposure to ransomware.

The rest of the paper proceeds thusly. First, we discuss the phenomenon of cybersecurity under consumption, particularly as it relates to externalities. In this section, we derive two empirically testable hypotheses. In the second section, we present our methods, data, and diagnostic results for our matching models. The third section presents the results of the optimal full matching models and the estimated ATEs for both data breach and ransomware exposure. The section also plots the weighted potential outcome means for both attack modalities and then inductively investigates a modest morphological change that ensure following ransomware exposure. The fourth section discusses the results. The fifth presents the limitations of the study. We then conclude.

1.0 - Consumption Externalities and Cost Internalization in Cybersecurity Settings

Cybersecurity is often plagued by an under-consumption problem [12-14]. The cause of this under consumption is, at least in part, a function of network effects and the interconnection of people, platforms, and devices online [10, 12]. Cybersecurity outcomes, in this sense, are a lot like private/public health outcomes surrounding illnesses such as the common cold. An individual who falls ill ends up bearing a certain amount of personal cost (e.g., cough, running nose, missed work, etc.). But, since that person has some positive chance of getting others ill as they go about their day, the total social costs of a single person's illness are usually higher than those born by the person themselves. When making production or consumption choices, individual tend to not factor in negative externalities (external costs) as meaningful in the context of a given choice—or at least heavily discount social costs relative to the costs they directly bear. When making choices about going to work versus using a sick day, for example, the person might consider primarily how they feel and think only secondarily, if at all, about the costs that might result from their making other people sick (of course the increased salience of disease spread during the Covid-19 pandemic might change this balance somewhat).

Similar cost-externalizing effects often happen in cybersecurity settings [6, 10]. If a device has become part of a botnet, for example, the owner might experience some small internalized costs (such as a reduction in processing capabilities) but this person's device might also impose costs on others by targeting web services and other users in a distributed denial of service (DDoS) attack. The individual owner of the device might invest to clean up their machine, but the choice will be largely a function of their knowledge of the problem and the estimated costs of remediation versus the costs of leaving the device infected by malware. Worry about the social costs imposed by a DDoS attack upon others is often an afterthought in an individual's decision to invest in security [13]. In some instances, such as with many industrial control systems, efforts to address security deficits through remedial steps such as software updates can even result in voided manufacturer warranties, creating an active disincentive toward investment in security. In other words, even though aggregate social costs (i.e., individual plus the costs to others) might far exceed the costs of prevention or remediation, decision makers might still choose to let vulnerabilities linger, malware lurk, or leave a patch unimplemented because the personal gains to be had from an investment in security are simply less than the direct personal costs of the expenditure.

1.1. - A Novel Framework to Gage a Priori the Cost Internalization Quotient of Attack Modalities

Negative externalities are examples of market failure. Hence, they are often addressed by government intervention. One way to address externality problems is to internalize costs to individual decision makers. Environmental regulations, for example, can impose costs on producers so that they bear a larger share of the previously externalized social costs that can accrue from air, water, or soil contamination. Government can implement cybersecurity regulation to affect similar outcomes by internalizing costs to decision makers through fiat (i.e., a non-market mechanism), leading to a higher level of security consumption. For example, Mondschein and Monda describes this internalization specifically as a regulatory ideal of GDPR: "The underlying regulatory ideal is to scale compliance to ensure that potential externalities created by the processing of personal data are internalized by the entities conducting these processing operations" [16].

In practice, regulatory rules might fail to actually result in improved security outcomes, depending on whether the restrictions govern security input choices or security output metrics [17], but forcing costs onto individual decision makers can correct for underlying externality problems in aggregate security consumption to some degree. Changes to consumer preferences can also compel greater social cost internalization, if, for example, people start buying goods known to be produced in socially desirable ways. And producers themselves might opt to internalize higher levels of social costs, depending upon the rate at which they discount social costs.

Unlike in the realm of regular economic activity where the interaction of producers, consumers, and government largely determines the degree of cost internalization that occurs, cybersecurity settings are more innately antagonistic. In this domain, the choices made by malicious actors, often quite independent of the preferences of consumers, producers, or government, can also

have direct material implications for the degree of cost internalization that occurs because of a cyberattack.

The degree of cost internalization inherent to an attack modality is largely a function of the interaction of two factors: 1) responsibility and 2) proportion. *Responsibility* refers broadly to the locus of obligation for the targeted device, information, or service. *Proportion* refers to the share of damage a potentially responsible party must content with as a result of having a given device, information, or service targeted by a malicious actor. When the degree of responsibility is low, contending with the aftermath of a security event is effectively conveyed by the nature of the attack to another responsible party. For example, if a worker at a firm has their device targeted by ransomware, their employer might be ultimately responsible for the costs associated with the attack, depending upon the company's IT use policy and the non-negligent behavior of the worker. When the proportion of the costs from a malicious event is low, a potentially responsible party needs to pay only some fraction of the total costs in remediation and clean up.

Scenarios with a low degree of target responsibility and a low proportion of total costs plausibly have a *high cost externalization* quotient, which would suggest that exposure this sort of adverse event might create few new incentives for additional security investment. Data breaches would be one such attack variant, especially for the individuals whose records are compromised as a part of a larger data breach. Given prevailing regulatory structures, for example, the choice by a malicious actor to commit a data breach against a large company, exfiltrating and then selling the financial information of the firm's clients, often imposes some direct costs on the firm but few on the individual's whose records are actually affected by the breach [18]. Instead, individuals affected by a data breach often pay little (beyond worry) and receive some admixture of event notification, free credit monitoring, and liability coverage in instances where the compromised financial information is used to commit fraud. In the language of the cost-internalization framework, individuals affected by a data breach of a firm's database have low responsibility for the attack and what costs they do bear are a small proportion of the whole. The implication here would be that an individual affected by a data breach ought to have little additional incentive to invest in security, since most of the costs that do accrue from the attack reside elsewhere in the system, largely external to the victim. This logic gives rise to H1.

> H1: Individual exposure to a data breach *should not* lead to greater levels of cybersecurity investment intentionality.

Attack types with a comparatively high degree of target responsibility and a concentrated proportion of damages often involve very *high cost internalizing* dimensions. Ransomware attacks are a preeminent example. Ransomware affects a person's device, encrypting files, and locks uses out of their system. Once done, the malicious actor behind the attack then demands that the afflicted user pay a ransom, often totaling a few hundred dollars—although the distribution of ransom payments tends to exhibit scale-free properties similar to income and can be considerably higher than the unstable average [19]. In contrast to many other potential attack types, ransomware is distinctive in that the cost of the ransom is largely carried only by the affected party – with instances wherein data is also exfiltrated (a hybrid ransomware/data breach) or the ransomware propagates from one afflicted machine to another (a worm) imposing

additional external costs elsewhere in the system. In the quintessential ransomware event where the ransom imposes a finite choice set on the owner of the device (e.g., decrypt; recover/restore files from a back up; pay; or give up the machine), the full cost of the attack is almost perfectly internalized to the affected party. In the language of the cost internalization framework, ransomware targets a victim's device and files, affecting a high degree of responsibility for remediation and subsequent prevention, and also involves a high proportion of concentrated costs, assuming the ransomware does not propagate to widely from an initially affected device to others. The high rate of cost internalization associated with ransomware attacks should produce a greater incentive to invest for those individuals affected by this cyberattack modality. This reasoning gives rise to H2.

> H2: Prior individual exposure to a ransomware attack *should* lead to greater levels of cybersecurity investment intentionality.

2.0 – Methods, Data, and Diagnostics

2.1 – Methods

In this study, we use optimal full matching implemented in R using the MatchIt package to estimate the average treatment effect (ATE) of both self-reported ransomware and self-reported data breach exposure within the past year on an individual's current cybersecurity investment intentionality in the present [20-22]. With observational data, exposure to a cybersecurity incident of either type is likely non-random, as people differ widely in terms of the online behaviors, technical competence, and so forth. As a result, simple cross tabulations or mean comparisons between perceived prior exposure to an adverse event and current cybersecurity investment intentionality are prone to spuriousness and omitted variable bias. Matching methods are a class of estimation tools, based upon the Rubin potential outcome causal framework [23, 24], developed to estimate "treatment effects from nonexperimental or observational data" [25].

Optimal full matching predicts propensity scores based upon an initial generalized linear model and then assigns units to subclasses based upon these values. The method then uses subclass membership to construct statistical weights that ideally produce a balanced sample between the 'treatment' and 'control' groups (see Section 2.3 for diagnostics). While propensity score matching has limitations [26], optimal full matching is less sensitive to the functional form of the propensity score model, because of its use of subclass assignment to create the resultant statistical weights [27]. Optimal full matching also minimizes data loss by using most (if not all) of the available observations and allows for the estimation of an average treatment effect (ATE).

2.2 – Data

2.2.1 – Sample and Controls

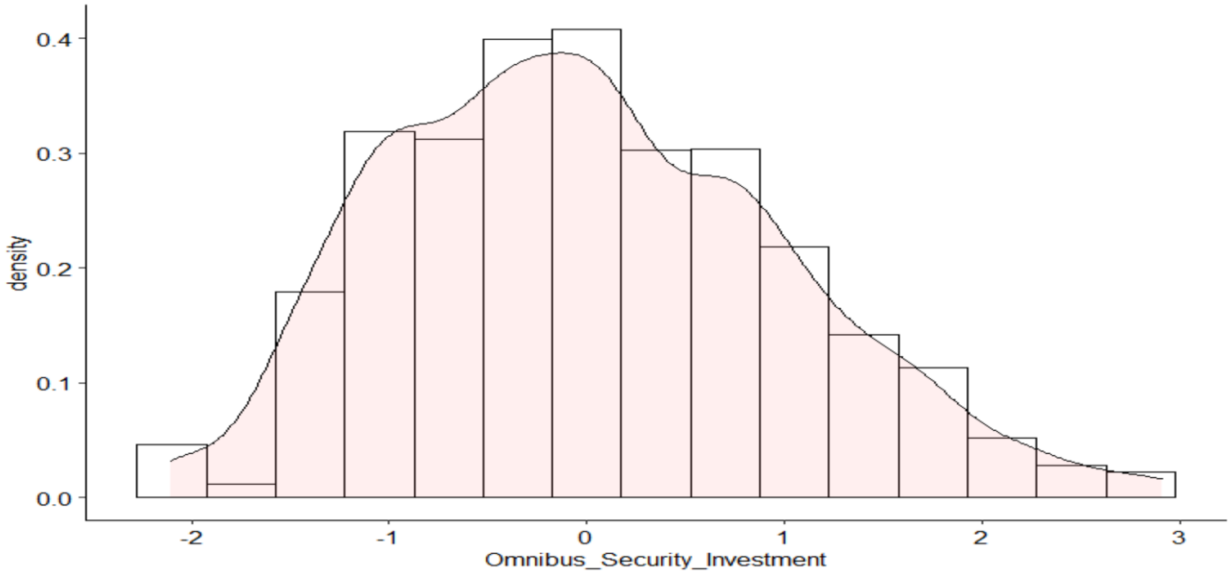
The data for this study consist of an initial representative sample of 2,028 US residents. We collected the data using the Qualtrics online panel between 8/3/2020 and 10/11/2020 and quota sampled on gender, age, household income, race, and region of the country in order to match the demographics of the general US population. The data also include a number of substantive questions that we transformed into mean unstandardized indexes to capture respondent 1) the

familiarity with types of cybersecurity attacks (α = 0.954), 2) attitudes toward various data types (α = 0.965), 3) expressed levels of computer competence (α = 0.923), and 4) online self-efficacy scores (α = 0.946). The cyberattack familiarity and data attitude measures were developed for this study. The online self-efficacy (α =) and computer competence scores (α =) were taken in modified form from previously validated scales [28, 29].

2.2.2 – The Outcome Measure

The outcome measure of interest in this study is cybersecurity investment intentionality. Crucially, since we do not measure actual security behaviors at a device level, we are only able to approximate a respondent's level of behavioral intention, which is a widely used outcome measure across numerous studies [15, 30-32]. We asked six discrete cybersecurity-related questions involving various forms of investment, including monetary expenditure, time spent on security, and four separate indicators of cognitive load related to common cybersecurity tasks. For the analysis, these questions were aggregated into a mean unstandardized item (α = 0.869). As the original scale for this measure is arbitrary, we transformed it into a z-score. Z-scores place the mean of the sample at zero and convert the units into more inherently meaningful units of standard deviation. Figure 1 plots the distribution of the z-score for the cybersecurity investment intentionality measure.

Figure 1. Cybersecurity Investment Intentionality
Z-Score Measure



2.2.3 – The Treatment Variables

Lastly, the survey also includes two separate self-reported measures of previous exposure to an adverse cybersecurity event. One question asked respondents to self-report exposure to ransomware in the past year. Another similarly asked for a self-reported indication whether a person's personal records had been involved in a data breach. In both cases, we gave respondents an option to select "I don't know." We dropped these cases from the analysis, with 277 people selecting this option for ransomware and 239 doing so for the data breach question. Additionally,

since we are interested in the isolated treatment effect of exposure to either ransomware or a data breach, we parsed the data into those who were afflicted by one of the attack modalities but not the other. This procedure allows use to isolate for individuals exposed to ransomware but not a data breach and those afflicted by a data breach but not ransomware. Absent this procedure, some proportion of the sample would be, in effect, doubly exposed to the effects of two analytically separate attack modalities, contaminating any subsequent estimation of effects. This procedure leads to two effective data subsets. One includes all respondents who indicated unique exposure to ransomware in the past year and all those respondents who reported no exposure to any attack modality (n= 1,127). The second includes all respondent who indicated they were unaffected by any attack type and those indicating unique exposure to a data breach (n= 1,528).

Table 1 summarizes the descriptive parameters of the data.

Table 1. Descriptive Statistics Summary (Unweighted Data)

	Data Breach Exposure (N=494)	No Exposure (N=1034)	Ransomware Exposure (N=93)	Overall (N=2013)
Omnibus Security Investment (z-score)				
Mean (SD)	0.0396 (0.936)	-0.0858 (1.00)	0.571 (0.942)	-0.000 (1.00)
Median [Min, Max]	-0.138 [-2.11, 2.91]	-0.138 [-2.11, 2.91]	0.578 [-1.57, 2.73]	-0.138 [-2.11, 2.91]
Missing	1 (0.2%)	23 (2.2%)	0 (0%)	39 (1.9%)
Unique Ransomware Exposure				
Mean (SD)	NA (NA)	0 (0)	1.00 (0)	0.0745 (0.263)
Median [Min, Max]	NA [NA, NA]	0 [0, 0]	1.00 [1.00, 1.00]	0 [0, 1.00]
Missing	494 (100%)	0 (0%)	0 (0%)	765 (38.0%)
Unique Data breach Exposure				
Mean (SD)	1.00 (0)	0 (0)	NA (NA)	0.303 (0.460)
Median [Min, Max]	1.00 [1.00, 1.00]	0 [0, 0]	NA [NA, NA]	0 [0, 1.00]
Missing	0 (0%)	0 (0%)	93 (100%)	385 (19.1%)
Gender: Male				
Female	269 (54.5%)	497 (48.1%)	32 (34.4%)	1014 (50.4%)
Male	225 (45.5%)	537 (51.9%)	61 (65.6%)	999 (49.6%)

Table 1. Descriptive Statistics Summary (Unweighted Data)

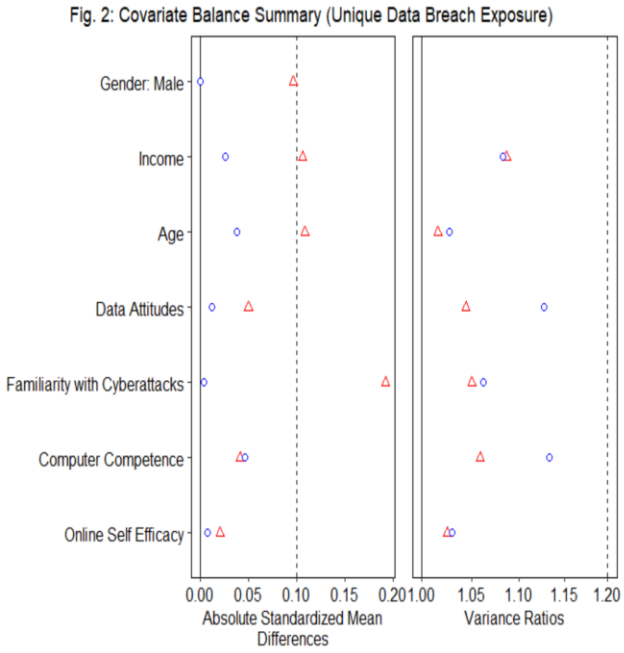
	Data Breach Exposure (N=494)	No Exposure (N=1034)	Ransomware Exposure (N=93)	Overall (N=2013)
Income				
Mean (SD)	2.94 (0.954)	2.87 (0.990)	2.89 (0.938)	2.83 (1.00)
Median [Min, Max]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]	3.00 [1.00, 4.00]
Age				
Mean (SD)	3.83 (1.66)	3.66 (1.67)	2.89 (1.35)	3.59 (1.66)
Median [Min, Max]	4.00 [1.00, 6.00]	3.00 [1.00, 6.00]	3.00 [1.00, 6.00]	3.00 [1.00, 6.00]
Data Attitudes				
Mean (SD)	34.5 (21.9)	33.6 (22.8)	41.2 (26.5)	35.2 (22.8)
Median [Min, Max]	30.7 [0, 99.7]	28.7 [0, 100]	38.3 [0.0714, 100]	30.7 [0, 100]
Missing	0 (0%)	2 (0.2%)	0 (0%)	3 (0.1%)
Familiarity with Cyberattacks				
Mean (SD)	2.68 (0.967)	2.52 (0.998)	3.04 (0.999)	2.57 (1.01)
Median [Min, Max]	2.67 [1.00, 5.00]	2.42 [1.00, 5.00]	3.00 [1.17, 5.00]	2.50 [1.00, 5.00]
Computer Competence				
Mean (SD)	4.67 (0.634)	4.66 (0.607)	4.47 (0.757)	4.59 (0.677)
Median [Min, Max]	5.00 [1.00, 5.00]	5.00 [1.00, 5.00]	4.88 [1.00, 5.00]	5.00 [1.00, 5.00]
Online Self-Efficacy				
Mean (SD)	3.54 (0.874)	3.59 (0.865)	3.73 (0.788)	3.53 (0.876)
Median [Min, Max]	3.67 [1.00, 5.00]	3.75 [1.00, 5.00]	3.75 [1.25, 5.00]	3.67 [1.00, 5.00]

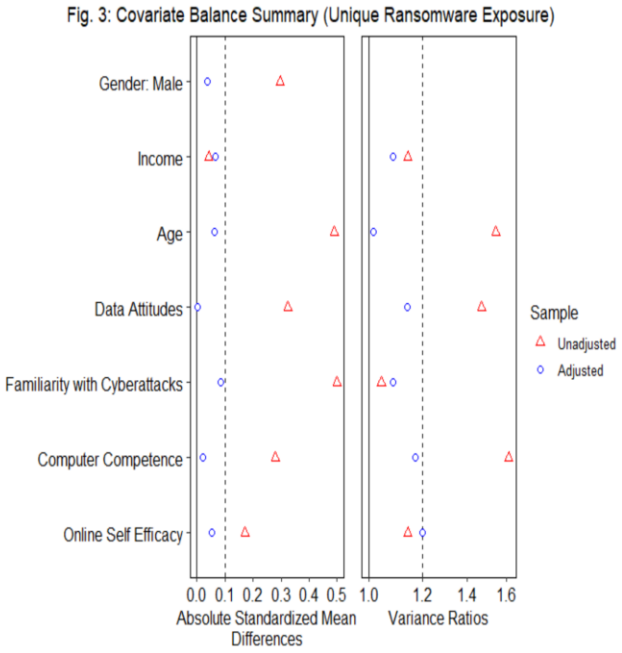
2.3 – The Matching Models and Diagnostic Balance Summaries

The optimal matching models were implement using MatchIt [22]. The initial matching model used to predict subclass membership included all the demographics and dispositional variables contained in Table 1. The caliper common support threshold was manually tooled in each

instance to produce an ideal sample balance that minimized data loss. In each case, the caliper remained at or below the suggested 0.2 units of standard deviation of the distance measure [25, 33, 34]. In the data breach subset, the caliper was set to 0.2 and the sample was exact matched on respondent Gender in order to affect good sample balance. This process led to a loss of 12 observations in the untreated group and 1 in the treated group. Within the ransomware matching model, the caliper was manually tooled downward from 0.2 to 0.085, leading to a loss of 22 observations from the control group and 3 from the treatment group.

Sample balance was assessed use the Cobalt package in R [35]. Figure 2 presents the absolute standardized mean differences and the variance ratios for the data breach exposure model and Figure 3 does the same for the ransomware subsample. In each instance, absolute standardized mean differences were less than 1 in the adjusted (i.e. matched) sample, as is recommended. Likewise, variance ratios in each case were within 0.2 of the ideal of 1, suggesting an acceptable level of balance in the matched samples. In total, Figures 2-3 indicates that the matching procedure resulted in an effective balance of covariates between treatment and control groups, which should allow for an unbiased estimation of the ATE for each type of adverse cybersecurity event.





3.0 – Results

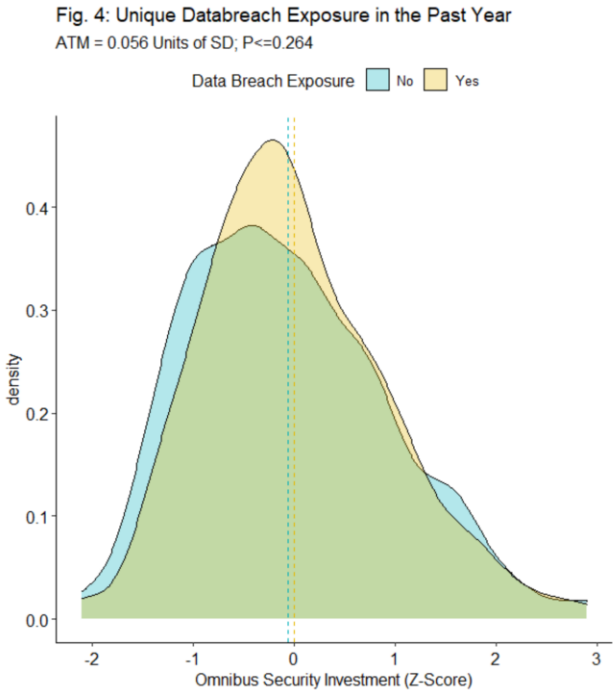
Table 2 summarizes the ATE results of the two matching models. In support of H1, unique self-report exposure to a data breach in the last year is not associated with changed cybersecurity investment intentionality (ATE = 0.056; 95% CI = -0.035 – 0.147; p<= 0. 264). In contrast, and supporting H2, exposure to ransomware in the past year is associated with a statistically significant increase in respondent cybersecurity investment intentionality (ATE = 0.474; 95% CI = 0.293 – 0.656; p <= 0.000). In substantive terms, ransomware exposure increases cybersecurity

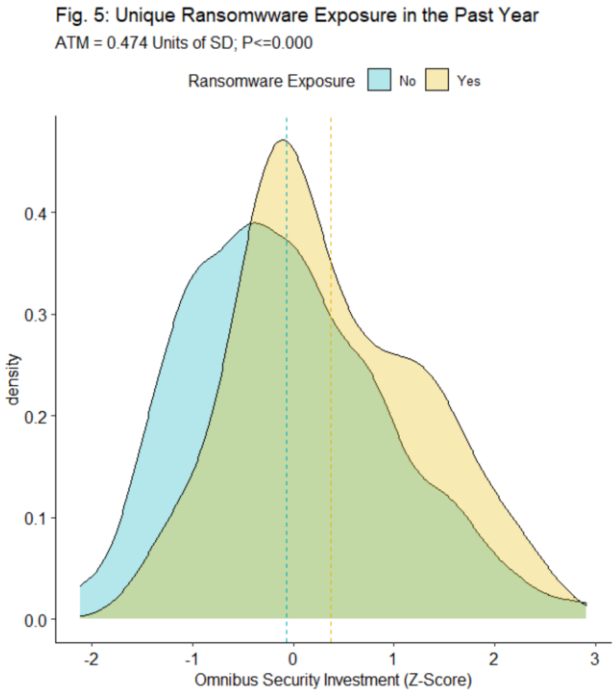
investment intentionality by between 1/3 and 2/3 of a unit of standard deviation on the z-score scale.

Table 2. Self-reported Security Event Exposure and Cybersecurity Investment Intentionality (z-score)

Estimated Average Treatment Effects (ATE)		
Yes vs No	**ATE**	**95% Confidence Interval**
Unique Data Breach Exposure	0.056 (0.050)	-0.035 – 0.147
Unique Ransomware Exposure	0.474*** (0.111)	0.292 – 0.655

Note: Clustered Robust Standard Error in Parentheses.
*** = p<0.001; ** = p <0.01; * = p <0.05



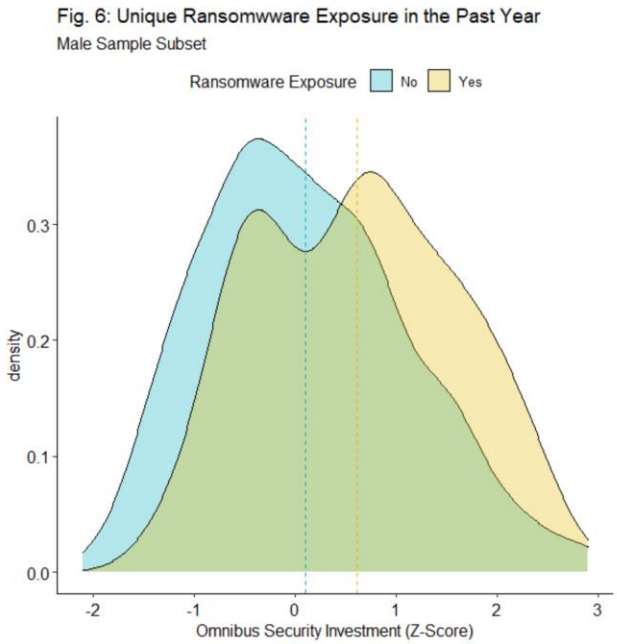


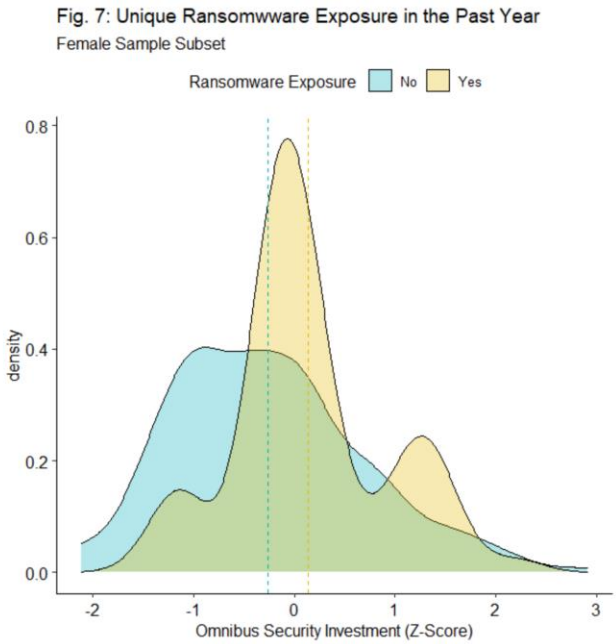
Figures 5-6 plot the weighted potential outcome means for both the data breach and ransomware samples. In line with the output of the matching model, data breach exposure has little visually discernable effect on the distribution of effort across the omnibus cybersecurity investment intentionality measure and the weighted means for the two samples are almost identical (μ = -0.058 for the unexposed sample and μ = 0.003 for the exposed group). In contrast, as shown in Figure 6, self-reported ransomware exposure shifts mean investment intentionality upward in a sizeable way (from μ = -0.069 to μ = 0.376).

Interestingly, ransomware exposure also seems to lead to modest morphological changes to the distribution of cases across the omnibus security investment intentionality scale. In particular,

ransomware exposure tends to lead to a greater density of security investment intentionality at two separate points on the scale (i.e. at both 0 and at 1 or more units of standard deviation on the z-score). These results support the potential operation of a clear attack type/cost internalization dynamic, but also raise an unanticipated question of potentially heterogeneous responses to a common treatment. Inductively investigating the source of these potential differences further suggests that human traits could drive the differing morphology of the ransomware exposure distribution [15, 32].

As shown in Figures 5-6 and consistent with other investigations of human factors in the security investment [15], gender may play a role in the observed changes to the shape of the distribution of security investment intentionality following ransomware exposure. Filtering the ransomware exposure sample by respondent gender and re-plotting the weighted potential outcome means shows that there are two separate shifts occurring in the distribution of security investment intentionality as a result of self-reported exposure to ransomware in the past year. The results of a two-way ANOVA using the statistical weights derived from the matching model suggests that both ransomware exposure ($F = 17.086$; $P <= 0.000$) and gender ($F= 42.174$; $P <= 0.000$) associate with security investment intentionality levels. However, the interaction term (i.e. ransomware exposure by gender) is not significant ($F= 0.263$; $p<=0.608$).





The non-significant interaction term in the two-way ANOVA suggests that men and women respond statistically similar to the treatment (hence there is no heterogenous treatment effect *per se*). Instead, the observable morphological differences at play are likely an artifact of base-rate effects. Men in the unexposed sample tend to have a higher overall level of security investment intentionality at the mean than women (weighted $\mu = 0.105$ vs weighted $\mu = -0.259$, respectively). Additionally, both men and women increase their security investment intentionality if exposed to ransomware (Male weighted $\mu = 0.613$ vs. Female weighted $\mu =$

0.138). The clustering of respondents by gender across the cybersecurity investment intentionality scale, therefore, is a function of start point, not response type or severity.

4.0 – Discussion

The result of the optimal full matching models show that self-reported exposure to some types of malicious events cause a significant increase in security investment intentionality while others do not. In particular, exposure to a data breach—an attack modality that used to be a prevailing mode of malicious actor activity and that has, for individually affected parties, mostly cost externalizing effects—tends to lead to little discernable change to how seriously individuals take cybersecurity. In contrast, ransomware exposure tends to increase significantly the security investment intentionality of affected parties. Interestingly, gender potentially predicts starting location on the cybersecurity investment scale and so contributes to modest morphological changes to the density of activity across the scale, as seen in other studies [15]. These results have a number of interesting implications.

The broadest point of note is that not all attack modalities that are available to malicious actors have the same cost in/externalizing effects. This simple notion has profound implications for malicious actors. Potentially, and over the long run, organic and voluntarily undertaken malicious activity might inadvertently generate a familiar ‘tragedy of the commons’ dynamic [36]. To the extent that individually preferred attack modalities (on the part of the attacker) also have a higher cost internalization quotient than available alternatives, independent choices made by malicious actors could gradually reduce the pool of vulnerable points of attack as investment rises due to the correction of a market failure through greater cost internalization [6]. The familiar analogy would be everyone choosing to stand at a hockey game (analogous to selecting the individually preferred, high cost internalization attack modality) because it is the optimal choice for one party, only to discover that when everyone does it, all become worse off.

The findings also raise a little-considered additional variable for regulators who might want to address perceived deficiencies in cybersecurity by altering the cost structures for responsible and affected parties through regulation [16]. It is plausible that greater perceived cybersecurity risk correlates positively with a larger “policy window” through which “policy entrepreneurs” with plans for how to regulate cybersecurity might choose to leap [37]. Our findings suggest, however, that worse cybersecurity outcomes and forecasted trajectories do not necessarily warrant a greater degree of government (regulatory) involvement. If, instead, emerging attack modalities *de jure* have both higher damage output and greater cost internalizing effects for affected parties, then it becomes a more open question whether the ecosystem will adapt organically to reduce cybersecurity risk to socially optimal levels over the medium-to-long run or whether government needs to intercede. Since the effectiveness of regulation for improving cybersecurity outcomes is not guaranteed [17] and government programs can often result in unintended negative effects [38], determining if the system will efficiently self-correct (due to high cost internalization attributable to certain attack types) or if markets for security will continue to fail is a useful first-order question for regulators.

A similar logic might prevail for cyber insurance as well. Premium pricing is at least partially a function of demonstrated security controls, client asset values, and ecosystem trends in malicious events [39, 40]. It can also be more simply stated as a function of perceived future risk. Over time, pricing of premiums will change for afflicted parties as a function of exposure to cybersecurity incidents. In an automobile insurance setting, an at-fault accident will likely lead to an increased premium, at least in part because it can signal greater risk-accepting behaviors by the driver. Similar premium increases could plausibly follow cybersecurity events. However, the results here nuance when and to what degree pricing schedules should change following malicious events. Core to the point is the extent to which exposure to a cybersecurity event could be taken as a future risk signal that would warrant higher premium pricing versus the exact opposite. In the case of high cost-externalizing cyberattacks (e.g., data breaches affecting individuals in our sample), past behaviors with regards to cybersecurity should predict future behaviors well (i.e., investment levels do not change). Exposure to a cyberattack that has a high degree of cost internalization, on the other hand, likely predicts less future risk and so might even warrant a premium reduction following an adverse event—unless, of course, the motive behind a premium increase is to recoup costs and not use past events as a predictive variable of future risk.

The base-rate investment levels in the weighted sample are quite strongly associated with the gender of the respondents, with women declaring lower levels of cybersecurity investment intentionality on average than men. These results are similar to other studies, but it is doubtful that gender as such matters. Instead, as detailed in other studies [32], gender is likely affecting other factors such as exposure to STEM fields, perceived online self-efficacy, and self-reported computer competence levels in ways that reduce the mean base-rate security investment intentionality levels of women versus men. An interesting implication of this finding is that greater inclusion of women within STEM-fields could actually work to improve cybersecurity to the extent that such exposure increases perceptions of self-efficacy and leads to greater investment intentionality. Framed conversely, marginalization of people from STEM fields—again, to the extent that exposure breeds self-efficacy and self-efficacy predicts security investment intentionality—might be construed as a security issue and not just a matter of fairness.

Finally, the validation of hypotheses about cost internationalization and security investment behaviors that we derived from the cost internalization framework suggests some utility to the simple bi-variate framework. Two points are particularly relevant in this case. First, the model provides variables that are independent of known target reactions to any given attack, which might allow for a degree of prediction as possible and preferred attack modalities continue to evolve over time. The broadest level prediction would be that new attack modalities that target those with a high degrees of responsibility for the targeted device, information, or service and that have high concentrations of damage would be highly cost internalizing attacks and should lead to greater subsequent security investment. Low scores on these values would entail a low cost internalizing attack modality and should be unassociated with subsequent increases in investment. In other words, the framework might provide a couple of leading indicators of investment behaviors by future targets of cyber-attacks.

Perhaps the broadest implication for the demonstrable utility of the cost internalization framework is that externalities, so often considered a major source of security under consumption [6, 9, 10, 12-14], might instead by a source of market failure for security only in the case of some attack modalities and not others. Framed another way, there might be market failure in some zones of security but efficient markets for security in others [41]. Extrapolating from the framework to other attack types, the market for CDN provided DDoS protection would be an efficient market, characterized by comparatively little evidence of market failure and security under consumption. While latent capacity for DDoS and rDDoS attacks remains high globally [42], demonstrable disruptions to service delivery via web services seem, on the face of it at least, to be fairly negligible today. While the ease and feasibility of remediation and prevention methods varies by attack modality, the results suggest that ransomware might become, in the future, a lot like DDoS attacks (largely a well-counter nuisance) as the market for security effectively mobilizes due to its high cost-internalizing quotient.

5.0 – Limitations

The study presents an estimation of the average treatment effect of two different cybersecurity attack modalities. Several limitations to the study exist, however.

First, we measured behavioral intentions and self-reported exposure to a cybersecurity event in place of real security practices or forensically confirmed malware infection or independently corroborated data breach exposure. These are indeed limits. Yet a number of studies have shown that security behavioral intentions can correlate with practice, making this a useful measure. Likewise, security events—especially ransomware and data breach exposure—are highly salient for respondents, so the reporting error for this class of phenomenon is plausibly lower than might be the case.

Second, while optimal full matching is doubly robust to misspecifications in either the matching or effect estimation, the model might still be prone to omitted variable bias. Since the model works by matching treated and untreated respondents on important covariates, then the exclusion of an important covariate might bias the effect estimations. We controlled for a number of demographic (i.e., human []) traits and also important distortional characteristics such as perceived computer competency and online self-efficacy. It is always possible that we omitted an important control from the models, but we strongly suspect that the results are robust.

Third, we investigated the potential operation of cost in/externalization of two separate cyberattack modalities at the individual level. The results as they pertain to individual respondents are clear, but it is less obvious that what applies at the individual level would scale to, say, firms or government, each of which might be nested in organizational and regulatory webs in which the individuals in our sample are not (though we believe this to be a limitation of the data and not of the methodology or analytical framework, and it would certainly be possible to survey boards in a similar way). Nevertheless, we suspect that our results are likely generalizable across individuals and potentially across scales, at least to some degree. We showed that the incentives generated by a given cyber-attack are endogenous to the selected

attack modality and cost in/externalization is a general enough effect that it likely applies to individuals, enterprises and non-profits.

Future research measuring known attacks, tracking de facto security behaviors, and doing so across scales would help to address these limitations.

6.0 – Conclusions

Some cyber-attack modalities internalize costs to affected parties better than others. The implication is that some individuals (and potentially larger institutions) might come out of an adverse cybersecurity incident with a greater or lesser incentive to investment more in security, depending upon the cost in/externalizing nature of the attack itself.

References

[1] M. J. Schwartz, "Reported US Data Breaches Declined by 19% in 2020," in *Bank Info Security* vol. 5/5/2021, ed, 2021.

[2] J. Jang-Jaccard and S. Nepal, "A survey of emerging threats in cybersecurity," *Journal of Computer and System Sciences,* vol. 80, no. 5, pp. 973-993, 2014/08/01/ 2014.

[3] D. Geer, E. Jardine, and E. Leverett, "On market concentration and cybersecurity risk," *Journal of Cyber Policy,* vol. 5, no. 1, pp. 9-29, 2020/01/02 2020.

[4] R. von Solms and J. van Niekerk, "From information security to cyber security," *Computers & Security,* vol. 38, pp. 97-102, 2013/10/01/ 2013.

[5] E. Jardine, "A Continuum of Internet-Based Crime: How the Effectiveness of Cybersecurity Policies Varies Across Cybercrime Types," in *Research Handbook on Digital Transformations*, F. X. Olleros and M. Zhegu, Eds. Cheltenham, UK: Edward Elgar, 2016, pp. 421-444.

[6] R. Anderson and T. Moore, "The Economics of Information Security," *Science,* vol. 314, no. 5799, pp. 610-613, 2006.

[7] D. G. Arce, "Cybersecurity and platform competition in the cloud," *Computers & Security,* vol. 93, p. 101774, 2020/06/01/ 2020.

[8] D. G. Arce, "Malware and market share," *Journal of Cybersecurity,* vol. 4, no. 1, 2018.

[9] J. M. Bauer and M. J. G. van Eeten, "Cybersecurity: Stakeholder incentives, externalities, and policy options," *Telecommunications Policy,* vol. 33, no. 10, pp. 706-719, 2009/11/01/ 2009.

[10] H. Varian, "System reliability and free riding," in *Economics of information security*: Springer, 2004, pp. 1-15.

[11] L. A. Gordon, M. P. Loeb, and W. Lucyshyn, "Sharing information on computer systems security: An economic analysis," *Journal of Accounting and Public Policy,* vol. 22, no. 6, pp. 461-485, 2003.

[12] H. Kunreuther and G. Heal, "Interdependent security," *Journal of risk and uncertainty,* vol. 26, no. 2, pp. 231-249, 2003.

[13] M. Lelarge, "Coordination in network security games: a monotone comparative statics approach," *IEEE Journal on Selected Areas in Communications,* vol. 30, no. 11, pp. 2210-2219, 2012.
[14] L. A. Gordon, M. P. Loeb, W. Lucyshyn, and L. Zhou, "Externalities and the magnitude of cyber security underinvestment by private sector firms: a modification of the Gordon-Loeb model," *Journal of Information Security,* vol. 6, no. 01, p. 24, 2014.
[15] M. Gratian, S. Bandi, M. Cukier, J. Dykstra, and A. Ginther, "Correlating human traits and cyber security behavior intentions," *computers & security,* vol. 73, pp. 345-358, 2018.
[16] C. F. Mondschein and C. Monda, "The EU's General Data Protection Regulation (GDPR) in a Research Context," in *Fundamentals of Clinical Data Science*, P. Kubben, M. Dumontier, and A. Dekker, Eds. Cham: Springer International Publishing, 2019, pp. 55-71.
[17] L. A. Gordon, M. P. Loeb, W. Lucyshyn, and L. Zhou, "Increasing cybersecurity investments in private sector firms," *Journal of Cybersecurity,* vol. 1, no. 1, pp. 3-17, 2015.
[18] S. Romanosky, "Examining the costs and causes of cyber incidents," *Journal of Cybersecurity,* vol. 2, no. 2, pp. 121-135, 2016.
[19] E. Leverett, E. Jardine, E. Burns, A. Gangwal, and D. Geer, "Averages Don't Characterise the Heavy Tails of Ransoms," in *eCrime 2020 Symposium on Electronic Crime Research*, Zoom, 2020: APWG.
[20] B. B. Hansen and S. O. Klopfer, "Optimal full matching and related designs via network flows," *Journal of computational and Graphical Statistics,* vol. 15, no. 3, pp. 609-627, 2006.
[21] E. A. Stuart and K. M. Green, "Using full matching to estimate causal effects in nonexperimental studies: examining the relationship between adolescent marijuana use and adult outcomes," *Developmental psychology,* vol. 44, no. 2, p. 395, 2008.
[22] D. Ho, K. Imai, G. King, and E. A. Stuart, "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference," *2011,* vol. 42, no. 8, p. 28, 2011-06-14 2011.
[23] D. B. Rubin, *Matched sampling for causal effects*. Cambridge University Press, 2006.
[24] D. B. Rubin, "Direct and indirect causal effects via potential outcomes," *Scandinavian Journal of Statistics,* vol. 31, no. 2, pp. 161-170, 2004.
[25] S. Guo and M. W. Fraser, *Propensity score analysis : statistical methods and applications*. 2015.
[26] G. King and R. A. Nielsen, "Why propensity scores should not be used for matching," 2019.
[27] P. C. Austin and E. A. Stuart, "The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes," *Statistical methods in medical research,* vol. 26, no. 4, pp. 1654-1670, 2017.
[28] M. C. Howard, "Creation of a computer self-efficacy measure: analysis of internal consistency, psychometric properties, and validity," *Cyberpsychology, behavior, and social networking,* vol. 17, no. 10, pp. 677-681, 2014.
[29] A. J. Van Deursen, E. J. Helsper, and R. Eynon, "Measuring digital skills. From digital skills to tangible outcomes project report," 2014.

[30] A. T. Shappie, C. A. Dawson, and S. M. Debb, "Personality as a predictor of cybersecurity behavior," *Psychology of Popular Media Culture*, 2019.
[31] Y. John, "Generational Differences: A Quantitative Study in Employees Information Security Knowledge, Attitudes, and Behavioral Intentions," Capella University, 2021.
[32] M. Anwar, W. He, I. Ash, X. Yuan, L. Li, and L. Xu, "Gender difference and employees' cybersecurity behaviors," *Computers in Human Behavior*, vol. 69, pp. 437-443, 2017/04/01/ 2017.
[33] M. Lunt, "Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching," (in eng), *American journal of epidemiology*, vol. 179, no. 2, pp. 226-235, 2014.
[34] P. R. Rosenbaum and D. B. Rubin, "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score," *The American Statistician*, vol. 39, no. 1, pp. 33-38, 1985.
[35] N. Greifer, "cobalt: Covariate Balance Tables and Plots. R package version 4.3.1.," ed, 2021.
[36] G. Hardin, "The tragedy of the commons," *Journal of Natural Resources Policy Research*, vol. 1, no. 3, pp. 243-253, 2009.
[37] J. W. Kingdon and E. Stano, *Agendas, alternatives, and public policies*. Little, Brown Boston, 1984.
[38] S. M. Gillon, *That's not what we meant to do: Reform and its unintended consequences in twentieth-century America*. WW Norton, 2000.
[39] D. Woods, T. Moore, and A. Simpson, "The County Fair Cyber Loss Distribution: Drawing Inferences from Insurance Prices," in *Workshop on the Economics of Information Security*, 2019.
[40] D. Woods and T. Moore, "Does insurance have a future in governing cybersecurity?," *IEEE Security and Privacy Magazine*, 2019.
[41] V. Garg, "Covenants Without the Sword: Market Incentives for Cybersecurity Investment," in *The 49th Annual Research Conference on Communications, Information, and Internet Policy*, Washington, DC. , 2021: TPRC.
[42] E. Leverett and A. Kaplan, "Towards estimating the untapped potential: a global malicious DDoS mean capacity estimate," *Journal of Cyber Policy*, vol. 2, no. 2, pp. 195-208, 2017/05/04 2017.

Acknowledgments

Funding for this project was provided by a Comcast Innovation grant. Grant number: 2019-145 || Incrementally Tailoring a Better Cyber Risk Score