Age and gender as independent risk factors for malware victimisation

Fanny Lalonde Lévesque
École Polytechnique de Montréal
www.polymtl.ca
fanny.lalonde-levesque@polymtl.ca

José M. Fernandez
École Polytechnique de Montréal
www.polymtl.ca
jose.fernandez@polymtl.ca

Dennis Batchelder
AppEsteem Corporation
www.appesteem.com
denbatch@appesteem.com

This paper presents the results of an empirical study we designed to investigate the independent effect of age and gender as potential risk factors for malware victimisation. Using data collected from Microsoft's Windows Defender on a sample of three million devices running Windows 10, we found that both age and gender are contributing factors for malware victimisation. Men, and young men in particular, were more likely to encounter malware than women, and younger users were more at risk of encountering malware than their older counterparts. However, our findings suggest that the effect of age and gender is not constant across different types of malware. We also discuss potential causes and implications of these age and gender differences in malware victimisation.

Human factor, Computer security, Malware, Field study

1. INTRODUCTION

Human factors (e.g. demographics, characteristics, behaviour) are known to play a significant role in information security. While the literature on user behaviour and cyberattacks is very extensive, there is significantly less work that focus on user demographics. So far many studies have investigated how user demographics relate to cyberattacks; only a few studies have focused on the risk of malware victimisation (Ngo and Paternoster 2011; Bossler and Holt 2009; Yen et al. 2014; Lalonde Lévesque et al. 2013). Their findings essentially suggest that age and gender could be contributing factors in the success (or failure) of malware infections.

On the one hand, cybercriminals are increasingly employing varied monetization schemes that target specific regions of the world and categories of users, for example with targeted banking fraud and ransomware attacks. It is conceivable that cybercriminals may be targeting particular groups to maximize success and revenues, in a similar fashion as Internet publicity campaigns are now targeting specific groups using profiling information based on computer usage behaviour. On the other hand, the psychological traits and level of awareness of users can affect their decision making in the context of computer usage, hence affecting both the likelihood of exposure and the effectiveness of the infection mechanisms. In the first case, there is sufficient circumstancial evidence from the analysis of malware and cybercrime campaigns to believe that users may be targeted according to age and gender. In the second case, previous research has shown that computer usage behaviour varies significantly with age and gender. For these reasons, it is reasonable to hypothesize that age and gender could be actual contributing factors related to the risk of malware victimisation

A better understanding of gender and age differences in the risk of malware victimisation could enable researchers, practitioners and policy makers to better design gender and age-differentiated interventions in cybersecurity. However, rigorous evidence of gender and age differences in malware victimisation are still relatively scarce. Consequently, there is a need to conduct empirical studies of actual malware victimisation based on large and representative sample of computer users. It is therefore essential to try to empirically confirm that age and gender 1) are indeed risk factors, and that 2) they are involved in the causal pathway leading to malware victimisation.

This paper concentrates on the first question, as a precursor for eventually addressing the second one. In particular, we present a large scale empirical study specifically designed to evaluate age and gender as independent risk factors for malware victimisation. Inspired by the epidemiology approach, we design a field study based on a large sample of millions of Windows 10 devices protected by Microsoft's Windows Defender. We use stratification and regression to investigate the effect of age and gender as risk factors of malware victimisation. Our results contribute to existing literature

by shedding light on age groups and gender differences in malware victimisation and how their effect vary depending on the type of malware (e.g. ransomware, adware, infostealer).

The remainder of the paper is organised as follows. In Section 2 we review previous work on age and gender differences in malware victimisation. Section 3 describes the study in terms of design, data collection and analysis. In Section 4 we present our results. We discuss our observations in Section 5 and limitations of our study in Section 6. We conclude and discuss potential implications of our findings in Section 7.

2. PREVIOUS STUDIES

There have not been, to the best of our knowledge, other empirical studies specifically designed to evaluate age and gender differences in the risk of malware victimisation. In this section, we present a review of past work that studied how age and gender correlate with malware victimisation; though it was not their primary interest. We also highlight a few studies that investigated the effect of users' demographics on the risk of other types of computer threats (e.g. phishing, spam, identity theft).

2.1. Demographics and malware victimisation

Some researchers have investigated the effect of users' demographics on malware victimisation by adopting subjective research methods. Mostly based on surveys, interviews, and observations, these methods seek to understand why and how users interact with computer systems. For example, Ngo and Paternoster (2011) applied the general theory of crime and lifestyle/routine activities framework to assess the effects of individual and situational factors on seven types of cybercrime victimization, including computer virus infection. They conducted a self-assessment survey using a sample of 295 college students and correlated users' demographics and characteristics (gender, age, race, marital status) with self-reported cybercrime victimization. The authors deduced that the effect of gender was not significant, while age was identified as a significant predictor for self-reported computer virus infection, with older respondents being less likely to get infected. In another study, Bossler and Holt (2009) applied a routine activities framework to explore the causes and correlates of self-reported data loss from malware infection. The authors administered a survey on a sample of 788 college students and investigated, among others, the effect of gender, age, race, and employment status. They found that being a female increases the odds of malware victimization by 1.827 times compared to male. However, age was not identified as a significant predictor of self-reported malware victimization.

Other studies investigated the effect of age and gender as potential risk factors of malware victimisation based on objective research methods. In comparison with the studies cited above, they are based on real-life data, and not on self-reported malware victimisation and users' behaviour. Lalonde *et al.* (Lalonde Lévesque et al. 2013, 2014) did a 4-month field study of 50 users based on the clinical trial approach used in medecine to assess the impact of human and technological factors on the risk of malware exposure. The authors found no significant differences based on gender or age. Also inspired by the epidemiology approach, Yen et al. (2014) conducted a study of malware encounters in a large, multi-national enterprise. They coupled malware encounters with web activities and demographic information, and found that males were more likely to encounter malware than females.

Although some studies suggest that age (Ngo and Paternoster 2011) and gender (Bossler and Holt 2009; Yen et al. 2014) could be significant correlates of malware victimisation, prior work has yielded mixed results in terms of identifying the direction of the aforementioned correlations. For example, Bossler and Holt (2009) found that females are more at risk of malware victimisation, while Yen et al. (2014) found that males are at higher risk. Moreover, all research previously cited performed a global analysis of malware, the exception being the work of Ngo and Paternoster (2011) that limited their study to one type of malware (virus). Our research goes beyond as we also evaluate how the direction and magnitude of age and gender vary between different types of malware. Finally, most of these studies offer surprisingly little or no discussion of how the results should be interpreted in terms of causality. In contrast, we also discuss potential underlying causes of how age and gender may affect the risk of malware victimisation —that is whether they have a direct or indirect effect or whether they are confounded by other factors that were not included in our study.

2.2. Demographics and other computer threats

There is also a number of research that studied the effect of users' demographics on other types of computer threats. For instance, several studies investigated the impact of demographic factors on phishing susceptibility (Sheng et al. 2010; Jagatic et al. 2007; Kumaraguru et al. 2009; Oliveira et al. 2017). Another set of related efforts attempted to examine how demographic factors relate to spam susceptibility (Grimes et al. 2007) or to Internet theft victimization (Reyns 2013). The overwhelming evidence, however, from all these studies suggests that age and gender are significant correlates for computer threats victimization.

3. STUDY DESIGN AND METHODS

Since research on demographic factors associated with malware victimisation is relatively sparse, we will derive our hypotheses from past research on demographics and risk in other domains (e.g. finances, career, sports, health). In other words, we are making the assumption that prior studies on age and gender differences in specific domains can perhaps be extrapolated to the risk of malware victimisation. Hence, by extension, we can hypothesize that (H1) gender and (H2) age are independent risk factors for malware victimisation.

3.1. Case-control study design

We designed a case-control study to test if (H1) gender and (H2) age are independent risk factors of malware victimisation. Commonly used within epidemiology, a case-control study is a type of comparative study where a group of individuals who have a disease (cases) is compared to a group of individuals who do not have the disease (controls). This kind of study is often used to determine whether there is an association between an exposure to a risk (or protective) factor and a disease. In contrast to experimental studies, case-control studies are observational; they do not attempt to alter the course of the disease. Moreover, they are usually, but not exclusively, retrospective by design. They look backwards to learn which individuals in each group (cases and controls) were exposed to the risk (or protective) factor. In other words, once the cases have been identified, the controls are selected from the same population independently of their exposure status.

The frequency of the exposure between the two groups is then compared based on their respective odds of exposure to the potential risk (or protective) factor. From there, the ratio of these odds, the odds ratio (OR), is computed. The confidence interval (CI) in which the true value of the OR is likely to be has to be taken into account when interpreting the OR. An OR larger than 1 indicates that the exposure is a risk factor; the odds of being exposed to the risk factor is higher for the cases than for the controls. To the opposite, an OR smaller than 1 means that the exposure is a protective factor. However, if the OR is equals to 1, or if 1 is included in the CI, nothing can be said on the association between the exposure and the risk of developing the disease.

3.2. Target population

In order to conduct a case-control study as previously described, we must first select a population on which we will base our study. As our focus is the effect of gender and age as potential independent risk factors (exposure) for malware victimisation (disease), we must also consider any other variables that may affect the risk of malware victimisation. To limit the effect of such extraneous factors that are not of primary interest, we decided to limit our population to one operating system (OS) and one antimalware product. More specifically, our *target population* was limited to Windows 10 devices protected by Microsoft's Windows Defender —an antimalware engine included with Windows that helps detect and mitigate malware on computers. We also added the geographical region where the device is located to control for any potential geographical or cultural effects.

3.3. Data collection

As our target population was protected by an antimalware product, malware victimisation was computed based on malware encounters reported on devices protected by Microsoft's Windows Defender. As such, we included both known malware attempting to be installed, and malware already installed on the device. Data on malware encounters was collected from October to November 2015 by Microsoft's Windows Defender. Encounters were recorded on all Windows 10 devices that consistently reported with up-to-date antimalware signatures for the entire study; representing a target population of 30+ million devices. All types of devices were included (for example, desktop PCs, notebooks, and tablets) except mobile devices.

Information on those devices was coupled with demographic data from Microsoft Account, a single sign-on web service that allows users to log into various services provided by Microsoft (for example, Outlook, Skype, OneDrive). For each account, associated gender and age group were used. Gender could be male, female or unknown, and age was grouped in six categories (0-17, 18-24, 25-34, 35-49, 50+, unknown). Accounts that had unknown age or gender or more than three devices associated were excluded from the analysis. However, data on malware encounters is collected on the devices and cannot be uniquely associated with a particular user. For example, if the detection happened due to an action by a particular user, e.g. user-triggered scan, the event that initiated the encounter might be attributable to another previous user. To limit this problem, we decided to consider only data from devices that had only one user account. In particular, we excluded devices that had more than one single account associated. In the end, combining the single-account and known-gender/age criteria, we were left with a *sample population* of 3 019 671 million devices. Further, Internet Protocol (IP) geolocation was used to identify the location of those devices. Locations were grouped into the following six categories: North America, Europe, South and Central America, Australia, Asia and Pacific, Africa and Middle East.

3.4. Ethical and privacy considerations

The telemetry data used in this study was collected by Microsoft in complies with its security and privacy

policies (Microsoft 2017), as well as international laws and regulations. Data was reported to Microsoft only on devices on which open, or blanket, consent was obtained when installing Windows 10; with the possibility to withdraw, or opt out. For the purpose of this study, only anonymous telemetry data limited to the factors identified in the paper was used.

3.5. Statistical analysis

The risk analysis was determined based on the calculation of the odds ratio (OR), with a confidence interval of 95%. Stratified and multivariate analysis through logistic regression was also performed between the dependent variable (malware victimisation) and the independent variables (age, gender, region). The statistical analysis was conducted using Statistica 12.7.

4. RESULTS

4.1. Population

The study lasted two months, and in this period we collected telemetry data from 3 019 671 Windows Defender devices running Windows 10. Table 1 presents the basic demographics of each population by factor.

Case population
Of the total population, 809 426 devices (26.81%) reported malware encounters during the study. Among all cases, we found a male:female ratio of 3.68:1. Similarly to the total population, the 0-17 age group was the less prevalent with only 58 876 users (7.27%). The distribution between the other groups varied from 15.70% (50+) to 29.34% (18-24). For the region, most of the cases were also either from Europe (33.35%) or North America (31.25%).

Control population
Of the 2 210 245 devices (73.19%) in the control population, 27.52% were associated with female users and 72.48% with male users. The less frequent age group was 0-17 (5.35%), with other groups ranging from 20.72% (18-24) to 25.43% (35-49). Approximately 80% of controls were either from Europe (30.46%) or North America (49.03%).

4.2. Malware encounter risk factors

The gender distribution (see Table 1) shows that the proportion of male was greater in the case group (78.64%) than in the total population (74.13%); suggesting that being a male may contribute to increase the risk of malware encounter. With respect to the age groups, an increase in the frequency in the case population was seen for the 0-17, 18-24, and 25-34 age groups; indicating that younger users could be more at risk of encountering malware.

Odds ratio
In order to test the effect of age and gender as risk factors, we computed their respective odds ratio (OR) and confidence interval (CI) at 95%. The effect of gender was investigated with female as a reference level, meaning that male was compared to female. For the effect of age, the age group 50+ was selected as the reference level —all other age groups were tested against this reference. All results in Table 2 were statistically significant at $p-$value < 0.001 when analysed separately.

Table 2: *Odd ratios by factor*

Factor	Description	OR (95% CI)
Gender	Male	1.40 (1.39-1.41)***
Age	0-17	2.04 (2.02-2.06)***
	18-24	2.12 (2.10-2.14)***
	25-34	1.59 (1.57-1.60)***
	35-49	1.25 (1.24-1.26)***
Region	Africa & Middle East	4.02 (3.97-4.07)***
	Asia & Pacific	2.48 (2.46-2.50)***
	Australia	1.18 (1.15-1.20)***
	South & Central America	3.36 (3.33-3.39)***
	Europe	1.72 (1.71-1.73)***

*Statistically significant at 0.05 level; **at 0.01 level; ***at 0.001 level.

Gender was found to be a significant factor associated with malware encounters. More specifically, being a male was identified as a potential risk factor; males were 1.40 times more likely to encounter malware than females. All age groups were shown to be statistically significant risk factors (OR > 1) when compared to the reference level (50+). The groups 0-17 and 18-24 were identified as being the most at risk, followed by the groups 25-34 and 35-49. Overall, results suggest that younger users (0-24) were nearly twice more likely to encounter malware than older users (50+). When analysing if any of the regions were associated with the risk of malware encounter, we found that they were all significant risk factors (OR > 1) when compared to the reference region (North America). Africa & Middle East and South & Central America had the highest odds, while Europe and Australia presented the lowest odds. This suggest that all regions are statistically significantly more at risk of malware encounter than North America. Although those results are of inherent interest, the understanding of these geographical variations in malware exposure is out of scope of this paper. Rather, we will focus our analysis and discussion on age and gender variations in malware exposure.

Stratified and multivariate analysis
To investigate the independent effect of each factor, we used stratification —division of the population in separate groups— to allow the analysis of one factor when controlling for other factors.

Table 1: *Population demographics by factor*

Factor	Description	Total population (N=3 019 671)	Case population (N=809 426)	Control population (N=2 210 245)
Gender	Female	25.87%	21.36%	27.52%
	Male	74.13%	78.64%	72.48%
Age	0-17	5.86%	7.27%	5.35%
	18-24	23.03%	29.34%	20.72%
	25-34	25.39%	26.51%	24.98%
	35-49	24.29%	21.17%	25.43%
	50+	21.43%	15.70%	23.53%
Region	Africa & Middle East	3.12%	5.64%	2.20%
	Asia & Pafific	12.19%	16.68%	10.54%
	Australia	2.37%	1.90%	2.54%
	South & Central America	6.82%	11.18%	5.22%
	North America	44.27%	31.25%	49.03%
	Europe	31.24%	33.35%	30.46%

Results of the stratified analysis (see Table 3) support our initial hypotheses that (H1) gender and (H2) age are independent risk factors for malware encounters. Although being a male was identified as an independent risk factor, the magnitude of its effect was smaller for users in the 35-49 and 50+ age groups. The age was also found to be a significant independent factor after stratification by gender. Interestingly, the impact of age was stronger on male users between 0-34 than on female, which could suggest a potential interaction between the two factors.

Table 3: *Stratified analysis by studied factors*

Risk factor	Stratifying factor	OR (95% CI)
Male gender	Age 0-17	1.53 (1.46-1.60)***
	Age 18-24	1.68 (1.64-1.72)***
	Age 25-34	1.42 (1.39-1.47)***
	Age 35-49	1.17 (1.13-1.21)***
	Age 50+	1.16 (1.11-1.21)***
Age 0-17	Female gender	2.17 (2.06-2.30)***
	Male gender	2.87 (2.79-2.95)***
Age 18-24	Female gender	2.69 (2.69-2.93)***
	Male gender	4.05 (3.97-4.13)***
Age 25-34	Female gender	2.17 (2.07-2.27)***
	Male gender	2.66 (2.61-2.72)***
Age 35-49	Female gender	165 (1.57-1.73)***
	Male gender	1.66 (1.62-1.70)***

*Statistically significant at 0.05 level; **at 0.01 level; ***at 0.001 level.

A logistic regression model was also developed to study the independent effect of age and gender. This kind of regression was selected as our dependent variable (DV) is binary —it can only take two values. The DV was represented by either 1 or 0, where 1 indicates that the device reported at least one malware encounter over the study. The age was considered as an ordinal discrete independent variable and gender was included as a binary independent variable. Region was also included in the regression as a control variable to account for potential cultural and geographical effects. Similarly to our previous analysis, female gender, age group 50+, and North America were used as the reference levels. We report for each factor the Wald statistic and the $p-$value associated. The Wald statistic is used to test the statistical significance of each regression coefficient in the model; the higher the value, the stronger is the effect of the coefficient. The $p-$value indicates if the null hypothesis can be rejected, meaning that the coefficient is relevant in the regression model.

Table 4: *Multiple logistic regression model*

Factor	Wald stat.	$p-$value
Intercept	107 596.70	< 1.00e-16
Gender	4 705.30	< 1.00e-16
Age	26 414.20	< 1.00e-16
Region	88 251.70	< 1.00e-16

Results in Table 4 show that all factors are significant at $p-$value < 1.00e-16 in the regression model, which support our initial hypotheses (H1) and (H2). We also computed from the regression the odds ratio and the 95% CI for each factor. The results in Table 5 also confirm that being a male is a risk factor; males were 1.24 times more likely to encounter malware than female. The associations between malware encouters and age groups were also identified as significant risk factors. The odds of malware encounter increase with age until 18-24, after which they decrease; indicating that users in the group 50+ are less likely to encounter malware than the other age groups.

4.3. Risk factors by malware types

We further wanted to investigate the independent effect of age and gender for different types of malware. Each malware encounter was classified by Microsoft's Windows Defender into a specific malware category. For the purpose of the analysis, malware were grouped in the following 10 categories: adware, virus, cracks, hack, exploit, rogue malware, infostealer, ransomware, bot, and rootkit. See Appendix A for a complete

Table 5: *Odds ratios from multiple logistic regression*

Factor	Description	OR (95% CI)
Gender	Male	1.24 (1.23-1.25)***
Age	0-17	1.74 (1.72-1.76)***
	18-24	1.78 (1.76-1.79)***
	25-34	1.34 (1.33-1.35)***
	35-49	1.13 (1.12-1.14)***
Region	Africa & Middle East	3.65 (3.60-3.70)***
	Asia & Pacific	2.22 (2.20-2.24)***
	Australia	1.12 (1.10-1.14)***
	South & Central America	3.01 (2.98-3.04)***
	Europe	1.63 (1.61-1.64)***

*Statistically significant at 0.05 level; **at 0.01 level; ***at 0.001 level.

definition of each type of malware. Adware (50.04%) represented half of all the encounters, followed by cracks (16.40%), other (15.75%), and virus (9.40%). All the other categories had proportions smaller than 3%: hack (0.77%), exploit (1.45%), rogue malware (1.85%), infostealer (2.77%), ransomware (0.76%), bot (0.65%), and rootkit (0.16%).

Odds ratio

The OR and the 95% CI were computed by studied factors for each type of malware (see Appendix B). Male gender appeared to be a significant risk factor for 8 types of malware: virus, cracks, hack, exploit, infostealer, ransomware, bot, and rootkit. To the opposite, being a male was found to be a weak protective factor for adware encounter (OR=0.98; CI 95%=0.97-0.99); meaning that females were slightly more at risk for this specific type of malware. Moreover, gender was not found to be a significant factor associated with the risk of rogue malware encounter (OR=0.98; CI 95%=0.94-1.02). The same analysis was performed for age by types of malware (see Appendix B). Results show that the effect of all age groups is significant for every types of malware, except for bot and rootkit, where the age groups 0-17 and 35-49 are not statistically significant. Age groups were found to be risk factors —when compared to the reference level (50+)— for 7 types of malware: adware, virus, cracks, hack, exploit, infostealer, bot, and rootkit. To the opposite, all age groups were identified as protective factors for rogue malware and ransomware, meaning that older users (50+) were the most at risk for these specific types of malware. Moreover, results show that the level of risk by age group is function of the type of malware. For example, while users in the 18-24 group are 7.14 times more likely to encounter virus than users in the 50+ group, they are only 1.36 more likely to encounter adware.

Multivariate analysis

Similarly to our previous analysis, we conducted a logistic regression in order to study the effect of age and gender as independent risk factors for different types of malware while controlling for potential regional effect. Results (see Appendix C) show that gender is a significant contributing factor for all types of malware. Interestingly, being a male was found to be a risk factor, expect for adware, where it was found to be a weak protective factor (OR=0.94; CI 95%=0.93-0.95). With respect to age, the effect of all age groups was significant for adware, virus, exploit, and infostealer. However, only one age group was found to be significant for bot and rootkit; suggesting that age may not be an important factor for those types of malware. The odds of infostealer and hack encounters were found to decrease with age. Whereas the odds of virus and cracks encounters exhibited an inverted U-shape trend with age; encounters increase from teenagers (0-17) to young users (18-24), before reducing with age. To the opposite, ransomware and rogue malware encounters were found to increase with age; users in the 50+ age group the most at risk. Hence, hypothesis (H1) is supported for all types of malware and (H2) is only partially supported, as not all age groups were found to be significant.

5. DISCUSSION

In this section, we give our interpretation of the findings previously reported, focusing on the most interesting results we found. We also compare our results to those reported in prior studies where possible, and highlight instances in which our findings corroborate or refute theirs.

Overall, we found that age and gender are independent risk factors for malware victimisation; males were found to be more at risk of being exposed to malware than females, and younger users at higher risk than older users. As we discuss below, however, the direction and magnitude of the effect of age and gender vary in some surprising ways depending on the type of malware.

5.1. Gender difference

The risk analysis allowed to identify gender as a significant independent factor related to malware victimisation. Males were found to be 1.24 times more likely to encounter malware than females. This gender difference was most marked in the population under the age of 25 years, but was also evident among older users. Similarly, Yen et al. (2014) found that males were more at risk of encountering malware than females. To the opposite, Bossler and Holt (2009) found that females were more susceptible to malware victimization, as measured by self-reported data loss from malware. However, direct comparison with our results is not possible, as previous work used different study design and target population; Yen et al. (2014) studied malware encounters of corporate users within a large enterprise, and Bossler and Holt (2009) based his study on self-reported malware victimisation from college students.

When performing the risk analysis for different types of malware, we also found that being a male was a risk factor, except for adware. For this specific type of malware, being a male was a weak significant protective factor; meaning that females were slightly more susceptible to encounter adware than males. We present in the following text potential underlying causes that could explain such difference across gender and types of malware.

Risk attitude
A first possibility for this gender difference could be that males are more susceptible to malware victimisation than females because of their attitude towards cybersecurity-related risk. This could be plausible as gender differences in risk attitude has been identified across various contexts, such as car driving, financial matters, health, social decisions, sport and leisure, and career (Byrnes et al. 1999; Weber et al. 2002; Dohmen et al. 2011; Harris et al. 2006). Though there is extensive evidence to show that males are more risk seeking than females overall, the direction and magnitude of the gender effect tend to depend of the domain. For example, while male are more likely to exhibit risky behaviors in car driving, researchers found that female report greater propensity than male to engage in risky behaviors when it comes to social decisions Weber et al. (2002); Johnson et al. (2004). These variations across domains have been attributed, among others, to gender differences in (1) the perceived probability of negative consequences, (2) the perceived severity of a potential negative consequences, and (3) the enjoyment of engaging in risky behaviors (Harris et al. 2006; Emond et al. 2009; Wang et al. 2011; Hogarth et al. 2007). However, the extent to which they are the product of genetic, social, developmental, or experimental factors is still lacking strong consensus in the research litterature.

Similarly, one could argue that males are more likely to encounter malware than females because of gender differences in risk perception and enjoyment of risky behaviors in cybersecurity. This could imply that (1) males have lower perceptions of the probabilities and severity of negative consequences from engaging in risky behaviors in cybersecurity, and (2) they expect higher enjoyment than females from these behaviors. Those gender differences could explain, for example, why males were found to be 1.65 times more likely to encounter cracks –tools often used to engage in software piracy– than females.

Computer usage
With car driving, we know that both driving behaviors and time spent on the road are significant contributing factors to the risk of car injury. Similarly, a second explanation could be the difference in frequency and type of computer usage behavior between male and female (Hu et al. 2007; Joiner et al. 2012; Goel et al. 2012). This is consistent with previous work that identified gender differences in frequency and patterns of Internet use. For example, Joiner *et al.* conducted a survey of 501 students and found males to be heavier Internet users than females (Joiner et al. 2012). Males were more likely to use the Internet for games and entertainment, to bet online, to visit web sites with adult content, and to download music and videos. On the other hand, females were more likely to use the Internet for communication (e.g. email, telephone), and visit social network sites. In another study, Goel *et al.* examined the Web histories of 250,000 anonymized individuals paired with user-level demographics. They found that females spend considerably more time online on social media sites, and that visits to sports sites are highly predictive of being male (Goel et al. 2012).

Moreover, several research found empirical evidence of associations between the frequency and type of computer usage and the risk of malware exposure. Carlinet et al. (2008) performed a case-control study based on network traffic of a large set of real ADSL customers. They found that surfing the web a lot and high usage of streaming applications are risk factors to being infected with malware. In another study, Lalonde Lévesque *et al.* (Lalonde Lévesque et al. 2013, 2014) found evidence that installing many applications and visiting many web sites may increase the risk of malware encounters. They also identified specific categories of web sites, most of which were legitimate, that were more likely to be associated with increased risk of malware encounters. Similar results were also obtained by Canali et al. (2014). The authors developed a risk model of malware encounter based on users' web browsing behavior. They used a large telemetry dataset collected by a major antimalware vendor and identified specific web sites categories, and the total number of web sites visited, as good predictors of the likelihood of encountering malware. This last finding was also supported by the work of Yen et al. (2014), which identified a positive correlation between the volume of user activity (as measured by the number of distinct domains visited by a host) and the probability of encountering malware.

Overall, the studies cited above support the existence of a relationship between web browsing behavior and the risk of malware victimisation. This trend is consistent with recent observations and reports by the antivirus (AV) industry. In particular, a recent report by Microsoft (Anthe and Chrzan 2015) identifies *web browsing* as being the most frequent transmission vector used by malware for the first quarter of 2015 (Anthe and Chrzan 2015), the period just 6 months ahead of our study. Although results are not limited to Windows 10 users, they provide strong evidence that most users encountered malware because they either visited a malicious or compromised web page, or

downloaded a malicious application (voluntarily or not). For example, users can get infected through malvertising —malicious advertising— by clicking on an innocuous-looking banner ads containing malicious code (Sood and Enbody 2011; Xing et al. 2015). Other attacks, such as drive-by downloads (Mavrommatis and Monrose 2008; Provos et al. 2007), can download malware without any user intervention required, by either operating malicious web sites or by injecting malicious content into compromised legitimate web sites. Finally, users can also get infected by downloading a piece of software (e.g. free games, media players, screen savers, keygens) that comes bundled with spyware, adware or malware.

In light of this discussion, males could be more at risk of encountering malware than females because (1) they are heavier computer users (e.g. they visit more web sites, they install more applications), and (2) they are more prone to engage in computer behaviors that may, intentionally or not, increase their likelihood of encountering malware. Similarly, females could be more at risk of adware encounter as a result of differences in their computer behavior (e.g. categories of web sites visited, type of applications installed). Although these hypotheses are plausible, additional research should be conducted in order to gain a better understanding of how computer usage behavior affect the risk of malware victimisation, and establish sound causation.

5.2. Age difference

Results suggest that age is a significant independent risk factor for malware victimisation. Young users (0-24 years), in particular users in the 18-24 age group, were the most likely to encounter malware. To the opposite, older users (50+) were found to be the less susceptible to encounter malware. This supports the findings of Ngo and Paternoster (2011) that suggest that older users are less likely to get infected by malware.

Our risk analysis by types of malware reveals, however, that the direction and magnitude of the age effect is a function of the type of malware. Although increasing age was associated with reduced malware encounters overall, its effect was particularly strong for virus and infostealer encounters, and relatively small for bot and rootkit encounters. Moreover, while older users (50+) were found to be less at risk of encountering malware overall, they were the more susceptible to encounter rogue malware and ransomware. We present in the following text potential causes for these age differences.

Risk attitude
Similarly to gender, age differences in malware victimisation could be attributed to variations across age groups in risk attitude towards cybersecurity. In comparison, age differences in risk-taking behaviors have also been identified in multiple risk domains (Rolison et al. 2013; Dohmen et al. 2011, 2005). There is an overwhelming consensus that young age is associated with higher willingness to take risks than older age (Dohmen et al. 2005). However, studies also reveal that age differences in risk-taking may depend on the domain. For example, Rolison et al. (2013) found that risk taking in the financial domain reduces steeply with older age, while in the social domain, it increases slightly from young to middle age, before reducing sharply in later life. A number of possible underlying causes, such as (1) changes in life circumstances, (2) motivational factors, and (3) cognitive decline, have been advanced to explain such variations (Rolison et al. 2013; Mather 2006). While these causes might be relevant for risk tendencies in specific domains (e.g. financial, social, recreational), risk attitude in cybersecurity may differ, and point to different underlying causes. As with gender, we believe that age differences in malware victimisation may be, to some extent, attributed to age changes in risk perceptions and expected enjoyment of engaging in risky behaviors. Specifically, this could imply that (1) younger users have lower perceptions of the probabilities and severity of negative consequences from engaging in risky behaviors in cybersecurity, and that (2) they expect higher enjoyment than older users from engaging in risky behaviors.

Another possibility could be that malware encounters differ across age groups as a result of changes in emotional processing. This is consistent with previous research in psychology, sociology and economics, that identified emotion to be a major determinant of risk perception and risk taking that changes with age (Figner et al. 2009). While emotions are found to act as an *advisor* for risk taking in situations of low level of emotional intensity, they seem to inhibit cognitive processes in situations of high level of emotional intensity (Bieberstein 2013). Emotional differences could therefore explain why older users (50+) are more likely to encounter rogue malware and ransomware than younger users. As those categories of malware are known to use deceptive fear to trick users into downloading a malicious software (a trial version of a bogus security software or a fake software update), older users could be more likely to act by emotions rather then by cognitive processes when exposed to such trickery. Hence, older users would be more susceptible to rogue malware and ransomware because of emotional differences when faced with persuasive messages that attempt to scare them.

Computer usage
Another likely reason could be age differences in frequency and type of computer usage. This is supported by prior studies that identified differences in volume and type of computer activities across age. By analysing the web histories of 250,000 individuals, Goel et al. (2012) found that younger users spend much more time online relative to their older counterparts.

Their results also reveal that older users spend a smaller fraction of their online time on social media web sites. In another study, Teo (2001) conducted a web-based survey of 1,370 respondents to examine how demographics variables and motivation variables correlate with Internet usage activities (messaging, browsing, downloading, and purchasing). Their results show that younger users engage in messaging and downloading activities to a greater extent than older users.

Taken together with previous findings of the relationship between computer usage and risk of malware victimisation (as presented in Section 5.1), we can hypothesize that younger users could be more likely to encounter malware because (1) they are heavier computer users, and (2) they engage in computer activities that could contribute to increase (intentionally or not) their risk of malware victimisation. Furthermore, older users could be more likely to encounter rogue malware and ransomware because of their computer activities. This is possible, as rogue malware and ransomware are known to target specific countries, OSes, programs, companies, or web site categories. Similarly, older users could engage in computer activities (e.g. visiting specific categories of web sites, installing/using specific types of applications) that would increase their likelihood of encountering such attacks. However, the extent to which older users are more exposed as a result of their computer activities, or because they are seen as attractive targets (lack of Internet savvy, potential access to life savings, and impaired decision making due to ageing) remains unknown.

5.3. Summary of findings

We presented in this paper a number of interesting findings related to gender and age differences in the risk of malware encounters. The key findings of our study can be summarized as follow:

- Age and gender are significant independent factors of malware encounter.
- Male, and young male in particular, are more likely to encounter malware than female.
- Female are slightly more at risk of encountering adware than male.
- The gender difference is most marked in the population under the age of 25 years, but is also evident among older users.
- Increasing age leads to decreasing risk of malware encounter; younger users (0-24) are more at risk of encountering malware than older users (50+);
- Older users (50+) are the most susceptible to encounter rogue malware and ransomware.

6. STUDY LIMITATIONS

Although case-control studies allow determination of whether an exposure is associated with an outcome, their results can be highly sensitive to bias, confounding variables, and chance circumstances. Hence, our study and its conclusions are subject to a number of limitations and potential bias that may affect its internal and external validity. Internal validity refers to the strength of the inferences from the study, that is the extent to which no other variables except the one we studied caused the results. While external validity refers to the ability to generalize the results to a more *universal population*.

First, malware encounters are limited to the malware families detected by Microsoft's Windows Defender. While these malware may represent some of the most significant malware families on Windows, they do not cover targeted attacks and zero-day attacks. Moreover, the encounters reported depend on the efficacy of Windows Defender, which may lead to an underestimation of malware encounters. Nevertheless, given the significance of the malware families covered by Windows Defender, these encounters are also of inherent interest, whether or not they are representative of all computer threats on Windows 10.

Second, the sample population is limited to devices that have known age/gender and a single-account associated. Hence, the exclusion of devices with multiple accounts, or with missing demographic information may have introduced a sampling bias. In order to estimate this potential bias, we compared our sample population (3+ million) against our target population (30+ million). We found that both populations were similar in terms of geographical distribution and malware encounters. However, this does not imply that the two populations are similar in terms of other factors, such as demographics or risk attitude. Moreover, our sample population may not be representative of the target population for other time frames. As security data are known to be dynamic, a sample population drawn from the same target population at another time-period may be different. This could be particularly true as our study was conducted few months (Oct.-Nov. 2015) after the official release of Windows 10 (July 2015); meaning the target population may evolve over time as more users adopt Windows 10.

Another limitation of our study is its susceptibility to confounding. Although the region factor was included in our analysis to account for potential geographical or cultural effect, and multivariate analysis was used, we cannot guarantee that our results were not affected by other unknown extraneous variables that may confound the results. For example, it may be possible that in some cases several human users shared the same user account on the same single-account device, which may have introduced a bias that we were not able to

control nor measure. It would be interesting in future work to consider additional extraneous variables, such as education or social status.

Finally, a significant limitation to our external validity derives from our target population –Windows 10 devices protected by Microsoft's Windows Defender. As our analysis was limited to Windows 10 devices, it does not provide insight into other versions of Windows (e.g. Windows Mobile, Vista, XP, etc.), and it does not give insight into the encounter rates on non-Windows systems such as MacOS and Unix-based OS. Furthermore, the analysis was limited to Windows 10 devices running Windows Defender. Thus, it does not cover users protected by other antimalware products. However, given that Defender was running on more than 40% of all Windows 10 devices during the period covered by our study, we believe our findings are important on their own, whether or not they are representative of patterns in devices protected by other antimalware products. Though we agree that a study including multiple antimalware products is interesting and would provide additional insights, such analysis was outside the scope of this study.

7. CONCLUSION

We presented the results of a large scale empirical study specifically designed to evaluate gender and age as potential independent risk factors of malware victimisation. While our work corroborates some findings in earlier research, our results support our initial hypothesis, that both (H1) gender and (H2) age are contributing independent factors correlated to the risk of malware victimisation. Those results were also robust after stratification and multivariate analysis. Male and younger users were found to be more at risk of malware encounter overall, though the direction and magnitude of the gender and age differences varied depending on the type of malware. Interestingly, certain types of malware were associated with nontrivial age differences (e.g. ransomware and rogue malware), whereas others were associated with gender differences that shifted from risk factor to protective factor (e.g. adware).

It is clear from the evidence that differences between the age groups and gender exist in the context of malware victimisation. The remaining question concerns the origins of these associations, i.e. their causality. We have discussed potential underlying causes that could explain why age and gender are risk factors. In particular, we hypothesize that differences in attitude towards risk taking and differences in computer and Internet usage, which have been reported to change with age and gender, could explain the differences in malware victimisation. Verifying these causal hypotheses is essential for the design of successful targeted, age and gender differentiated interventions aimed at preventing or reducing the risk of malware victimisation.

In particular, this study and its findings may help support the development of user-differentiated human-computer systems. As systems designed to suit the *average user* may not accommodate all user groups (Egelman and Peer 2015), security systems could be tailored to users' risk of victimisation. For instance, one recent study provided preliminary evidence that antivirus effectiveness differs significantly across demographic factors; antivirus had lower performance for female users and the 0-17 age group (Lalonde Lévesque et al. 2016). Demographic factors could then be used to infer the risk of malware victimisation, and personalize systems (default security settings, human-computer interfaces, etc.) in order to maximize protection for all user groups.

In addition, this could have potential implications for the cyberinsurance industry as well. For example, in the car insurance industry personal characteristics such as age, gender, and marital status are often used as proxies of driver behavior (accelerating, braking, etc.) and driving characteristics (where, when, etc.). Although finer-grained data on driving can be collected through vehicle telematics, the use of such devices is not always available for drivers and insurers. Besides, the collection of such data brings up a number of privacy concerns and other ethical issues, especially concerning computer and Internet behaviour and usage, which is potentially much more privacy-invasive than driving data. Thus, we believe that it could be useful to develop predictive user risk models that use coarse non-invasive information, in order to address these privacy concerns, while supporting the risk-selection needs of the cyber insurance industry.

Furthermore, more studies are needed based on alternate observational data sources, other time frames and different analysis methods in order to confirm that our findings are robust across different populations. Finally, we believe it is important to try to identify and validate the potential causality of other risk factors that may be associated with malware victimisation, such as other personal traits, and socio-economical and cultural factors. Determining *who* is more susceptible to malware victimisation and *why* is paramount to improve security for *all* users.

8. ACKNOWLEDGEMENTS

The authors would like to thank the Microsoft Malware Protection Center (MMPC) for supporting this work and granting us access to the Microsoft's Windows Defender telemetry data.

REFERENCES

Anthe, C. and P. Chrzan (2015). Microsoft Security Intelligence Report January-June 2015.

Bieberstein, A. (2013). *An Investigation of Women's and Men's Perceptions and Meanings Associated with Food Risks*. Springer Science & Business Media.

Bossler, A. M. and T. J. Holt (2009). On-line activities, guardianship, and malware infection: An examination of routine activities theory. *International Journal of Cyber Criminology 3*(1), 400.

Byrnes, J. P., D. C. Miller, and W. D. Schafer (1999). Gender differences in risk taking: A meta-analysis. *Psychological bulletin 125*(3), 367.

Canali, D., L. Bilge, and D. Balzarotti (2014). On the effectiveness of risk prediction based on users browsing behavior. In *Proceedings of the 9th ACM symposium on Information, computer and communications security*, pp. 171–182. ACM.

Carlinet, Y., L. Mé, H. Débar, and Y. Gourhant (2008). Analysis of computer infection risk factors based on customer network usage. In *Second International Conference on Emerging Security Information, Systems and Technologies (SECURWARE'08)*, pp. 317–325.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Dohmen, T. J., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey. *Social Science Research Network*.

Egelman, S. and E. Peer (2015). The myth of the average user: Improving privacy and security systems through individualization. In *Proceedings of the 2015 New Security Paradigms Workshop*, pp. 16–28. ACM.

Emond, C., W. Tang, and S. Handy (2009). Explaining gender difference in bicycling behavior. *Transportation Research Record: Journal of the Transportation Research Board 2125*, 16–25.

Figner, B., R. J. Mackinlay, F. Wilkening, and E. U. Weber (2009). Affective and deliberative processes in risky choice: age differences in risk taking in the columbia card task. *Journal of Experimental Psychology: Learning, Memory, and Cognition 35*(3), 709.

Goel, S., J. M. Hofman, and M. I. Sirer (2012). Who does what on the web: A large-scale study of browsing behavior. In *ICWSM*.

Grimes, G. A., M. G. Hough, and M. L. Signorella (2007). Email end users and spam: relations of gender and age group to attitudes and actions. *Computers in Human Behavior 23*(1), 318–332.

Harris, C. R., M. Jenkins, and D. Glaser (2006). Gender differences in risk assessment: why do women take fewer risks than men? *Judgment and Decision Making 1*(1), 48.

Hogarth, R. M., M. Portell, and A. Cuxart (2007). What risks do people perceive in everyday life? A perspective gained from the experience sampling method (esm). *Risk Analysis 27*(6), 1427–1439.

Hu, J., H.-J. Zeng, H. Li, C. Niu, and Z. Chen (2007). Demographic prediction based on user's browsing behavior. In *Proceedings of the 16th international conference on World Wide Web*, pp. 151–160. ACM.

Jagatic, T. N., N. A. Johnson, M. Jakobsson, and F. Menczer (2007). Social phishing. *Communications of the ACM 50*(10), 94–100.

Johnson, J., A. Wilke, and E. U. Weber (2004). Beyond a trait view of risk taking: A domain-specific scale measuring risk perceptions, expected benefits, and perceived-risk attitudes in german-speaking populations. *Polish Psychological Bulletin 35*, 153–172.

Joiner, R., J. Gavin, M. Brosnan, J. Cromby, H. Gregory, J. Guiller, P. Maras, and A. Moon (2012). Gender, internet experience, internet identification, and internet anxiety: a ten-year followup. *Cyberpsychology, Behavior, and Social Networking 15*(7), 370–372.

Kumaraguru, P., J. Cranshaw, A. Acquisti, L. Cranor, J. Hong, M. A. Blair, and T. Pham (2009). School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, pp. 3. ACM.

Lalonde Lévesque, F., J. M. Fernandez, and A. Somayaji (2014). Risk prediction of malware victimization based on user behavior. In *Malicious and Unwanted Software: The Americas (MALWARE), 2014 9th International Conference on*, pp. 128–134. IEEE.

Lalonde Lévesque, F., J. Nsiempba, J. M. Fernandez, S. Chiasson, and A. Somayaji (2013). A clinical study of risk factors related to malware infections. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, pp. 97–108. ACM.

Lalonde Lévesque, F., J. M. Fernandez, D. Batchelder, and G. Young (2016, October). Are they real? real-life comparative tests of anti-virus products. In *26th Virus Bulletin International Conference*, pp. 1–7.

Mather, M. (2006). A review of decision-making processes: Weighing the risks and benefits of aging. *When I?m 64*(145), 145–173.

Mavrommatis, N. P. P. and M. A. R. F. Monrose (2008). All your iframes point to us. In *USENIX Security Symposium*, pp. 1–16.

Microsoft (2017). Microsoft Privacy Statement. `https://privacy.microsoft.com/en-gb/privacystatement`.

Ngo, F. T. and R. Paternoster (2011). Cybercrime victimization: An examination of individual and situational level factors. *International Journal of Cyber Criminology 5*(1), 773–793.

Oliveira, D., H. Rocha, H. Yang, D. Ellis, S. Dommaraju, M. Muradoglu, D. Weir, A. Soliman, T. Lin, and N. Ebner (2017). Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 6412–6424. ACM.

Provos, N., D. McNamee, P. Mavrommatis, K. Wang, N. Modadugu, et al. (2007). The ghost in the browser: Analysis of web-based malware. *HotBots 7*, 4–4.

Reyns, B. W. (2013). Online routines and identity theft victimization further expanding routine activity theory beyond direct-contact offenses. *Journal of Research in Crime and Delinquency 50*(2), 216–238.

Rolison, J. J., Y. Hanoch, S. Wood, and P.-J. Liu (2013). Risk-taking differences across the adult life span: a question of age and domain. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, gbt081.

Sheng, S., M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 373–382.

Sood, A. K. and R. J. Enbody (2011). Malvertising–exploiting web advertising. *Computer Fraud & Security 2011*(4), 11–16.

Teo, T. S. (2001). Demographic and motivation variables associated with internet usage activities. *Internet Research 11*(2), 125–137.

Wang, M., C. Keller, and M. Siegrist (2011). The less you know, the more you are afraid of?a survey on risk perceptions of investment products. *Journal of Behavioral Finance 12*(1), 9–19.

Weber, E. U., A.-R. Blais, and N. E. Betz (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of behavioral decision making 15*(4), 263–290.

Xing, X., W. Meng, B. Lee, U. Weinsberg, A. Sheth, R. Perdisci, and W. Lee (2015). Understanding malvertising through ad-injecting browser extensions. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 1286–1295. International World Wide Web Conferences Steering Committee.

Yen, T.-F., V. Heorhiadi, A. Oprea, M. K. Reiter, and A. Juels (2014). An epidemiological study of malware encounters in a large enterprise. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1117–1130. ACM.

A. DEFINITIONS BY MALWARE TYPE

Adware: Software that shows you extra promotions that you cannot control as you use your PC.

Bot: Small, hidden programs that are often controlled by a malicious hacker. Bots can be installed on your PC without you knowing.

Cracks: A type of tool that can be used to activate an unregistered copy of a software.

Exploit: A piece of code that uses software vulnerabilities to access information on your PC or install malware.

Hack: A type of tool that can be used to allow and maintain unauthorized access to your PC.

Infostealer: A type of malware that is used to steal your personal information, such as user names and passwords.

Ransomware: A type of malware that can stop you from using your PC, or encrypt your files so you cannot use them. You may be warned that you need to pay money, complete surveys, or perform other actions before you can use your PC again.

Rogue: Software that pretends to be an antivirus program but doesn't actually provide any security. This type of software usually gives you a lot of alerts about threats on your PC that don't exist. It also tries to convince you to pay for its services.

Rootkit: A program that is designed to hide itself and other malware from detection while it makes changes to your PC.

Virus: Type of malware that spread on their own by attaching their code to other programs, or copying themselves across systems and networks.

B. ODDS RATIOS BY MALWARE TYPE

Table 6: *Odds ratios for gender by malware type*

Malware	OR (95% CI)	$p-$value
Adware	0.98 (0.97-0.99)	2.17e-10
Virus	1.66 (1.63-1.68)	< 1.00e-16
Cracks	2.01 (1.97-2.04)	< 1.00e-16
Hack	3.13 (2.88-3.40)	< 1.00e-16
Exploit	1.73 (1.65-1.82)	< 1.00e-16
Rogue	0.98 (0.94-1.02)	4.11e-01
Infostealer	1.97 (1.89-2.04)	< 1.00e-16
Ransomware	1.32 (1.24-1.40)	< 1.00e-16
Bot	2.09 (1.73-2.5)	< 1.00e-16
Rootkit	2.25 (1.92-2.64)	< 1.00e-16

Table 7: *Odds ratios for age by malware type*

Malware	Age	OR(95% CI)	$p-$value
Adware	0-17	1.75 (1.72-1.77)	< 1.00e-16
	18-24	1.36 (1.35-1.38)	< 1.00e-16
	25-34	1.01 (0.99-1.02)	0.00e-01
	35-49	0.99 (0.98-1.01)	< 1.00e-16
Virus	0-17	4.38 (4.21-4.56)	< 1.00e-16
	18-24	7.14 (6.93-7.37)	< 1.00e-16
	25-34	3.91 (3.79-4.03)	< 1.00e-16
	35-49	2.37 (2.29-2.46)	< 1.00e-16
Cracks	0-17	2.38 (2.31-2.45)	3.82e-11
	18-24	3.78 (3.70-3.86)	< 1.00e-16
	25-34	3.16 (3.09-3.22)	< 1.00e-16
	35-49	1.95 (1.91-1.99)	< 1.00e-16
Hack	0-17	4.40 (3.94-4.91)	< 1.00e-16
	18-24	3.18 (2.90-3.49)	< 1.00e-16
	25-34	2.45 (2.23-2.69)	8.15e-03
	35-49	1.88 (1.71-2.08)	1.93e-14
Exploit	0-17	0.76 (0.69-0.84)	< 1.00e-16
	18-24	1.10 (1.04-1.16)	6.49e-03
	25-34	1.27 (1.21-1.35)	< 1.00e-16
	35-49	1.16 (1.10-1.22)	1.37e-08
Rogue	0-17	0.51 (0.47-0.55)	2.56e-08
	18-24	0.47 (0.44-0.48)	< 1.00e-16
	25-34	0.54 (0.51-0.56)	1.71e-11
	35-49	0.64 (0.61-0.67)	4.02e-04
Infostealer	0-17	4.21 (3.94-4.49)	< 1.00e-16
	18-24	4.57 (4.33-4.81)	< 1.00e-16
	25-34	3.26 (3.09-3.44)	< 1.00e-16
	35-49	1.99 (1.88-2.10)	< 1.00e-16
Ransomware	0-17	0.68 (0.60-0.77)	1.70e-03
	18-24	0.65 (0.61-0.71)	4.60e-12
	25-34	0.76 (0.71-0.82)	9.02e-02
	35-49	0.93 (0.87-0.99)	2.15e-10
Bot	0-17	1.20 (1.05-1.38)	1.07e-01
	18-24	1.69 (1.55-1.84)	< 1.00e-16
	25-34	1.46 (1.34-1.59)	1.90e-05
	35-49	1.27 (1.17-1.39)	3.93e-01
Rootkit	0-17	1.19 (0.89-1.60)	8.61e-02
	18-24	1.76 (1.47-2.09)	1.62e-04
	25-34	1.93 (1.62-2.29)	6.85e-09
	35-49	1.48 (1.23-1.77)	5.24e-01

C. MULTIPLE LOGISTIC REGRESSION BY MALWARE TYPE

We present for each type of malware the results of the logistic regression. The odds ratio (OR) and its associated confidence interval (CI) at 95% were computed and are shown. We used the $p-$value as an indicator of whether the difference in exposure between the cases and the controls is statistically significant: * indicates that the effect is statistically significant at 0.05 level; ** at 0.01 level; and *** at 0.001 level.

Table 8: *Multiple logistic regression for adware*

Factor	Description	OR (95% CI)
Gender	Male	0.94 (0.93-0.95)***
Age	0-17	1.59 (1.57-1.61)***
	18-24	1.28 (1.27-1.29)***
	25-34	0.95 (0.94-0.96)***
	35-49	0.96 (0.95-0.97)***
Region	Africa & Middle East	2.43 (2.39-2.47)***
	Asia & Pacific	1.07 (1.05-0.08)***
	Australia	1.15 (1.13-1.17)***
	South & Central America	1.82 (1.80-1.84)***
	Europe	1.40 (1.39-1.41)

Table 9: *Multiple logistic regression for virus*

Factor	Description	OR (95% CI)
Gender	Male	1.21 (1.19-1.23)***
Age	0-17	3.27 (3.14-3.40)***
	18-24	4.51 (4.37-4.66)***
	25-34	2.52 (2.44-2.60)***
	35-49	1.86 (1.80-1.93)***
Region	Africa & Middle East	14.67 (14.27-15.07)***
	Asia & Pacific	8.34 (8.20-8.58)***
	Australia	0.48 (0.43-0.54)***
	South & Central America	6.30 (6.13-6.47)***
	Europe	1.28 (1.24-1.31)***

Table 10: *Multiple logistic regression for cracks*

Factor	Description	OR (95% CI)
Gender	Male	1.65 (1.62-1.67)***
Age	0-17	1.87 (1.81-1.92)
	18-24	2.86 (2.80-2.92)***
	25-34	2.41 (2.36-2.46)***
	35-49	1.68 (1.64-1.72)***
Region	Africa & Middle East	4.91 (4.79-5.03)***
	Asia & Pacific	3.30 (3.24-3.36)***
	Australia	1.37 (1.31-1.43)***
	South & Central America	4.81 (4.72-4.90)***
	Europe	2.16 (2.13-2.20)***

Table 11: *Multiple logistic regression for hack*

Factor	Description	OR (95% CI)
Gender	Male	2.68 (2.47-2.91)***
Age	0-17	3.35 (2.30-3.74)***
	18-24	2.40 (2.18-2.63)***
	25-34	1.81 (1.65-1.99)
	35-49	1.58 (1.43-1.74)***
Region	Africa & Middle East	7.64 (6.94-8.41)***
	Asia & Pacific	2.39 (2.18-2.61)***
	Australia	1.35 (1.08-1.68)***
	South & Central America	4.97 (4.55-5.42)***
	Europe	2.90 (2.70-3.12)***

Table 12: *Multiple logistic regression for exploit*

Factor	Description	OR (95% CI)
Gender	Male	1.48 (1.41-1.55)***
Age	0-17	0.67 (0.61-0.74)***
	18-24	0.88 (0.83-0.93)*
	25-34	1.04 (0.99-1.10)***
	35-49	1.05 (0.99-1.10)***
Region	Africa & Middle East	1.53 (1.37-1.70)
	Asia & Pacific	3.55 (3.38-3.73)***
	Australia	0.71 (0.60-0.85)***
	South & Central America	1.95 (1.81-2.09)***
	Europe	1.38 (1.31-1.45)**

Table 13: *Multiple logistic regression for rogue*

Factor	Description	OR (95% CI)
Gender	Male	1.35 (1.30-1.40)***
Age	0-17	0.82 (0.76-0.89)
	18-24	0.66 (0.63-0.69)***
	25-34	0.69 (0.66-0.72)***
	35-49	0.74 (0.71-0.77)*
Region	Africa & Middle East	0.03 (0.02-0.04)***
	Asia & Pacific	0.007 (0.005-0.009)***
	Australia	0.27 (0.24-0.31)***
	South & Central America	0.009 (0.006-0.014)***
	Europe	0.05 (0.05-0.06)

Table 14: *Multiple logistic regression for infostealer*

Factor	Description	OR (95% CI)
Gender	Male	1.52 (1.47-1.58)***
Age	0-17	2.64 (2.47-2.82)***
	18-24	2.52 (2.39-2.66)***
	25-34	1.92 (2.39-2.65)***
	35-49	1.48 (1.40-1.57)***
Region	Africa & Middle East	7.53 (7.04-8.06)***
	Asia & Pacific	7.85 (7.47-8.25)***
	Australia	1.59 (1.38-1.84)***
	South & Central America	23.73 (22.64-24.87)***
	Europe	2.04 (1.93-2.15)***

Table 15: *Multiple logistic regression for ransomware*

Factor	Description	OR (95% CI)
Gender	Male	1.40 (1.32-1.49)***
Age	0-17	0.71 (0.62-0.80)***
	18-24	0.70 (0.65-0.75)***
	25-34	0.78 (0.72-0.84)
	35-49	0.93 (0.87-1.00)***
Region	Africa & Middle East	1.17 (1.02-1.33)***
	Asia & Pacific	0.47 (0.43-0.53)***
	Australia	0.90 (0.76-0.07)
	South & Central America	0.46 (0.40-0.53)***
	Europe	1.06 (1.00-1.13)***

Table 16: *Multiple logistic regression for bot*

Factor	Description	OR (95% CI)
Gender	Male	1.53 (1.42-1.65)***
Age	0-17	0.98 (0.85-1.12)
	18-24	1.18 (1.09-1.29)***
	25-34	1.06 (0.97-1.16)
	35-49	1.07 (0.98-1.17)
Region	Africa & Middle East	4.94 (4.40-5.54)***
	Asia & Pacific	5.88 (5.45-6.34)***
	Australia	1.40 (1.13-1.76)***
	South & Central America	3.03 (2.73-3.36)***
	Europe	1.51 (1.39-1.64)***

Table 17: *Multiple logistic regression for rootkit*

Factor	Description	OR (95% CI)
Gender	Male	1.65 (1.41-1.94)***
Age	0-17	1.10 (0.82-1.48
	18-24	1.23 (1.02-1.47)
	25-34	1.43 (1.20-1.70)***
	35-49	1.29 (1.07-1.55)
Region	Africa & Middle East	0.71 (0.44-1.14)*
	Asia & Pacific	7.80 (6.83-8.92)***
	Australia	1.74 (1.20-2.51)*
	South & Central America	0.43 (0.28-0.66)***
	Europe	0.69 (0.57-0.83)***