Estimating The Number Of Ransomware Attacks

January 11, 2025

Abstract

Accurate crime measurement is essential for scientists, policymakers, and the public. Traditional methods, such as self-reporting and official statistics, face challenges in reliability, validity, and sampling, particularly with the rise of online crime. This study estimates the prevalence of ransomware attacks, a growing concern due to their high costs and societal impact. Using data from police reports, incident response companies, and leak pages, we apply capture-recapture methodology to estimate ransomware incidents among large, medium, and small businesses in the Netherlands from 2019 to 2022. We estimate that there were 138 ransomware attacks against large companies, over the four years period used (2019-2023), 219 attacks among medium companies and 2373 attacks against small companies. The estimate for small companies, however, is judged to be too large to be reliable. We calculate that there is an average annual risk of 1.3% for large companies and 0.6% for medium companies of becoming a ransomware victim. Our results show significant underreporting, with only 41.4% of attacks observed for large companies and 40.2% for medium companies reported to the police. However, this level of police reporting is still larger than for victims of other types of cybercrime. Despite these limitations, our findings align closely with the Dutch large-scale victimization survey, the Statistics Netherlands Cybersecurity Monitor, reinforcing the robustness of both approaches. The results highlight the value of crime-specific datasets but underscore the need for other data sources of ransomware attacks on small companies.

Index terms— Ransomware - Measurement - Capture-Recapture Methodology - Police - Victimization Surveys - Cybercrime

1 Introduction

Knowing how much crime there is in a country is important for a number of reasons, among them government accountability, informs public awareness and research, and aids in resource allocation [61]. Traditionally there are three main sources of crime statistics: self-report victimization, self-report offending, and official statistics on recorded crime by law enforcement agencies [46]. Measures based on interviews, whether they focus on offending [41, 63] or victimization

[7, 18, 27], have problems to solve with respect to the reliability and validity of their instruments, and they have to deal with sampling problems, such as an increase of nonresponse over time [44, 62]. For instance, the Dutch victimization survey has response percentages of around 32% [2]. Furthermore, police reports include only a selection of victims as victims do not always report an incident to the police [65, 66].

While offline crime is not very easy to measure, measurement problems become even more complicated with online crime [24]. The anonymity of the internet makes it hard to identify offenders, and online crimes are more likely to go unnoticed compared to traditional crimes. For example, data theft might not be detected immediately, making it challenging to measure the true extent of the crime. The hidden nature of specific online crimes adds to these measurement challenges, as they are not as physically visible as traditional crimes. Finally, according to [27], one of the main problems of measuring cybercrime is the relative absence of official data. However, this is not true to the same extent for all online crime.

The present study focuses on estimating the prevalence of ransomware. A ransomware attack is an example of online crime, which involves malicious software that encrypts a victim's data, with the attacker demanding a ransom for the decryption key. In recent years, ransomware has become a significant societal concern [6, 12, 20, 21]. This concern comes, among other things, from the high costs to victims and the significant disruptions to daily life, as exemplified by the Colonial Pipeline incident that led to widespread fuel shortages in the United States [6].

Measuring the prevalence of ransomware attacks is crucial for understanding their impact. There are three primary sources that provide data on ransomware attacks: police reports, incident response companies, and leakpages. Police reports provide information on incidents brought to the attention of law enforcement. Incident response companies offer insights from their operations assisting victims in recovering from ransomware attacks. Leakpages are websites where attackers publish data of victims who do not pay the ransom.

By linking individual victims in these datasets, its combination provides a way to measure ransomware prevalence, taking into account that every dataset in itself might be biased, as described previously. Using this combination we apply capture-recapture methodology, or multiple system estimation (MSE), to compute estimates of the total number of ransomware attacks for large, average, and small businesses [68]. Accordingly, our main research question is:

> **How many ransomware attacks are there in the Netherlands in 2019 - 2022?**

Multiple systems estimation (MSE) is a methodology used in official statistics, particularly with population censuses and administrative data sources. MSE, also known as capture-recapture, is widely used to estimate the size of populations that cannot be completely observed [36]. This method links multiple data sources, or 'lists,' to estimate the number of unobserved cases. By definition, the number of cases that is missed by all lists is unknown. By analyzing the overlap between these lists, it is possible to estimate this number, and once we have this estimate, we can infer the total number of incidents.

The outline of this paper is as follows: in §2 we consider the background literature on traditional crime rate estimation methods and potentially new data sources based on the ransomware crime script. In §3 we present our data and the methodology. Afterwards, §4 presents the results on the amount of ransomware attacks in the Netherlands. In Section §5 we compare our results with the Dutch Victimization Survey of the Statistics Netherlands [10]. Subsequently, we discuss our findings and conclude in §6 and §7, respectively.

2 Background

Having basic information on crime is essential for nation-states. Citizens of developed countries usually have at least some concerns about crime levels in their community [16, 30, 57]. Knowledge about the amount of crime and its characteristics matters to citizens and policymakers. Accordingly, adequate crime statistics are important. A commission of the UK government [61] listed five major reasons why a nation needs crime statistics at a national level:

1. **Government accountability:** To provide reliable quantitative measurements of criminal activity and trends that enable parliament to fulfill its democratic function of holding the government accountable for this aspect of the state of the nation.

2. **Public awareness and research:** To keep the public, media, academia, and relevant special interest groups informed about the state of crime in the country, and to provide (access to) data that informs wider debates and non-governmental research agendas.

3. **Resource allocation:** To inform relevant aspects of short-term resource allocation, both within government and for external related bodies, e.g., for policing and Victim Support.

4. **Performance and accountability:** To inform performance management and accountability at the national level for agencies such as the police.

5. **Strategic policy development:** To provide an evidence base for longer-term government strategic and policy developments [61].

A common measurement tool is victimization (and offender) self-report surveys. Victimization surveys provide a valuable perspective on the level of crime as experienced by the population, capturing incidents that are not reported to or recorded by the police. Victimization (and offender) surveys have been conducted in the Netherlands since 1980 by the Statistics Netherlands (CBS), offering a long-term view of crime trends [2,35]. By sampling private households and asking individuals aged 15 years and older about their experiences with various crimes, victimization surveys can uncover hidden crime figures, especially for offenses that victims may choose not to report to the police.

Since 2017, Statistics Netherlands (CBS) introduced a victimization survey specifically focused on online crime that focused on businesses: the Dutch Cybersecurity Monitor [9]. Data is collected through the annual ICT survey, involving around 20,000 randomly selected companies and 22,000 self-employed individuals. Specific questions about ransomware have been included since 2021. In 2022, Statistics Netherlands reported that 15% of Dutch residents were victims of online crime, with 80% of them not reporting incidents to the police [9]. In 2021, 6,300 ransomware attacks were reported, including 4,000 incidents among self-employed individuals and 2,300 targeting businesses. By 2022, this increased to 8,310 attacks, with 6,000 involving self-employed individuals. Larger companies were disproportionately affected, with 4% of businesses with 250+ employees reporting attacks in 2021, compared to 0.3% of self-employed individuals. This trend continued in 2022, when larger companies were still more affected by ransomware than smaller ones.

Business victimization surveys have the advantage, like victimization surveys of individuals, of measuring crime that is not necessarily reported to the police (see below). However, alongside advantages, business victimization surveys also have problems and issues. The sampling process is complex. For example, who to interview from a large company, how to achieve representation from all economic sectors and companies of different sizes, are issues that need to be satisfactorily resolved [28, 29]. Non-response is a problem with only around 50% of companies participating in the English/Welsh Commercial Victimisation Survey [29, 37]. Also, business victimization surveys are based on information from a single respondent, and the percentage of victimized companies who responded with "don't know" or "no answer" is high (30.8%) [39]. Furthermore, operationalizing the various concepts that make up 'online crime' is not straightforward. There is some overlap with different categories of online crime [39] and respondents may not be aware of the types of online crime and terminology used in the surveys [38].

Another traditional source of crime statistics are police reports. Police reports contain recorded incidents reported to or discovered by law enforcement. In the Netherlands, these records have been systematically collected since 1950, providing a long-term dataset for crime trend analysis [66]. They also provide legally verified information on crimes, making them a reliable source for serious offenses.

Nevertheless, police reports are limited by underreporting, as was mentioned above. This has been shown in surveys of individuals [65, 66] and of businesses [22, 29, 39]. This matters as underreporting is related to crime characteristics such as whether the perpetrator was a known person [59], the type and impact of the incident [39], and fear of reputational damage [1].

Many crimes, especially online crime, go unreported because victims may feel that law enforcement cannot help, or because the crime is not recognized as serious enough to report [52]. Furthermore, few victims report online crime to the police, compared to offline crime [40, 65], although this may be an effect of the type of crime and not a difference between online and offline crime. For example, [65] found a willingness to report of 8-10% of victims of online fraud and [52] found a willingness to report of 2-5% of victims of a particular ransomware variant. Additionally, not all reported crimes are officially recorded due to investigative priorities or legal policies [66]. All these aspects of commercial victimization surveys introduce selection biases into the police data. Furthermore, changes in laws, public awareness campaigns, and administrative practices can influence the consistency and comparability of police data over time. Thus, while useful, police reports are not representative of the mix of crimes experienced by victims [60].

The modus operandi of ransomware may provide potential new data sources to measure the prevalence of ransomware attacks. The modus operandi can be described using a crime script, which breaks down the steps involved in executing an attack [13, 34]. Crime scripts might reveal potential new data sources to measure ransomware incidents. The ransomware crime script [45, 48] includes (1) developing infrastructure and malware, (2) buying ransomware malware from other malicious actors, defined as Ransomware-as-a-Service (RaaS), (3) gaining access via methods like phishing or brute force attacks, (4) moving laterally within the network, (5) exfiltrating sensitive data for extra extortion, (6) encrypting files, (7) communicating with victims for ransom negotiation, (8) deciding on ransom payment, (9) applying blackmail, and (10) laundering ransom and providing decryption keys [11, 26, 31, 32, 42, 43, 47, 55, 56, 58].

This crime script suggests additional methods for measuring ransomware incidents beyond traditional approaches, such as using leak pages where victims are exposed for non-payment, and data from incident response companies that assist with recovery, negotiations, and ransom management. Other potential sources, like negotiation pages, bitcoin payment records, and the market for initial access brokers, are beyond the scope of this paper.

Incident response companies offer valuable insights into ransomware attacks that are often not reported to law enforcement [49, 67]. These companies assist victims in recovering from attacks, negotiating with attackers, and managing ransom payments. However, their data tends to overrepresent larger organizations, as only companies with sufficient financial resources can typically afford these services, leading to a bias in the dataset.

Leak pages, where ransomware groups publish the names or data of victims who refuse to pay the ransom, provide another source of unreported incidents. Monitoring these sites can reveal additional ransomware cases. However, this data is also biased. Not all victims are exposed; attackers may withhold data if a ransom was paid, or may focus on high-profile targets to boost their reputation [50]. Some attackers also lack the resources to publish all cases. As a result, leak pages tend to overrepresent larger companies, further skewing the distribution of reported victims [51].

In the present study, we integrate data from police reports, incident response companies, and leak pages to develop a comprehensive picture of ransomware incidents. By cross-referencing victim names, we can identify which victims appear across multiple datasets and which are unique to a single source. This approach enables us to estimate the number of unobserved ransomware attacks, producing independent estimates that we will compare with the victimization survey of the Statistics Netherlands, the Cybersecurity Monitor, in the discussion section [9].

3 Methodology

3.1 Data

From the study, the population size was based on observations from three datasets between 1 January 2019 and 31 December 2022.

1. **Police Reports (P):** Official reports of ransomware attacks targeting Dutch companies were filed with Dutch Law Enforcement. For a detailed report about the data collection process, we refer to [48,49]. From the 525 attacks, we excluded attacks on individuals and attempted attacks. We included 434 incidents in this study.

2. **Incident Response Companies (I):** Data from an Incident Response company based in the Netherlands, specialized in helping victims recover from ransomware attacks. From the 99 attacks, 30 incidents were outside the Netherlands and therefore left out of the analysis, since we do not know whether they reported to the Police. Since we need to match cases with the other two data sources, this makes it unfeasible to use this data. We included 69 incidents in this study.

3. **Leakpages (L):** Websites where attackers publish stolen data or victim names if the ransom is unpaid. From the 9200 attacks, 9139 attacks were outside the Netherlands and therefore not used in this study. The leakpage dataset was from ecrime.ch and provided to the researchers [14]. We included 61 attacks in this study.

This study aimed to estimate the prevalence of unreported ransomware attacks across different company sizes in the Netherlands, analyzing data from police reports (P), leak page data (L), and incident response data (I), categorized by small (K), medium (M), and large (G) companies. Companies between 1-50 employees are categorized as small, between 51-250 employees as medium, and 251+ employees as large.

A summary of the data is presented in Table 1. Observations were linked by considering company size and victim company name across observations. We considered the probability that two different victims have the same company name and size and are attacked at the same time period to be acceptably small. This procedure led to 477 unique observations.

Table 1: Dataset used for this study. The categories are one-hot encoded and categorized by data source (P, I, L) and company size (S).

P	I	L	S	Frequency
1	0	0	L	30
1	1	0	L	8
0	0	1	L	8
1	0	1	L	8
0	1	1	L	1
1	1	1	L	2
1	0	0	M	48
0	1	0	M	6
1	1	0	M	13
0	0	1	M	7
1	0	1	M	12
1	1	1	M	2
1	0	0	S	293
0	1	0	S	12
1	1	0	S	4
0	0	1	S	15
1	0	1	S	1
0	1	1	S	2
1	1	1	S	5

3.2 Analysis

To estimate the hidden number of ransomware attacks, we employ a method for the estimation of the size of a population known as multiple systems estimation (MSE). We follow the explanation that was provided earlier in [15], for the estimation of homeless. This estimation technique has its origins in biology and refers to the estimation of an unobserved part of a certain population, originally populations of animals. The approach has evolved into a useful technique with applications in epidemiological research and the social sciences. The methodology has proven to be especially useful for estimating hidden populations, such as drug users and homeless people. This method is well-known in statistics, as demonstrated by [5], with applications in public health by [36], homelessness by [17], official statistics by [64], and in human slavery by [17].

MSE of linked administrative sources has the advantage that it is cost-effective for a statistical bureau in need of a national estimate of the number of ransomware attacks. A major advantage is that this approach can deal with incomplete lists, which is an evident problem using registers of ransomware attacks. However, MSE relies on certain assumptions. When linking two sources, the method assumes that the inclusion of a ransomware attack in one source is independent of its inclusion in the other. If more than two sources are linked, this strict independence assumption is relaxed and replaced with the less restrictive condition that no significant k-factor interaction exists across k registers. Additionally, MSE assumes that attacks can be accurately linked across the registers. For this to hold, the registers must contain sufficient and relevant information for linking, and privacy regulations must not impede the process of cross-register matching.

We begin by explaining the method for two lists. This is dual systems estimation. Consider two lists, A and B. By linking these lists, we obtain the following counts: attacks found in A but not in B, attacks found in B but not in A, and attacks recorded in both A and B. These counts form a contingency table, denoted as $A \times B$, where the variable A represents 'inclusion in register A with levels 'yes' and 'no', and similarly for B. In this table, the cell corresponding to 'no, no' (attacks missing from both registers) has a count of zero by definition. The statistical challenge is to estimate this unknown value for the population. To derive an estimate of the total population size, the estimated number of missed attacks is added to the number of attacks observed in at least one of the registers.

The frequency of the missing 'no, no' cell can be estimated by fitting a log-linear model to the incomplete contingency table. Log-linear models express the (logarithm of) observed cell frequencies in terms of main effects and interaction effects of the variables included in the model. To differentiate between various log-linear models, we adopt the notation from [5]. In this notation, interacting variables are enclosed within a single set of square brackets, whereas non-interacting variables are placed in separate sets of square brackets.

For instance, consider a 2×2 contingency table for the registers A and B. The log-linear model $[AB]$ for these two registers is expressed as:

$$\log m_{ab} = \lambda + \lambda_a^A + \lambda_b^B + \lambda_{ab}^{AB}. \tag{1}$$

Here m_{ab} represents the expected frequency of cell a, b, with $a, b =$ 'yes', 'no'. The parameter λ is the intercept, λ_a^A and λ_b^B correspond to the main effects of A and B, respectively, and λ_{ab}^{AB} represents the interaction effect between A and B. The presence of λ_{ab}^{AB} in the model indicates that the probability of being included in A depends on whether the subject is also included in B, and vice versa. This model is referred to as saturated because it includes as many parameters as there are cell frequencies. However, since the cell $m_{no;no}$ is unobserved, the model $[AB]$ contains one parameter too many, making it non-identifiable and, therefore, not estimable.

On the other hand, the independence model $[A][B]$, as given by:

$$\log m_{ab} = \lambda + \lambda_a^A + \lambda_b^B, \tag{2}$$

has only three parameters, and the absence of the interaction parameter λ_{ab}^{AB} indicates that the inclusion probabilities of registers A and B are assumed to be independent. For a 2×2 contingency table with one unobserved cell, the model $[A][B]$ is considered saturated, as it has exactly as many parameters as there are observed cell frequencies. By fitting this model to the three observed cell frequencies, the parameter estimates can be used to derive an estimate for the frequency of the missing 'no, no' cell, and consequently, the total population size.

Independence is a highly restrictive and often unrealistic assumption. To make the model more realistic, we employ two approaches. The first approach involves including covariates, particularly those with levels that exhibit heterogeneous inclusion probabilities across both registers (see [5]). In our study, the covariate 'Size of the company' fulfills this role. For example, by introducing a covariate X, we can extend the two-way contingency table into a three-way contingency table and fit a log-linear model $[AX][BX]$, expressed as:

$$\log m_{abx} = \lambda + \lambda_a^A + \lambda_b^B + \lambda_x^X + \lambda_{ax}^{AX} + \lambda_{bx}^{BX}. \tag{3}$$

Here, the two-factor interaction parameters λ_{ax}^{AX} and λ_{bx}^{BX} represent the interactions between the covariate X and the registers A and B, respectively. The restrictive assumption of independence between A and B is replaced by a less restrictive assumption of conditional independence, given the covariate X. Subpopulation size estimates are then derived for each level of the covariate, and these estimates are summed to obtain the total population size estimate.

The second approach involves including a third register C and analyzing the resulting three-way contingency table using log-linear models that may incorporate one or more two-factor interactions. The saturated model in this case is given by:

$$\log m_{abc} = \lambda + \lambda_a^A + \lambda_b^B + \lambda_c^C + \lambda_{ab}^{AB} + \lambda_{ac}^{AC} + \lambda_{bc}^{BC}. \tag{4}$$

In shorthand notation, this is expressed as $[AB][AC][BC]$. This model allows for pairwise dependence between the registers but does not account for a three-factor interaction, as indicated by the absence of the parameter λ_{abc}^{ABC} . However, including a third register is not always feasible, either because such a register is unavailable or because there is insufficient information to link attacks in the third register to those in the other two registers.

In this study, we have access to both a third register and a covariate, allowing us to significantly relax the assumptions underlying population size estimation. With three registers, we can model pairwise dependencies between the registers by including the interaction terms λ_{ab}^{AB}, λ_{ac}^{AC} and λ_{bc}^{BC}, and test whether these terms are statistically significant. Additionally, the inclusion of a covariate removes the need to assume homogeneity of inclusion probabilities. As mentioned earlier, the use of a covariate also provides valuable insights into the characteristics of individuals who are not captured by any of the registers.

For model selection, we follow a standard approach in log-linear modeling by comparing models based on their relative fit. The relative fit is assessed using the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which are widely used measures for evaluating model performance. Both measures aim to prevent overfitting by penalizing overly complex models, allowing for the comparison of non-nested models. The AIC applies a penalty based on the number of parameters in the model, while the BIC includes an additional penalty that accounts for both the number of parameters and the sample size.

The model with the lowest AIC or BIC value is considered the best fit (for an example in the context of population size estimation, see [3]).

4 Results

A model search is carried out using the BIC, and this leads to the model $[PI][PS][ILS]$. I.e. there is an interaction between P and I, between P and S and between I, L and S. So, controlling for the other variables, in this model there is no direct relation between P and L. The estimated frequencies with 95% confidence intervals for the unobserved cases are presented below:

Table 2: Estimated ransomware incidents under model $[PI][PS][ILS]$

	Observed	Unobserved	Total	Observed	CI	CI
	Cases	Estimated		(%)	2.5	97.5
L	57	80.7	137.7	41.4%	55.1	122.6
M	88	130.7	218.7	40.2%	98.3	190.1
S	332	2373.4	2705.4	12.3%	1272.6	7057.2

For Large and Middle size companies the estimates of unobserved attacks are quite reliable with points estimates 80.7 (CI 55.1 – 122.6) and 130.7 (CI 98.3 – 190.1), but for Small companies the number of unobserved attacks is not reliable, with estimate 2,373.4 (CI 1,272.6 – 7,057.2). Given the large number of observed cases for Small companies, which is 332, we can only conclude that for Small companies the number of ransomware attacks is larger than for Middle and Large companies.

The estimated total number of ransomware attacks for Large and Medium companies is 137.7 and 218.7, respectively, with significant underreporting in both categories. Observed cases totaled 145, while unobserved cases were estimated at 211.4, making the overall total 356.4 attacks. This indicates that 40.7% of ransomware attacks on Large and Medium companies are reported, while 59.3% go unreported. For Large companies, 41.4% of attacks are observed (57 incidents) and 58.6% unobserved (80.7 incidents), while for Medium companies, 40.2% of attacks are observed (88 incidents) and 59.8% are unobserved (130.7 incidents).

We study the model search procedure, in order to have more confidence in this outcome. See Table 3. Model $[PI][PS][ILS]$ has the smallest BIC of 1,638.2. It has 16 parameters. Adding the term PL to the model leads to a higher BIC of 1,640.7, but lowers the AIC. For this model the estimate for Small companies increases considerably, and becomes unrealistically large. For other models adding or deleting terms lead to suboptimal AIC and BIC values.

If we consider model 2 in more detail, by fitting models on the table where we left out the counts for small companies, we find estimates 81 for Large and 130 for Middle Sized companies. We conclude that the estimates for Model 2 found in Table 3 are due to the inclusion of the Small companies, that lead to instability

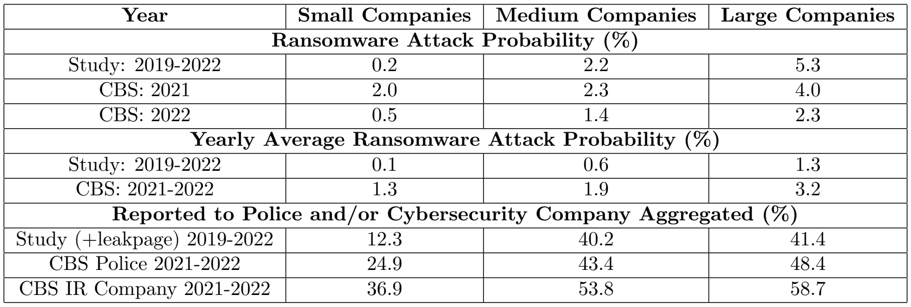
Table 3: Model search using three levels of Size

Model	Logl	pars	AIC	BIC	Large	Middle	Small
1. $[PI][PS][ILS]$	-770.3	16	1572.6	1638.2	81	131	2,373
2. 1. + PL	-768.4	17	1570.9	1640.7	169	274	11,978
3. 2. + PIS	-768.5	18	1573.0	1646.8	87	116	4,725
4. 1. - PI	-774.5	15	1579.0	1640.5	72	111	1,182
5. 4. -PS	-780.9	14	1589.8	1647.2	100	157	751
6. 4. -ILS	-784.3	14	1596.8	1654.0	82	138	1,204

of all estimates. We conclude that we can safely use the estimates in Table 3. In summary, our analysis indicates that a significant number of ransomware attacks remain unobserved through conventional reporting methods.

5 Comparing with Cybersecurity Monitor

In this section, we compare our estimates with a victimization survey from Statistics Netherlands in 2021 and 2022, the Cybersecurity Monitor [9] (see Table 4). Our models estimate that large companies experienced 138 ransomware attacks, while medium-sized companies faced 218 attacks between 2019 and 2022. Combining these estimated number of total ransomware attacks with the number of companies in the Netherlands in 2021 for different company sizes, extrapolated from the Cybersecurity Monitor [9], we calculate the ransomware attack risk for large companies at 5.3% and for medium-sized companies at 2.2% between 2019 and 2022. These figures translate to an average annual risk of 1.3% for large companies and 0.6% for medium companies of becoming a ransomware victim. Although there may be some uncertainty in these estimates due to fluctuations in the number of companies between 2019 and 2022, we believe they reflect the correct order of magnitude. In comparison, the Cybersecurity Monitor reported ransomware attack rates of 4.0% for large companies



Year	Small Companies	Medium Companies	Large Companies
Ransomware Attack Probability (%)			
Study: 2019-2022	0.2	2.2	5.3
CBS: 2021	2.0	2.3	4.0
CBS: 2022	0.5	1.4	2.3
Yearly Average Ransomware Attack Probability (%)			
Study: 2019-2022	0.1	0.6	1.3
CBS: 2021-2022	1.3	1.9	3.2
Reported to Police and/or Cybersecurity Company Aggregated (%)			
Study (+leakpage) 2019-2022	12.3	40.2	41.4
CBS Police 2021-2022	24.9	43.4	48.4
CBS IR Company 2021-2022	36.9	53.8	58.7

Table 4: Ransomware Attacks and Reporting Percentages by Company Size according to the present study and Cybersecurity Monitor of CBS (Statistics Netherlands) [9]

in 2021 and of 2.3% for medium-sized companies, dropping to 2.3% and 1.4%, respectively, in 2022 [9].

Our estimates appear to be relatively lower than those from the Cybersecurity Monitor, which could be due to several factors. First, our analysis focuses on direct victims, excluding indirect victims affected through interdependence of companies. The Statistics Netherlands dataset may include both direct and indirect victims, inflating their numbers. Second, our data does not account for attempted ransomware attacks, which are likely underreported to the police, incident response companies, and leakpages, but may be included in victimization surveys. Lastly, calculation limitations could lead to discrepancies in outcomes; for instance, the exact number of companies per size category is only available for 2021, and we had to extrapolate data for other years. Furthermore, only the percentage of ransomware attacks for 2021 and 2022 are available from CBS.

Despite these limitations, our estimates for the risk of ransomware attacks fall within the confidence intervals (CI) of our study (Table 4). Specifically, the CBS estimate for large companies (4.0% in 2021 and 2.3% in 2022) aligns with our CI of 2.1% to 4.7%. For medium-sized companies, CBS estimates (2.3% in 2021 and 1.4% in 2022) fall within our CI of 1.0% to 1.9%. For small companies, CBS estimates (2.0% in 2021 and 0.5% in 2022) are consistent with our CI of 0.8% to 4.6%. This alignment suggests that both CBS and our estimates provide reliable estimates of risk of ransomware attacks, demonstrating the robustness of our findings.

6 Discussion

The present study estimates the total number of ransomware attacks on businesses in the Netherlands between 2019 and 2022. According to our estimates, 138 large companies, 219 medium companies, and 2706 small companies suffered from a ransomware attack, suffered from a ransomware attack. While the estimates for large and medium companies are reliable, those for small companies carry high uncertainty due to wide confidence intervals. As a result, we present the findings for large, medium, and small companies separately, acknowledging the limitations for small companies. Based on our estimates, we calculated that there is an annual risk of 1.3% for large companies and 0.6% for medium companies of suffering a ransomware attack. This is in line with previous figures of the Cybersecurity Monitor published by Statistics Netherlands in 2021 and 2022 [9].

Our analysis shows significant underreporting of incidents to the police across all company sizes. For large companies, about 41.4% of attacks are observed, while 58.6% go unreported. Similarly, 40.2% of medium-sized company attacks are captured, leaving 59.8% unobserved. However, it should be noted that about 40% of attacks reported to the police, incident response company and/or leakpage, is considerably more than police reporting of online crime in general, like online fraud. Previous research found police reporting rates for online fraud of 11.5% in the UK [54], 14% in the US [53], 13.4% in Portugal [23], and in the

Netherlands, percentages ranging from 11.8% [40] to 13 and 14% [65].

One reason for higher reporting rates in our findings compared to prior research, might be the more severe impact of ransomware attacks on medium and large companies [48]. Serious online crimes are generally reported more often, as supported by prior research [49, 52]. For instance, Deadbolt ransomware, which primarily targets individuals and small businesses, had low reporting rates of 2.8% to 5.1% [52]. Smaller companies may choose not to report due to lower perceived financial loss or other factors. In contrast, larger companies are more likely to report ransomware attacks, potentially due to operational impacts or insurance requirements [49].

The estimated percentage of ransomware attacks observed (or reported) in our study aligns with the Cybersecurity Monitor's reporting figures (see Table 4). According to the Cybersecurity Monitor, 37% of companies with two or more employees sought help from cybersecurity firms after an attack, while only 18% reported the incident to the police, with reporting rates decreasing for smaller businesses. These percentages are close to the 40% observed in our dataset from the three data sources. This is noteworthy given the limitations of our data, such as relying on only one incident response (IR) company, while the Cybersecurity Monitor includes victims who used any cybersecurity or IR service. Despite these limitations, the consistency between the datasets highlights the robustness of our findings.

Finally, our study has several other limitations that affect the generalizability of our findings. Firstly, the willingness of victims to report ransomware attacks to the police may vary across countries due to cultural and moral differences. Since this study focused only on the Netherlands, the estimates may differ when using data from other countries. The representation of victims on leak pages might also vary internationally, influenced by differing tendencies to pay ransoms. Additionally, our study is based on data from a single incident response company, which may not be representative of the broader industry. Finally, as mentioned before, we do not include data on individuals who become victim of ransomware, attempted ransomware and indirect victims. These numbers would provide a more reliable estimation of the victimization of ransomware.

Despite these limitations, we believe our results are significant for several reasons. Firstly, our methodology allows us to extract valuable information from multiple data sources and understand the interaction between these sources. Secondly, while the exact figures may vary, we expect the general trend of higher underreporting rates among small companies to hold true across different contexts. This is likely due to small companies being less represented in various data sources compared to medium and large companies. However, this hypothesis needs to be tested in follow-up research.

7 Conclusion

This study highlights the importance of using multiple data sources to measure the full scope of ransomware attacks. To answer our main research question:

How many ransomware attacks occurred in the Netherlands between 2019 and 2022?, we applied the capture-recapture methodology. Our analyses indicate that, for large companies, 57 (41.4%) ransomware attacks were reported, with 80.7 (58.6%) of the attacks unobserved. For medium-sized companies, 88 (40.2%) ransomware attacks were reported, with 130.7 (59.8%) of the attacks unobserved. Overall, 137.7 large companies, 218.7 medium companies, and 2705.4 small companies suffered from a ransomware attack. We noted that the estimate small companies is unreliable. The average annual risk of a ransomware attack is 1.3% for large companies and 0.6% for mid-sized companies.

Our results align closely with the Statistics Netherlands Cybersecurity Monitor [9]. This has several implications: First, the results are robust, as we obtain similar estimates using independent methods. Second, our approach may be more cost-efficient than a large-scale victimization survey, making it preferable for exploratory research or to reduce costs.

Future research should focus on small businesses, where uncertainty in our estimates remains high due to wide confidence intervals. The uncertainty could be reduced if more of the attacks reported to the police were also detected by Incident Response Companies and on Leakpages, increasing the overlap between sources. However, it is unclear how this can be achieved. Small companies often lack the resources to address cybersecurity threats and may underreport attacks due to perceived insignificance, resource limitations, or unawareness of reporting mechanisms. There is also a belief that police may not take small companies as seriously as larger ones, resulting in fewer police reports. Many small businesses cannot afford incident response services, further reducing detection. Offenders may also avoid posting small firms on leak pages to maintain their reputation. This underreporting suggests many ransomware incidents go undetected, highlighting the need for additional datasets of ransomware targeting small businesses. However, estimates for medium and large companies are encouraging, as higher-than-expected reporting rates implies a more accurate picture of ransomware than previously assumed.

References

[1] Abhishta, A., van Rijswijk-Deij, R., & Nieuwenhuis, L. J. (2019). Measuring the impact of a successful DDoS attack on the customer behaviour of managed DNS service providers. *ACM SIGCOMM Computer Communication Review, 48*(5), 70-76.

[2] Akkermans, M., Kloosterman, R., Moons, E., Reep, C., & Aa, M. T.-v. d. (2022). Veiligheidsmonitor 2021 (the safety monitor 2021) (pp. 99). Retrieved from https://www.cbs.nl/-/media/pdf/2022/09/veiligheidsmonitor.pdf

[3] Anderson, D. R., & Burnham, K. P. (2002). Avoiding pitfalls when using information-theoretic methods. The Journal of Wildlife Management, 912-918.

[4] Beerman, J., Berent, D., Falter, Z., & Bhunia, S. (2023, May). A review of colonial pipeline ransomware attack. In 2023 IEEE/ACM 23rd International Symposium on Cluster, Cloud and Internet Computing Workshops (CCGridW) (pp. 8-15). IEEE.

[5] Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). Discrete multivariate analysis: Theory and practice. New York: McGraw-Hill.

[6] Blatchly, J. (2023). The Impact of Ransomware–A Comparison of Worldwide Governmental Policies and Recommendations for Future Directives (Doctoral dissertation, Utica University).

[7] Cantor, D., & Lynch, J. P. (2000). Self-report surveys as measures of crime and criminal victimization. Criminal Justice & Behavior, 4, 85-138.

[8] Centraal Bureau voor de Statistiek. (2023). *Online veiligheid en criminaliteit 2022*. Retrieved 19 June, 2024, from `https://www.cbs.nl/nl-nl/publicatie/2023/19/online-veiligheid-en-criminaliteit-2022`.

[9] Centraal Bureau voor de Statistiek. (2023). *Cybersecurity Monitor 2023*, Statistics Netherlands, 2023. Retrieved 23 August, 2024, from `https://www.cbs.nl/nl-nl/longread/rapportages/2024/cybersecuritymonitor-2023`, Accessed: August 2023.

[10] Centraal Bureau voor de Statistiek. (CBS). *Aantal bedrijven naar grootteklasse*. 2024. Retrieved 23 August, 2024, from `https://www.cbs.nl/nl-nl/cijfers/detail/81588NED`.

[11] Check Point Research (2022). Leaks of Conti Ransomware Group Paint Picture of a Surprisingly Normal Tech Start-Up. Retrieved August 31, 2022, from https://research.checkpoint.com/2022/leaks-of-conti-ransomware-group-paint-picture-of-a-surprisingly-normal-tech-start-up-sort-of/

[12] Connolly, A. Y., & Borrion, H. (2022). Reducing ransomware crime: analysis of victims' payment decisions. Computers & Security, 119, 102760.

[13] Cornish, D.B. (1994). "The procedural analysis of offending and its relevance for situational prevention," *Crime Prevention Studies*, vol. 3, no. 1, pp. 151–196, 1994.

[14] Cosin, C. (2022). Ecrime. Retrieved March 1, 2023, from https://ecrime.ch/

[15] Coumans, R.M., Cruyff, M.J.L.F., van der Heijden, P.G.M., & Schmeets, H. (2017). Estimating homelessness in the Netherlands using a capture-recapture approach. Social Indicators Research, 130, 189-212.

[16] Crawford, T. A. M., & Evans, K. (2016). Crime prevention and community safety. In A. Leibling, S. Maruna, & L. McAra (Eds.), Oxford Handbook of Criminology (6 ed.). Oxford, UK: Oxford University Press.

[17] Cruyff, M.J.L.F., van Dijk, J., & van der Heijden, P.G.M. (2017). The challenge of counting victims of human trafficking. Chance, 30, 41-49.

[18] Daigle, L. E., Snyder, J. A., & Fisher, B. S. (2016). Measuring victimization: Issues and new directions. In B. M. Huebner & T. S. Bynum (Eds.), The handbook of measurement issues in criminology and criminal justice (pp. 249-276). West Sussex, UK: John Wiley & Sons, Inc.

[19] De Haan, W.J.M. (1997). 't kon minder: geweldscriminaliteit, leefbaarheid en kwaliteit van gezondheidszorg. Deventer: Gouda Quint.

[20] Europol. (2021). Internet Organised Crime Threat Assessment (IOCTA) 2021. Luxembourg: Publications Office of the European Union. Retrieved August 31, 2022, from https://www.europol.europa.eu/publications-events/main-reports/internet-organised-crime-threat-assessment-iocta-2021

[21] Europol. (2023). Internet Organised Crime Threat Assessment (IOCTA) 2023. Luxembourg: Publications Office of the European Union. Retrieved August 31, 2023, from https://www.europol.europa.eu/publication-events/main-reports/internet-organised-crime-assessment-iocta-2023

[22] Flatley, J. (2023). Crime against businesses: findings from the 2022 Commercial Victimisation Survey. London, UK. Retrieved from `https://www.gov.uk/government/statistics/crime-against-businesses-findings-from-the-2023-commercial-victimisation-survey/`

[23] Fonseca, C., Moreira, S., & Guedes, I. (2022). Online consumer fraud victimization and reporting: A quantitative study of the predictors and motives. *Victims & Offenders*, 17(5), 756-780.

[24] Gibbon, J., Marjanov, T., Hutchings, A., & Aston, J. (2024, July). Measuring the Unmeasurable: Estimating True Population of Hidden Online Communities. In *2024 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (pp. 56-66). IEEE.

[25] Hagerty, M.R. et al. (2001). Quality of life indexes for national policy: review and agenda for research. Social Indicators Research, 55, 1-96.

[26] Hernandez-Castro, J., Cartwright, A., & Cartwright, E. (2020). An economic analysis of ransomware and its welfare consequences. Royal Society open science, 7(3), 190023.

[27] Holt, T. J. (2016). Cybercrime. In B. M. Huebner & T. S. Bynum (Eds.), The handbook of measurement issues in criminology and criminal justice (pp. 29-48). West Sussex, UK: John Wiley & Sons, Inc.

[28] Hopkins, M. (2016a). Business, victimisation and victimology: Reflections on contemporary patterns of commercial victimisation and the concept of businesses as 'ideal victims'. *International Review of Victimology, 22*(2), 161-178.

[29] Hopkins, M. (2016b). The crime drop and the changing face of commercial victimization: Reflections on the 'commercial crime drop' in the UK and the implications for future research. *Criminology Criminal Justice, 16*(4), 410-430.

[30] Hough, M., & Robert, J. V. (2007). Public Opinion and Criminal Justice: The British Crime Survey and Beyond. In J. Hough & M. Maxfield (Eds.), Surveying crime in the 21st century: commemorating the 25th anniversary of the British crime survey (Vol. 22, pp. 197-220). Monsey, NY: Criminal Justice Press.

[31] Huang, D.Y., Aliapoulios, M.M., Li, V.G., Invernizzi, L., McRoberts, K., Bursztein, E., Levin, J., Levchenko, K., Snoeren, A.C., & McCoy, D. (2018). Tracking Ransomware End-to-end. 39th IEEE Symposium on Security and Privacy, S & P, pp. 618-631.

[32] Huang, K., Siegel, M., & Madnic, S. (2018). Systematically Understanding the Cyber Attack Business: A Survey. ACM Comput. Surv. 51, 4, Article 70 (July 2019), 36 pages. https://doi.org/10.1145/3199674.

[33] Huls, F.W.M. et al. (2001). Criminaliteit en rechtshandhaving 2000: ontwikkelingen en samenhangen. Wetenschappelijk Onderzoek- en Documentatiecentrum/Centraal Bureau voor de Statistiek, Den Haag/Voorburg.

[34] Hutchings, A., & Holt, T.J. (2015). A crime script analysis of the online stolen data market. *British Journal of Criminology*, vol. 55, no. 3, pp. 596–614.

[35] Huys, H. & Rooduijn, J. (1994). A new survey on justice and security. Netherlands Official Statistics, 9, 47-51.

[36] International Working Group for Disease Monitoring and Forecasting (1995). Capture–recapture and multiple record systems estimation. Part I. History and theoretical development. American Journal of Epidemiology, 1995, 142, 1059-1068.

[37] IPSOS. (2023). Crime against businesses: findings from the 2023 Commercial Victimisation Survey. Technical Report 2023. Retrieved from London, UK: `https://assets-uk.ipsos.com/pa/cvs/2023/cvstechnicalreport.pdf`

[38] Junger, M., & Hartel, P. (2022). Crime Survey & Cybercrime: The State-of-the-art. In M. F. Aebi, S. Caneppele, & L. Molnar (Eds.), Measuring cybercrime in the time of Covid-19: the role of crime and criminal justice statistics. Proceedings of the conference 29-30 October 2020 (Version: 25.12.2021) (pp. 75-88). Strassbourg, France (online): Eleven Publisher, European Union and the Council of Europe.

[39] Kemp, S., Buil-Gil, D., Miró-Llinares, F., & Lord, N. (2021). When do businesses report cybercrime? Findings from a UK study. *Criminology and Criminal Justice*.

[40] Koning, L., Junger, M., & Veldkamp, B. P. (in press). Reporting fraud victimization to the police: factors that affect why victims do not report. Psychology, Crime and Law.

[41] Krohn, M. D., Thornberry, T. P., Gibson, C. L., & Baldwin, J. M. (2010). The development and impact of self-report measures of crime and delinquency. Journal of Quantitative Criminology, 26(4), 509-525.

[42] Leo, P., Işik, Ö., & Muhly, F. (2022). The Ransomware Dilemma. Mit Sloan Management Review MIT Sloan Management Review.

[43] Li, Z., & Liao, Q. (2021). Game theory of data-selling ransomware. Journal of Cyber Security and Mobility, 65-96.

[44] Luiten, A., Hox, J., & de Leeuw, E. (2020). Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys. Journal of Official Statistics, 36(3), 469-487.

[45] Matthijsse, S. R., van 't Hoff-de Goede, M. S., & Leukfeldt, E. R. (2023). Your files have been encrypted: A crime script analysis of ransomware attacks. Trends in Organized Crime, 1-27.

[46] Maxfield, M. G., & Babbie, E. R. (2010). Research methods for criminal justice and criminology: CengageBrain.com.

[47] Meland, P. H., Bayoumy, Y. F. F., & Sindre, G. (2020). The Ransomware-as-a-Service economy within the darknet. Computers & Security, 92, 101762.

[48] Meurs, T., Junger, M., Tews, E., & Abhishta, A. (2022). Ransomware: How attacker's effort, victim characteristics and context influence ransom requested, payment and financial loss. In 2022 APWG symposium on electronic crime research (eCrime) (pp. 1-13). IEEE.

[49] Meurs, T., Cartwright, E., Cartwright, A., Junger, M., Hoheisel, R., Tews, E., & Abhishta, A. (2023). Ransomware economics: A two-step approach to model ransom paid. In *2023 APWG Symposium on Electronic Crime Research (eCrime)*, IEEE, pp. 1–13.

[50] Meurs, T., Cartwright, E., Cartwright, A., Junger, M., & Abhishta, A. (2024). Deception in double extortion ransomware attacks: An analysis of profitability and credibility. *Computers & Security*, vol. 138, p. 103670. Elsevier.

[51] Meurs, T., Hoheisel, R., Junger, M., Abhishta, A., & McCoy, D. (2024). What To Do Against Ransomware? Evaluating Law Enforcement Interventions. In *2024 APWG Symposium on Electronic Crime Research (eCrime)*, IEEE, pp. 1–13.

[52] Meurs, T., Hoheisel, R., Junger, M., & Abhishta, A. (in press). NAS ransomware: Ransomware Targeting NAS Devices. *Proceedings of the Human Factors in Cybercrime Conference 2024*.

[53] Morgan, R. E. (2021). *Financial Fraud in the United States, 2017*. (NCJ 255817). Washington DC: Bureau of Justice Statistics. Retrieved from `https://bjs.ojp.gov/redirect-legacy/content/pub/pdf/ffus17.pdf`

[54] ONS. (2022). *Nature of fraud and computer misuse in England and Wales: year ending March 2022*. Retrieved from `https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/natureoffraudandcomputermisuseinenglandandwales/yearendingmarch2022`

[55] Oosthoek, K., Cable, J., & Smaragdakis, G. (2022). A Tale of Two Markets: Investigating the Ransomware Payments Economy. arXiv preprint arXiv:2205.05028.

[56] Oz, H., Aris, A., Levi, A., & Uluagac, A. S. (2021). A survey on ransomware: Evolution, taxonomy, and defense solutions. ACM Computing Surveys (CSUR).

[57] Pain, R. (2000). Place, social relations and the fear of crime: a review. Progress in Human Geography, 24(3), 365-387.

[58] Payne, B., & Mienie, E. (2021). Multiple-Extortion Ransomware: The Case for Active Cyber Threat Intelligence. In ECCWS 2021 20th European Conference on Cyber Warfare and Security (p. 331). Academic Conferences Inter Ltd.

[59] Shahbazov, I., Afandiyev, Z., & Balayeva, A. (2023). Some determinants of crime reporting among economic and financial crime victims: The case of Azerbaijan. *Journal of White Collar and Corporate Crime, 4*(1), 24-37.

[60] Skogan, W. G. (1984). Reporting crimes to the police: The status of world research. *Journal of Research in Crime and Delinquency, 21*(2), 113-137.

[61] Smith, A. (2006). Crime statistics: An independent review (Carried out for the Secretary of State for the Home Department). London, UK. Retrieved from `http:/rds.homeoffice.gov.uk/rds/pdfs06/crime-statistics-independent-review-06.pdf`

[62] Stoop, I. A. L. (2005). The hunt for the last respondent. Nonresponse in sample surveys. The Hague, The Netherlands: Social and Cultural Planning Office.

[63] Thornberry, T. P., & Krohn, M. D. (2000). The Self-Report Method for Measuring Delinquency and Crime. In D. Duffee (Ed.), Measurement and Analysis of Crime and Justice (Vol. 4, pp. 33-84). Washington, DC: National Institute of Justice.

[64] Van der Heijden, P. G. M., Cruyff, M., Smith, P.A., Bycroft, P., Graham, P, & Matheson-Dunning, N. (2022). Multiple system estimation using covariates having missing values and measurement error: estimating the size of the Māori population in New Zealand. Journal of the Royal Statistical Society, Series A, 185, 156-177.

[65] Van de Weijer, S. G., Leukfeldt, R., & Bernasco, W. (2019). Determinants of reporting cybercrime: A comparison between identity theft, consumer fraud, and hacking. European Journal of Criminology, 16(4), 486-508.

[66] Wittebrood, K., & Junger, M. (2002). Trends in violent crime: A comparison between police statistics and victimization surveys. Social Indicators Research, 59(2), 153-173.

[67] Woods, D. W., Böhme, R., Wolff, J. & Schwarcz, D. (2023) "Lessons lost: Incident response in the age of cyber insurance and breach attorneys," in *32nd USENIX Security Symposium (USENIX Security 23)*, pp. 2259–2273.

[68] Zhang, L.-C., & Dunne, J. (2018). Trimmed dual system estimation. In Böhning, D., Van der Heijden, P. & Bunge, J. (Eds.), Capture–recapture methods for the social and medical sciences (pp. 229-235). Boca Raton: CRC Press.