EFFECTIVENESS OF INFORMATION-SECURITY AWARENESS TRAINING TO PREVENT SUCCESS OF SOCIAL ENGINEERING IN HEALTHCARE: A META-ANALYSIS

The University of Twente
Enschede, The Netherlands
Anica Strating – a.j.strating@student.utwente.nl

Abstract

Creating information awareness is particularly important today. The great increase in digitalization over the past five years since the COVID-19 pandemic, increased the number of successful cyber-attacks in companies significantly, leading to excessive costs, especially in healthcare. To counter this, it is important that healthcare employees protect themselves against this type of crime. This quantitative study combines a systematic review method with a meta-analysis to express the current effectiveness of interventions to combat cybercrime and identify the main building blocks of these interventions. This meta-analysis yielded a moderate effect size, and the most effective building blocks are dynamic content, warnings and high intensity of training.

Keywords: Cybersecurity, Social engineering, Awareness training, Effectiveness, Healthcare.

Samenvatting

Het creëren van bewust omgaan met informatiebeveiliging is erg belangrijk in de huige samenleving. Door de grote toename van digitalisatie in de afgelopen vijf jaar, sinds de COVID-19 pandemie, is het aantal succesvolle cyber-aanvallen in bedrijven sterk toegenomen en dit leidt tot hoge kosten, met name in de zorg. Om dit tegen te gaan is het van belang dat zorgpersoneel beter wordt beschermd tegen deze vorm van criminaliteit. Dit kwantitatieve onderzoek combineert een systematische review methode met een meta-analyse, om te kunnen uitdrukken wat de huidige effectiviteit van interventies om cybercriminaliteit tegen te gaan is en wat de belangrijkste bouwstenen zijn van deze interventies. Deze meta-analyse leverde een gemiddelde effectgrootte op ten aanzien van de effectiviteit van training om cyber-aanvallen tegen te gaan. Hiernaast zijn de belangrijkste bouwstenen van de afgelopen 5 jaar dynamiek, waarschuwingen en intenstiteit.

1. Introduction

Currently, the internet represents the predominant medium for communication and information dissemination. This increases the risk of cybercrime victimization, in particular of social engineering [1]. The concept of "social engineering" encompasses a range of non-technical methods, including psychological elements, aimed at exploiting human vulnerabilities to obtain personal information and undermine security systems [2]. Because of the human element being the weakest link in information-security, experts in this domain classify social engineering as significant threats for companies [3, 4].

Companies suffer from financial losses resulting from data breaches caused by cybercrime in the past years, particularly the healthcare industry. Research shows that between 2015 and 2020 1,485 data breaches occurred affecting 141,252,797 medical records [5]. The term "data breach" denotes a security incident characterized by unauthorized individuals copying, hacking, viewing, or accessing sensitive data [6]. Sixty percent of these breaches are a result of social engineering [5]. To decrease the risks of data breaches, proactive innovations such as awareness-trainings are designed to assist users in the recognition of social engineering attacks [7].

Educating users in information-security awareness entails equipping them with skills to defend against social engineering [7]. The combination of using human vulnerability and obscurity to deceive someone, results in

difficulties to defend against [8]. Methodologies for innovations to combat social engineering range from didactic lectures and interactive workshops to the application of gaming techniques for educational purposes [9]. Bullée and Junger conducted a meta-analysis examining the effectiveness of interventions to combat social engineering up to the year of 2017 [10]. Their findings revealed that interventions aimed at reducing victimization from social engineering resulted in a statistically significant moderate effect. Moreover, it examines the influence of various training methodologies and demographic factors. This research serves as a follow-up to their research, aiming to provide enhanced insights into the effectiveness of interventions of the previous five years since the COVID-19 pandemic. This research seeks to offer a comprehensive analysis of the subject by employing a meta-analysis methodology.

2. Problem Statement

In recent years, the COVID-19 pandemic has highlighted the critical importance of digital transformation within the healthcare sector. Strategic and cost-effective adoption of digital healthcare technologies can significantly improve the quality of care delivered, and enhance the overall well-being of the population [11]. Recent developments, such as eHealth, can significantly improve the quality of life for a lot of people in our society. As digitalization becomes more integrated into healthcare practices, sensitive medical data is increasingly vulnerable to cyberattacks [12]. However, this involves challenges for the healthcare industry in information-security, because eHealth accelerates the exchange of personal information on digital devices [13]. This is why the healthcare sector faces an increasing threat from cyberattacks worldwide [14].

Healthcare organizations also suffer from higher costs as a result of these cyberattacks as compared to other sectors [5]. To adequately protect healthcare employees against cybercrime, particularly social engineering, it is crucial to compare as many outcomes from recent previous studies about information-security training as possible. This approach allows for a deeper understanding of measures that are effective in combating social engineering and reduce the risk of data breaches in the healthcare industry.

3. Research questions

This research aims to find an answer to two main research questions:

1. *How effective are information-security awareness trainings to prevent success of social engineering in the past 5 years?*
2. *What characteristics should an information-security awareness training for employees in healthcare have to reduce the risk of a data breach in the Electronic Health Record?*

To answer these questions the following sub-questions have been composed:

- *How do various types of interventions differ in their effectiveness at mitigating social engineering attacks?*
- *What impact do intervention characteristics have on the effectiveness of information-security awareness training?*
- *How do sociodemographic factors of a population affect the effectiveness of information-security training in the past 5 years?*

4. Theoretical framework

This section describes the theoretical background of this study according to literature.

4.1. Social Engineering

Social engineering refers to the use of non-technical techniques, including psychological elements, to exploit human vulnerabilities to obtain personal information and undermine security systems [2].

Phishing email and ransomware

An example of a social engineering technique is phishing email. Phishing emails target individuals, as they represent the vulnerability within the information system. A phishing email looks like a genuine email from a

company with which the victim has a relationship, for example the target's bank. This email contains a malicious URL (link). When the victim clicks the malicious URL, the social engineer achieves access to the personal data of the victim [15]. Upon clicking on these URLS, ransomware may also disseminate throughout a hospital network, rendering the Electronic Health Record (EHR) inaccessible. Given that this constitutes the primary cause of ransomware attacks, it is imperative to ensure that all healthcare personnel undergo cybersecurity training [16].

Data Breaches in healthcare

Health data breaches, which include loss, theft, unauthorized access, and hacking incidents, are often attributed to errors or negligence by hospital employees managing the data. Hospital data breaches can block the access or undermine the accuracy of patient information available to healthcare providers via the EHR [6]. To enhance the security of the EHR it is important, particularly for healthcare personnel, to protect themselves against social engineering. These breaches increase the risk of providers unintentionally accessing or altering the patient's medical information. Inaccuracies or delays can disrupt the care process and adversely affect patient outcomes. In extreme scenarios, are for example hospital employees limited to using the primitive means of communication. This leads to complication of patient care and reduces the quality of care [17].

4.2. Healthcare industry

To give appropriate advice about the training characteristics to successfully combat social engineering in the healthcare industry and answering RQ2, theoretical background about healthcare employees is necessary.

Employees in healthcare

The work environment in the healthcare industry is a critical determinant of an organization's success, influencing both employee performance and patient outcomes. It must be supportive and dynamic to enable healthcare organizations to meet their goals and achieve high productivity. A positive work environment directly impacts the safety, quality of care, and overall health outcomes for patients [18]. Elements that should be included in this environment are for example: internal work motivation, multidisciplinary collaboration, personal and professional development, and teamwork [19]. A well-structured and encouraging work environment benefits not only patients but also the broader healthcare organization, ensuring that all parties thrive in this complex and evolving field [18].

4.3. Information-security interventions

This section describes the building blocks of current interventions focused on information-security awareness that underpins this research. These building blocks are the foundation for the classification of variables in this study. The coding of the variables is described in section 5.2..

Characteristics of interventions

According to Bullée and Junger, the methods used to combat social engineering can be categorized into three types: static, spoken or dynamic [10]. These groups describe the modality of the intervention. The modality of the intervention tells something about the devices used to facilitate awareness at the participants in the training.

Static content: A document (for example pdf) which contains the learning material for the intervention without further explanations.

Spoken content: A spoken intervention means to have contact with users during a lecture for example [9]. In a lecture, the trainer explains rule-based training content to the participants.

Dynamic content: There are several types of dynamic interventions, two examples of dynamic interventions are games and mindfulness.

a. *Games*
   Gaming methods incorporate interactive scenarios and challenges to educate

participants on effectively identifying and responding to phishing attempts.

b. *Mindfulness*
   Mindfulness training encourages individuals to pause and consider the context of requests, engage in active questioning when evaluating emails for suspicious elements, seek advice from trusted sources, and gather evidence before making decisions [20].

The following characteristics of interventions influence the success of the training in Bullee and Junger's research [9].

Priming: Priming is a technique to indirectly warn users for danger. Priming can activate knowledge, without being conscious about it.

Warnings: Warnings prevent the user from immediately viewing the content of for example an email. There are 2 types of warnings: passive and active warnings. The difference between passive and active warnings is that active warnings completely hide the content of the malicious email, while passive warnings only tell the user that the link in the email is malicious.

Tips: Giving tips about for example the look of a malicious URL, helps people recognize URLs which lead to phishing websites.

Focus: According to literature, trainings with specific focus on for example the look of malicious URLs are reported to have a higher effect than general cybersecurity awareness-trainings [9].

Intensity: The intensity of the training can be expressed in the amount of effort needed for the subject to complete of the intervention.

- Low, only tips or warnings for example.
- Medium, contains a document including learning material.
- High, a game which includes a lecture or a mindfulness intervention for example.

According to literature, the higher the intensity of the training, the higher the effectiveness [10].

Research context

The context of research in combating cybercrime could have effect on the outcomes of the effectiveness of interventions. For this research, there are six key factors which can influence the outcomes of the interventions; type of social engineering, experimental design, pre-victimization application, the awareness of being tested, the environment of the population and if research is done in the healthcare section.

Experimental design: In this research, studies using pre-/post-training experimental design and intervention-/control-group experimental design are both included. In pre-/post-training design, the population is the same. In intervention-/control-group design, the populations in both groups could differ from each other. Therefore, it is important to look if the type of experimental design influences the effect sizes in the meta-analysis.

Type of social engineering: In section 4.2. several types of social engineering are described. The type of social engineering is based on the type of the device used to contact users. This could be face-to-face, email, phone, SMS or a website.

Time: The duration of an experiment could have influence on the performance of the participants. Participants can maintain attention on a task for a limited time consecutively before fatigue occurs and affects performance [21]. The time it takes to complete the training will be examined in this research.

Percentage phishing mail: There are many opportunities in the design of the experiment to test participants' phishing mail detection ability [22]. For example, only phishing mails could be sent to the participants, or it could be that some of the mails sent are phishing and some are legitimate [23]. Finally, it is possible to send only legitimate mails in the experiment [24].

Pre-victimization: To conduct research in the section cybercrime, pre-victimization is applied in interventions often. Pre-victimization influences the outcome of the effectiveness of the intervention. Pre-

victimization is called ‘embedded training’. Participants receive a social engineering attack for example a phishing e-mail before they have had training. This is performed to measure the baseline of the population performance. Afterwards, only participants which became victim, clicked the malicious link in the e-mail for example, qualify for training in the second phase [10].

Awareness: When people are aware of participating in an experiment, they are more likely to behave differently towards people who are unaware of participation. This may cause the Hawthorne-effect to occur [25].

Environment: People could perform different behavior when they are in laboratory environments towards field environment. In field research, people receive for example a mock-phishing email in their personal inbox. In laboratory experiments, is the behavior of the participant observed in a controlled environment [10]. For example, the participant has to detect phishing emails in a simulated environment [26].

Healthcare: The significance of cybersecurity extends beyond the healthcare sector and still the healthcare industry has shown a comparatively slower and inadequate response in safeguarding stakeholders' data [27]. For this research, it is interesting to look for differences in effects when training healthcare employees towards training employees in different sectors.

5. Methodology

This section describes the methodology used in this research. This is a quantitative study in which a systematic review is employed to gather and synthesize all relevant empirical evidence meeting the predetermined eligibility criteria, as utilized in previous research by Bullée and Junger [10]. Subsequently, a meta-analysis is conducted to consolidate the findings of these studies.

5.1. Data Collection

The Scopus database is used to obtain studies for the analysis in the first phase of the data collection; the identification. A query is composed which includes all the important characteristics that a study must have to possibly be included in this research. The query below is used to collect data from Scopus. This query returned 517 studies.

TITLE-ABS-KEY ("social engineering" OR "socialengineering" OR "social-engineering" OR phishing OR (phishing AND mail*) AND ((cybercrime OR disclosure OR victim* OR malicious) OR (prevent* OR reduc* OR combat OR detect*))) AND experiment* AND (survey* OR warning* OR interven* OR game* OR mindfulness OR workshop* OR solution* OR training* OR countermeasure) AND PUBYEAR > 2018 AND NOT TITLE-ABS-KEY ("neural network*" OR "deep learning" OR "machine learning" OR "AI" OR "artifical intelligence") AND (LIMIT-TO (PUBSTAGE,"final") AND (LIMIT-TO (PUBSTAGE,"ar") OR LIMIT-TO (PUBSTAGE,"re")) AND (LIMIT-TO (LANGUAGE,"English") OR LIMIT-TO (LANGUAGE,"Dutch"))*

Not all the studies will contain sufficient data suitable for inclusion in the meta-analysis. To select the studies which contain data suitable for answering the research questions, nine eligibility criteria are composed. The eligibility criteria are based on criteria in previous research of Bullée and Junger [10]. Criterium 9 is added to exclude every study published before 2019 as this research aims to analyze data from 2019 (the start of the COVID-19 pandemic) till April 2024.

Eligibility Criteria:

1. To be a published scientific paper or a PhD thesis;
2. The manuscript must be written in English or Dutch;
3. The study should involve human subjects;
4. An experimental design should be used, questionnaires or surveys that only measure, e.g. attitude or intention are excluded; it is of particular interest to observe how the subjects behave in the context of social engineering;
5. The experiment (and intervention) should aim to reduce victimisation by social engineering; there should be deception or a malicious part be involved;
6. There should be a comparison of at least two groups, i.e.: a control and training or awareness group; or a pre-training and

post-training group; the comparison of groups is required to state the effectiveness of an intervention;
7. No technical solutions (e.g. an algorithm that filters possible phishing emails); this analysis is about human behavior in social engineering; therefore, exclusively technical solutions are excluded;
8. There should be at least 20 observations per group; this was chosen to have sufficient strength in the analysis and reduce the possibility of the observations based on random chance;
9. The publication date is after 2018.

To report the studies for the meta-analysis, the PRISMA guideline is used [28]. PRISMA stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Figure 1 in the appendix shows the PRISMA flow-diagram of the data collection.

In the first phase, identification, studies are exported from Scopus to excel, and the duplicates are removed. Secondly, the articles are screened for eligibility. During this phase, eight studies are excluded based on criterium one and two. 468 Studies in total are excluded based on the other eligibility criteria see figure 2.

Lastly, the studies included in the screening phase are read thoroughly to extract the data which is necessary to conduct the meta-analysis and formulate an answer to the research questions. The data which will be extracted relates to the following topics:

1. General information about the study (author, year of publication and the language of the study);
2. Raw data about the experiment (number of clicks or the ClickRate (CR) in the intervention group and control group during the pre- and post-training)
3. Specific research context (type of experiment, type of social engineering, pre-victimization used, the time the experiment takes to complete, the quantity of phishing mails send to the participants, awareness of participants, environment of the experiment and population in healthcare sector);
4. Intervention characteristics (method of the intervention, specific focused, priming applied, warnings used, tips used and the intensity of the intervention); and
5. The last category includes the socio-demographics. These variables have been coded to analyze their possible influence on the effectiveness of training. Socio-demographic factors (population type, mean age of the population, proportion men/women and country).

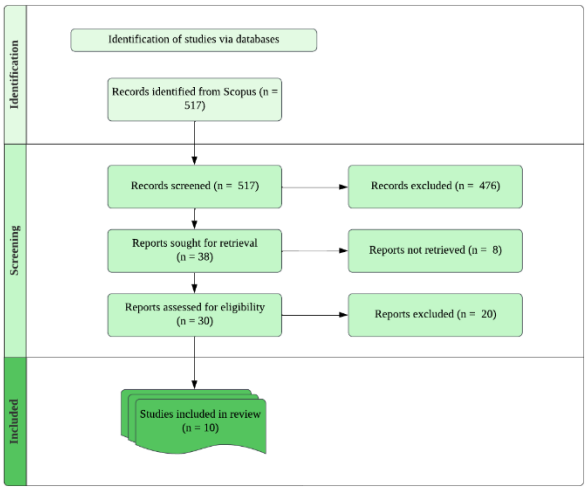

Figure 1 PRISMA flow diagram

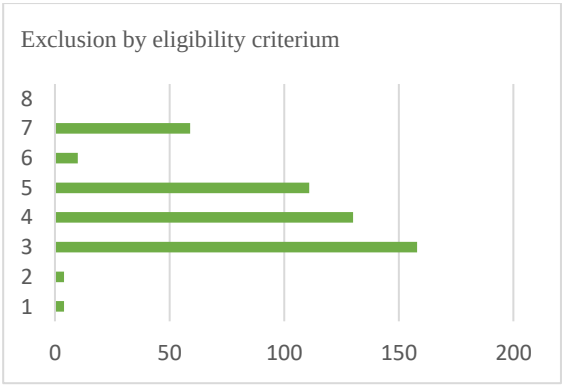

Figure 2 Excluded studies

5.2. Data-Analysis

This section describes the plan for meta-analysis and explains the coding of all variables analyzed in the data-analysis. The data-analysis is performed using the R-Studio program. R version 4.3.3 for Windows consists of a general package for meta-analysis, metafor, which is used to compute the statistical analysis [29]. Also, the packages

tidyverse, dplyr and escalc (effect size computation) are used for the analysis.

5.2.1. Meta-analysis

For the data-analysis, a meta-analysis is conducted. Meta-analysis integrates findings from multiple studies to compute an aggregate effect size, while employing subgroup-analysis to investigate potential sources of heterogeneity or variation in training program effectiveness. A random-effects model is applied in the meta-analysis due to its acknowledgement that the included studies are drawn from a broader population of potential studies [30]. This could cause true effect sizes to differ across the studies. For the subgroup-analysis a mixed-effects model is used to define differences in effect sizes of the variables extracted from the data. The mixed-effects model is necessary, because the effects between the subgroups are random and the data from the subgroups itself are considered to be fixed [29].

The effect sizes both in the meta- as the subgroup-analysis will be calculated as the standardized mean difference (SMD), in literature known as Cohen's d [29]. Cohen's d is defined as the difference between two means divided by the pooled standard deviation [31]. According to Cohen, a standardized mean difference of 0.20 would be considered small, 0.50 would be considered moderate and 0.80 would be considered large [31].

Following this, an examination is conducted to assess the extent of variability in the effects, known as heterogeneity. The variability between studies can be expressed in the Q-statistic, I-squared (percentage of variance in observed effects reflecting the variance in true effects rather than sampling error), the Tau-squared (variance of true effects), the Tau (standard deviation of true effects) and the R-squared (the proportion of the heterogeneity caused by the moderator) [32]. From an I-squared of forty percent, the variability between studies is substantial. Subgroup-analysis is important in this situation to recognize if there are true effects due to variables (moderators) or if the effect arises only from variability between the studies.

5.2.2. Power analysis

Before the meta-analysis, a power analysis is performed in R to check what number of included studies should be sufficient to detect meaningful effects [29]. Table 1 shows the parameters which are estimated a priori.

Table 1 Estimated parameters for power analysis

PARAMETER	ESTIMATE
μ	0.54
$\boldsymbol{\tau^2}$	0.20
k	10
α	0.05

These estimates are based on previous research [10]. The power for these estimates and inclusion of ten studies (k=10) is calculated (Power=0.939). This means that there is an 93.9% chance that the meta-analysis will detect a true existing medium effect (μ=0.54).

5.2.3. Coding variables

All variables described in section 5.1. are categorized and coded. The coding of all variables can be found in the appendix, Table A1.

6. Results

This section describes the results of the meta- and subgroup-analysis which is conducted to answer the research questions. All effect sizes, corresponding Q-statistics and R-squared values for the overall results and subgroups are explained.

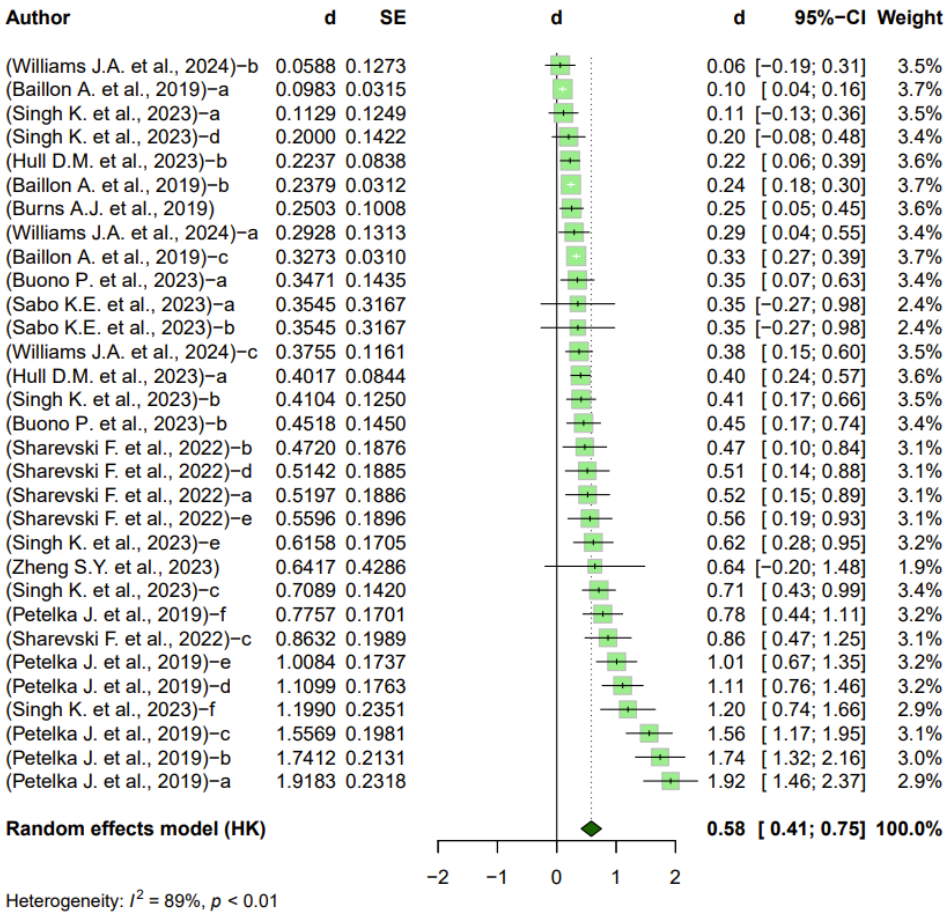

Figure 3 Forest plot of the meta-analysis with ten included studies and 31 effect sizes

6.1. Overall results

The meta-analysis included 10 studies (k=10) [20, 22-24, 26, 33-37] and 31 effect sizes (N=31). The total number of participants (n) of this meta-analysis is 14.087 and the mean age of the participants 34.0 years from birth. There are 5 studies published in 2023. In 2022 and in 2024 is both one study published and there are three studies published in 2019. There are 3 studies originating from Europe [24, 26, 37] and the other 7 studies are originating from the United States. There are 4 conference papers [24, 26, 34, 35] and 6 articles included.

The overall effect-size d=0.582 (95% CI [0.411; 0.753], t =6.96, p<.0001). The overall effect-size of this meta-analysis can be considered moderate according to Cohen [38]. The I-squared is 89.2% (95% CI [85.8%; 91.8%]). This indicates high heterogeneity across studies. Cochran's Q is calculated (Q=277.63) and this value is significant (p<.0001). This means that 89.2% of the variance in the estimated effect sizes are due to true differences between studies and not random error. Figure 3 provides a forest plot containing an overview of the effects in this meta-analysis.

6.2. Subgroup-analysis

The results of the subgroup-analysis can be found in Table A2 in Appendix A.

6.2.1. Research context

Email is the only type of social engineering used in the interventions which are included in the dataset. Therefore, no differences in effect sizes between the types of social engineering are found in this subgroup-analysis.

Experimental design

The dataset contains two studies, and nine effect sizes, which uses a pre-/post-training as experimental design [22, 33]. The effect size of pre-/post-training experimental design is negative d=-0.224 (p=.210). The effect size of

intervention/control design can be considered moderate d=0.649 (p<.0001). This indicates that the experimental design chosen could influence the observed outcome of the effectiveness of training. Cautious interpretation is necessary because the Q-statistic of this subgroup is not statistically significant (Q=1.571, p=.210).

Awareness

In seven studies, and nineteen effect sizes, participants knew they are participating in an experiment [20, 23, 26, 34-37]. The effect size of participants being aware can be considered small d=0.322 (p=.043). The effect size of participants being unaware is also considered small d=0.385 (p=.002). The Q-statistic of this subgroup is statistically significant (Q=4.082, p=.043). This indicates that there is a small difference in the effect of training between aware and unaware participants. If participants are unaware of being trained, the effectiveness of the training is slightly higher. However, the R-squared is 12.12, which means that this moderator can only explain approximately twelve percent of the heterogeneity.

Pre-victimization

One study pre-victimized the participants [36]. This resulted in one effect size. The effect of pre-victimization is negative d=-0.344 (p=.444). This indicates that pre-victimization does not contribute to a higher effectiveness of training. Cautious interpretation is necessary because the Q-statistic of this subgroup is not statistically significant (Q= 0.585, p=.444).

Environment

Four studies took place in field environment, resulting in thirteen effect sizes [22, 33, 36, 37]. Field environment results in a small effect size d= 0.357 (p=.001). The laboratory environment also results in a small effect d= 0.403 (p=.007). The Q-statistic of this subgroup is statistically significant (Q=7.370, p=.007). This indicates that training could be measured slightly more effective when the experiment is laboratory. However, the R-squared is 22.71. This means that only approximately twenty-three percent of the heterogeneity can be explained by this moderator.

Healthcare

The participants of one study were healthcare employees, this resulted in three effect sizes [33]. The effect size of healthcare is negative d=-0.378 (p=.156) which indicates that training could be less effective while the participants are healthcare employees. However, cautious interpretation is necessary because the Q-statistic of this subgroup is not statistically significant (Q=2.008, p=.156).

Phishing mail

The included studies send different percentages of phishing mails to the participants. For example, Singh K. et al [22] only send phishing mails to the participants to test their ability to detect phishing, while other studies send partly phishing mails and partly legitimate emails [23, 24, 34]. The effect size of the proportion of phishing mails send to the participants is negative d=-1.014 (p<.0001). The Q-statistic of this subgroup is statistically significant (Q=21.715, p<.0001) and the R-squared is 56.80, which means that a substantial amount of the heterogeneity can be explained by this moderator. This indicates that a higher the proportion of phishing mails send to the participants, results in less effective training.

Time

The effect size of long time (25+ minutes) to complete the experiment can be considered moderate d=0.500 (p=.003). The effect size of moderate time (16-25 minutes) to complete the experiment is also moderate d=0.553 (p=.019). The effect size of a short experiment time is negative d=-0.101 (p=.785). The Q-statistic of this subgroup is statistically significant (Q=6.594, p=.037). This indicates that experiments lasting longer than fifteen minutes have a moderate positive influence on the overall effectiveness of a training. However, the R-squared is 20.86, which means that this moderator can only explain approximately twenty-one percent of the heterogeneity.

6.2.2. Intervention characteristics

There were no studies included with the intervention type spoken, Therefore, no effect size for spoken interventions is found in this subgroup-analysis.

Intervention type

Four included studies used static content as intervention, this resulted in six effect sizes [33, 34, 36, 37]. The static content intervention type has a small negative effect size d=-0.460 (p=.014). This indicates no improvement and worse performance after training when participants receive static content as device. In contrast to static content, the effect size of a dynamic intervention type can be considered moderate d=0.674 (p<.0001). The Q-statistic is statistically significant (Q=6.095, p=.014). This indicates that dynamic interventions have a higher effectiveness. However, the R-squared is 17.88, which means that only approximately eighteen percent of the heterogeneity can be explained by this moderator.

Priming

One study applied priming in the intervention and this resulted in six effect sizes [22]. Use of priming has a negative effect size d=-0.075 (p=.721). The effect size of no priming applied can be considered moderate d=0.598 (p<.0001). This indicates that application of priming has a negative influence on the effectiveness of training. Cautious interpretation is necessary because the Q-statistic of this subgroup is not statistically significant (Q= 0.128, p=.721).

Warning

Two studies used warnings in the intervention and this resulted in eight effect sizes [26, 35]. The effect size of using warnings can be considered moderate d=0.670 (p<.0001). The effect size of no warning used can be considered small d=0.340 (p<.0001). The Q-statistic of this subgroup is statistically significant (Q=21.167, p<.0001) and the R-squared is 47.78, which means that a substantial amount of the heterogeneity can be explained by this moderator. This indicates that use of warnings in the intervention has a positive influence on the effectiveness of training.

Tips

Eight studies used tips in the intervention [20, 22-24, 33, 34, 36, 37]. This resulted in twenty effect sizes. Use of tips has a negative effect size d=-0.477 (p=.002). The effect size of not using tips can be considered large d=0.882 (p<.0001). The Q-statistic of this subgroup is statistically significant (Q=10.086, p=.002). This indicates that receiving tips decreases the effectiveness of training. However, the R-squared is 27.43 which means that only approximately twenty-seven percent of the heterogeneity can be explained by this moderator.

Specific focus

All the included studies experimented with at least one intervention with specific focus, this resulted in thirty effect sizes [22, 24, 26, 33-37]. The effect size of specific focus can be considered moderate d=0.542 (p=.228). A general training had almost no effect on the improvement d= 0.059 (p=.894). This indicates that specific focus increases the effectiveness of the training. Cautious interpretation is necessary because the Q-statistic of this subgroup is not statistically significant (Q=1.454, p=.228).

Intensity

There are three included studies that used a highly intensive intervention [20, 24, 33, 34]. The effect size of high intensity can be considered small d=0.360 (p=.062). Five studies applied a low intensity intervention [22, 26, 33, 35, 37]. The effect size of the low intensity can also be considered small d=0.326 (p=.149). Four studies applied a middle intensity intervention [20, 22, 23, 36]. The effect size of a middle intensity intervention is very small d=0.187 (p=.442). This indicates that to improve the effectiveness, a training should be high or low intensive. Middle intensity does hardly improve the effectiveness of the training.

6.2.3. Socio-demographics

None of the effect sizes of the socio-demographics have a statistically significant Q-statistic. This means that cautious interpretation of the effect sizes is necessary.

The effect size of differences in population resulted in a moderate effect for employees d= 0.596 (p<.0001). The effect size of the category students is negative d=-0.066 (p=.750). This indicates that the effectiveness of training is higher given among employees. The effect size of sample size is negligible (d=-0.000, p=.162) and the effect sizes of all categories of education can be considered small. Only middle level of education has a slightly higher effect size relative to the other levels (d=0.434, p=.120). Looking at the effect sizes for differences in gender, the proportion men have a large effect on the training outcome d=1.272 (p=.230). The proportion women have a negative effect size d=-1.247 (p=.239). It could be possible that training is less effective for women because literature indicates that the processing of information could differ between gender [39]. The effect of age is negligible (d=-0.012). Lastly, the effect size of country can be considered moderate if participants are residing in the continent of America d=0.645 (p<.0001). The effect of the participants residing in the continent of Europe is negative (d=-0.324, p=.105).

7. Conclusion

This section provides answers to the research questions starting with the first main research question, next the sub-questions and last the second main research question including the advice.

7.1. RQ1 – Effectiveness of information-security awareness training

The effectiveness of training employees to combat social engineering is moderate given an overall effect size of d=0.582. This indicates that the overall effectiveness of information-security awareness training from 2019 till now can be considered moderate and there is still room for improvement of training content. The fact that the 95%-CI ([0.411, 0.753]) interval does not contain a negative effect size indicates that it is ninety-five percent certain that the true effect size in the population is positive and training is effective. Cautious interpretation of this effect is important due to the high percentage of heterogeneity (89.20%) in this meta-analysis. Although moderators can partly explain this heterogeneity, there is still unexplained heterogeneity left. However, the relatively small number of included studies (k=10) makes it more difficult to detect heterogeneity. Including more studies may ensure that subtle heterogeneity is better detected. As a result, effect sizes caused for variables can also be better distinguished from effect sizes explained only by heterogeneity.

7.1.1. Sub-Q1 – Intervention type

The intervention type which is mostly effective is the dynamic content (d=0.674, p<.0001). This effect is statistically significant, which means that this finding supports the use of dynamic content as an effective measure in information-security awareness trainings.

7.1.2. Sub-Q2 – Intervention characteristics

First, according to the results of the subgroup-analysis, interventions with specific focus (d=0.542), warnings (d=0.670) and high intensity (d=0.360) have the largest impact. Second, the usage of priming in the intervention does not have impact at the effectiveness of the training. Using tips has a negative influence on the effectiveness of training. In short, an information-security awareness training should at least have a high intensity and include interactive elements or warnings to create a dynamic environment between user and trainer.

7.1.3. Sub-Q3 – Socio-demographics

In this research, employees moderately affect the effectiveness of training (d=0.596). Also, participants from America affect the outcomes of training moderately (d=0.645). Other socio-demographic factors do not affect the effectiveness of an information-security

awareness training. However, the effects in these subgroups are not statistically significant, which means that there is no certainty that these results are true effects in the population.

7.2. RQ2 – Training healthcare employees

According to the subgroup analysis of the variable healthcare, the effectiveness of training when the experiment is taking place in healthcare industry is negative (d=-0.378). This suggests that the environment of the healthcare industry could make it harder for an information-security training to be effective. Therefore, it is important to create a highly intensive and interactive training which specifically focuses on one subject or includes warnings. An example of an intervention in literature which meets almost all of these characteristics is the adversarial mindset training, in Zheng. S.Y. et al.[24]. The training includes a short video (5 minutes) in which an actor dressed as a stereotypical cybercriminal demonstrates how to set up phishing emails. Afterwards participants must detect phishing mails in different scenarios.

8. Discussion

This section describes reflection on the findings relative to relevant literature, the limitations of this research and future research.

8.1. Reflection on literature

First, the overall effect of this research (d=0.58) is larger relative to previous research (d=0.54) Bullée and Junger [10]. Comparing the results of this research, a similar trend emerges from the characteristics of training which are effective. Dynamic content has the largest effect on the performance of participants after training. However, a large proportion of the dataset consisted of dynamic interventions, which could cause the effect size to be affected by this. Also, no spoken intervention was present in the dataset, so it was not possible to compare effects with it. Usage of tips and priming is an exception to the trend. This research declares that tips have negative effect on the overall effectiveness of the training. Since the single effects of intervention characteristics are not compared to the effects of the combination of different intervention characteristics in this research, could it be possible that this effect originates from combined characteristics. Priming also deviates from the trend, as in this study priming has a negligible effect on the effectiveness of training (d=-0.075). This could be explained by the small number of studies in the dataset (k=10). Only one study used priming [22], therefore this effect is almost impossible to compare.

Second, the type of training which is advised to be effective for healthcare employees must be cautious interpret. According to literature, to make a training suitable for the work environment of healthcare personnel it is important to consider elements regarding to internal work motivation, multidisciplinary collaboration, personal and professional development, and teamwork [19]. From this, it follows that the employees must be intrinsically motivated to participate in the training. Further research is necessary to investigate the willingness to participate in this training.

8.2. Study limitations

The limitations of this study include time pressure, availability, and validity.

First, time pressure is the largest limitation of this study. Due to time pressure, only the Scopus database is used to collect studies. If more time is available, other databases could also be searched. This resulted in the inclusion of only ten studies. When the quantity of studies within a subgroup is limited, such as when $k \leq 5$ (in some subgroups in this research) the assessment of R-squared tends to be less precise and this creates a highly uncertain estimate of between-study heterogeneity within the subgroups [30]. However, the a priori power analysis simulated ten studies (k=10). This means that the current study has enough power to obtain 93.9% chance that the detection of the moderate effect (d=0.582) is truly existing.

Second, availability of information limits this study. There was only one study included about information-security training in the healthcare industry [33]. Next to lack of availability, most data about socio-demographics in healthcare is confidential and therefore unpublished. This prevented this research from a proper analysis of effects from socio-demographic data in healthcare.

Last, the extent of 'researcher degrees of freedom' can impact both the internal and external validity of research, as well as its reliability [40]. For instance, researchers may include or exclude certain studies based on undisclosed personal preferences, thereby potentially compromising the validity of the findings. To resolve this, a second observation from another researcher is recommended. Similarly, various methodological decisions and interpretation of outcome measures can lead to disparate results and interpretations, thus diminishing the reliability of the research. For example, the selection of the studies is done by screening on eligibility criteria. The researcher compiles these criteria.

8.3. Future research

This research has not compared the single effects of intervention characteristics to the effects of the combination of different intervention characteristics. For future research, it would be interesting to look at the effects of combining different intervention characteristics. Maybe when a single variable turned out to have a small effect on the effectiveness of the training now, it could have a high effect when combined with other variables.

Also, an outcome of the subgroup-analysis is that employees in the healthcare industry have a negative effect on the effectiveness of information-security awareness training. It would be interesting to figure out how this came to be, because the possibility exists that it is necessary to solve this problem first before information-security awareness training can be more effective for healthcare employees in the future. In conclusion, the advice to improve information-security awareness in healthcare is as follows. Investigate the underlying reasons why the healthcare industry has worse performance in information-security towards other sections and implement dynamic information-security awareness training for employees to make the organization more secure.

Bibliography

1. Krombholz K, Hobel H, Huber M, Weippl E. Advanced social engineering attacks. Journal of Information Security and Applications. 2015;22:113-22. doi: 10.1016/j.jisa.2014.09.005.

2. Fadhil HS. Social engineering attack techniques. International journal of progressive research in engineering management and science. 2023;03.

3. Sumner A, Yuan X, Anwar M, McBride M. Examining Factors Impacting the Effectiveness of Anti-Phishing Trainings. Journal of Computer Information Systems. 2021;62(5):975-97. doi: 10.1080/08874417.2021.1955638.

4. Singh K, Aggarwal P, Rajivan P, Gonzalez C. What makes phishing emails hard for humans to detect? Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2021;64(1):431-5. doi: 10.1177/1071181320641097.

5. Yeo LH, Banfield J. Human Factors in Electronic Health Records Cybersecurity Breach: An Exploratory Analysis. Perspect Health Inf Manag. 2022;19(Spring):1i.

6. Parkavi R, Jeya Iswarya MR, Kirithika G, Madhumitha M, Varsha O. Data Breach in the Healthcare System. Handbook of Research on Data Science and Cybersecurity Innovations in Industry 40 Technologies. Advances in Information Security, Privacy, and Ethics, 2023. Chapter chapter 21. p. 418-34.doi: 10.4018/978-1-6684-8145-5.ch021.

7. Ahlan AR, Lubis M, Lubis AR. Information Security Awareness at the Knowledge-Based Institution: Its Antecedents and Measures. Procedia Computer Science. 2015;72:361-73. doi: 10.1016/j.procs.2015.12.151.

8. Yasin A, Fatima R, Liu L, Wang J, Ali R, Wei Z. Understanding and deciphering of social engineering attack scenarios. Security and Privacy. 2021;4(4). doi: 10.1002/spy2.161.

9. Bullée J-W, Junger M. Social engineering: digitale fraude en misleiding. Justitiële verkenningen. 2020;46(2):92-110. doi: 10.5553/jv/016758502020046002009.

10. Bullee J-W, Junger M. How effective are social engineering interventions? A meta-analysis. Information & Computer Security. 2020;28(5):801-30. doi: 10.1108/ics-07-2019-0078.

11. Raimo N, De Turi I, Albergo F, Vitolla F. The drivers of the digital transformation in the healthcare industry: An empirical analysis in Italian hospitals. Technovation. 2023;121. doi: 10.1016/j.technovation.2022.102558.

12. Tariq MU. Enhancing Cybersecurity Protocols in Modern Healthcare Systems. Transformative Approaches to Patient Literacy and Healthcare Innovation. Advances in Healthcare Information Systems and Administration, 2024. Chapter chapter 11. p. 223-41.doi: 10.4018/979-8-3693-3661-8.ch011.

13. Ostern N, Perscheid G, Reelitz C, Moormann J. Keeping pace with the healthcare transformation: a literature review and research agenda for a new decade of health information systems research. Electron Mark. 2021;31(4):901-21. doi: 10.1007/s12525-021-00484-1.

14. Keogh RJ, Harvey H, Brady C, Hassett E, Costelloe SJ, O'Sullivan MJ, et al. Dealing with digital paralysis: Surviving a cyberattack in a National Cancer center. J Cancer Policy. 2024;39:100466. doi: 10.1016/j.jcpo.2023.100466.

15. Daengsi T, Pornpongtechavanich P, Wuttidittachotti P. Cybersecurity Awareness Enhancement: A Study of the Effects of Age and Gender of Thai Employees Associated with Phishing Attacks. Educ Inf Technol (Dordr). 2022;27(4):4729-52. doi: 10.1007/s10639-021-10806-7.

16. IBM: Cyber Resilient Organization Study. 2021 Accessed. Available from: https://www.ibm.com/resources/guides/cyber-resilient-organization-study/.

17. Johnson SCME. Do Hospital Data Breaches Reduce Patient Care Quality? , 2019.doi: 10.48550/1904.02058.

18. Al-Ghwary AA, Al-Oweidat IA, Al-Qudimat AR, Abu Shosha GM, Khalifeh AH, M AL. The Impact of Work Environment on Structural Empowerment among Nurses in Governmental Hospitals. Nurs Rep. 2024;14(1):482-93. doi: 10.3390/nursrep14010037.

19. Maassen SM, van Oostveen C, Vermeulen H, Weggelaar AM. Defining a positive work environment for hospital healthcare professionals: A Delphi study. PLoS One. 2021;16(2):e0247530. doi: 10.1371/journal.pone.0247530.

20. Hull DM, Schuetz SW, Lowry PB. Tell me a story: The effects that narratives exert on meaningful-engagement outcomes in antiphishing training. Computers & Security. 2023;129. doi: 10.1016/j.cose.2023.103252.

21. Seet MS, Bezerianos, A. . Neuroscience of Cognitive Functions: From Theory to Applications. Handbook of Neuroengineering. Singapore: Springer; 2023. doi:doi.org/10.1007/978-981-16-5540-1_73.

22. Singh K, Aggarwal P, Rajivan P, Gonzalez C. Cognitive elements of learning and discriminability in anti-phishing training. Computers & Security. 2023;127. doi: 10.1016/j.cose.2023.103105.

23. Sharevski F, Jachim P. "Alexa, What's a Phishing Email?": Training users to spot phishing emails using a voice assistant. EURASIP J Inf Secur. 2022;2022(1):7. doi: 10.1186/s13635-022-00133-w.

24. Zheng SY, Becker I. Phishing to improve detection. Proceedings of the 2023 European Symposium on Usable Security2023. p. 334-43.

25. Parsons HM. What Happened at Hawthorne?: New evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. Science. 1974;183(4128):922-32. doi: 10.1126/science.183.4128.922.

26. Buono P, Desolda G, Greco F, Piccinno A. Let warnings interrupt the interaction and explain: designing and evaluating phishing email warnings. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems2023. p. 1-6.

27. Nifakos S, Chandramouli K, Nikolaou CK, Papachristou P, Koch S, Panaousis E, et al. Influence of Human Factors on Cyber Security within Healthcare Organisations: A Systematic Review. Sensors (Basel). 2021;21(15). doi: 10.3390/s21155119.

28. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi: 10.1136/bmj.n71.

29. Harrer M, Cuijpers, P., Furukawa, T.A., & Ebert, D.D. Doing Meta-Analysis with R: A Hands-On Guide. Boca Raton, FL and London: Chapmann & Hall/CRC Press; 2021.

30. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods. 2010;1(2):97-111. doi: 10.1002/jrsm.12.

31. Cohen J. Statistical power analysis for the behavioral sciences. 2013.

32. Higgins JPT TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Chapter 10: Analysing data and undertaking meta-analyses. In: Deeks JJ HJ, Altman DG. , editor. Cochrane Handbook for Systematic Reviews of Interventions version 64 (updated August 2023). 2023.

33. Williams JA, Zafar H, Gupta S. Fortifying healthcare: An action research approach to developing an effective SETA program. Computers & Security. 2024;138. doi: 10.1016/j.cose.2023.103655.

34. D.M. SKEBJS. Developing IMPAWSTER: Improving Meaningful Phishing Awareness With Simulated Training and Email Roleplay. 2023.

35. Petelka J, Zou Y, Schaub F. Put Your Warning Where Your Link Is. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems2019. p. 1-15.

36. Burns AJ, Johnson ME, Caputo DD. Spear phishing in a barrel: Insights from a targeted phishing campaign. Journal of Organizational Computing and Electronic Commerce. 2019;29(1):24-39. doi: 10.1080/10919392.2019.1552745.

37. Baillon A, de Bruin J, Emirmahmutoglu A, van de Veer E, van Dijk B. Informing, simulating experience, or both: A field experiment on phishing risks. PLoS One. 2019;14(12):e0224216. doi: 10.1371/journal.pone.0224216.

38. LeCroy CW, Krysik J. Understanding and Interpreting Effect Size Measures. Social Work Research. 2007;31(4):243-8. doi: 10.1093/swr/31.4.243.

39. Lee YY, Gan CL, Liew TW. Susceptibility to instant messaging phishing attacks: does systematic information processing differ between genders? Crime Prevention and Community Safety. 2023;25(2):179-203. doi: 10.1057/s41300-023-00176-2.

40. Wicherts JM, Veldkamp CL, Augusteijn HE, Bakker M, van Aert RC, van Assen MA. Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. Front Psychol. 2016;7:1832. doi: 10.3389/fpsyg.2016.01832.

APPENDIX

Table A1 Coding of variables

CATEGORY	VARIABLE	CODING
Research context	Type of social engineering used (categorical)	Face-to-face, Email, Phone, SMS and Website.
	Pre-victimization (dichotome)	Yes (1), only vicitms are trained. No (0), all received training.
	Awareness (dichotome)	Yes (1), participants were aware of being trained; and No (0), participants were unaware of being trained.
	Environment (categorical)	Laboratory, participants are trained in controlled simulation. Field, participants receive content in their personal inbox.
	Healthcare (dichotome)	Yes (1), participants are healthcare employees; and: No (0), participants are not healthcare employees.
	Time (numeric)	The duration to complete the experiment in minutes.
	%-Phish (numeric)	The percentage phishing mails of all the emails sent to the participant in decimal numbers.
Intervention characteristics	Intervention type (categorical)	Static content Rule-based Dynamic
	Priming (dichotome)	Yes (1), priming is used. No (0), no priming is used.
	Warning (dichotome)	Yes (1), warnings are used. No (0), no warnings are used.
	Tips (dichotome)	Yes (1), tips are used. No (0), no tips are used.
	Specific focus (dichotome)	Yes (1), the training has one specific focus, for example recognition of malicious URLs. No (0), the training was general.
	Intensity (categorical)	Low, less effort for the participants. Middle, middle effort for the participants. High, high effort for the participants.
Socio-demographics	Population (categorical)	Employees (research conducted on people working for an organization). Students (research conducted on students at high school or University). Elderly people (people in retirement).
	Sample size (numeric)	The number of individuals participating in the study.
	Gender (numeric)	Proportion of men and women in the research population in decimal numbers.
	Age (numeric)	The mean age of the research population in terms of years after birth.
	Country (categorical)	Participants residing in the continent of America; Participants residing in the continent of Europe, Asia, Australia or Africa.

Table A2 Average effect size for sub-group analysis of 10 studies in the meta-analysis

Characteristic	Effect (d)	SE	z-value	p-value	CI.lb	CI.ub	Q	p-value	I-squared(%)	R-squared(%)
All	0.582				0.411	0.753	277.63	0.0001	89.20	
Research context										
Experimental design							1.571	0.210	95.51	
Pre-/post	-0.224	0.178	-1.254	0.210	-0.573	0.126				
Intervention/control	0.649	0.098	6.623	<.0001	0.457	0.841				
Awareness							4.082	**0.043**	94.09	12.12
Yes	0.322	0.160	2.021	0.043	0.010	0.635				
No	0.385	0.124	3.105	0.002	0.142	0.628				
Pre-victimization							0.585	0.444	95.72	
Yes	-0.344	0.450	-0.765	0.444	-1.227	0.538				
No	0.595	0.085	7.006	<.0001	0.428	0.761				
Environment							7.370	**0.007**	94.27	22.71
Field	0.357	0.109	3.267	0.001	0.143	0.572				
Laboratory	0.403	0.148	2.715	0.007	0.112	0.694				
Healthcare							2.008	0.156	95.51	
Yes	-0.378	0.267	-1.417	0.156	-0.900	0.145				
No	0.621	0.086	7.220	<.0001	0.452	0.789				
%-Phishing	-1.014	0.218	-4.660	<.0001	-1.441	-0.588	21.715	**0.000**	92.16	56.80
Time							6.594	**0.037**	88.96	20.86
Short	-0.101	0.367	-0.273	0.785	-0.826	0.624				
Moderate	0.553	0.236	2.337	0.019	0.089	1.016				
Long	0.500	0.166	3.015	0.003	0.175	0.826				
Intervention characteristics										
Intervention type							6.095	**0.014**	94.44	17.88
Static content	-0.460	0.186	-2.469	0.014	-0.826	-0.095				
Dynamic content	0.674	0.085	7.905	<.0001	0.507	0.841				
Priming							0.128	0.721	95.80	
Yes	-0.075	0.211	-0.357	0.721	-0.488	0.338				
No	0.598	0.094	6.376	<.0001	0.415	0.781				
Warning							21.167	**0.000**	92.02	47.78
Yes	0.670	0.146	4.601	<.0001	0.385	0.955				
No	0.340	0.073	5.509	<.0001	0.257	0.542				
Tips							10.086	**0.002**	94.06	27.43
Yes	-0.477	0.150	-3.176	0.002	-0.772	-0.183				
No	0.882	0.121	7.316	<.0001	0.646	1.119				
Specific focus							1.454	0.228	95.61	
Yes	0.542	0.449	1.206	0.228	-0.339	1.422				
No	0.059	0.441	0.133	0.894	-0.806	0.924				
Intensity							2.169	0.338	95.57	
Low	0.326	0.225	1.445	0.149	-0.116	0.767				
Middle	0.187	0.243	0.769	0.442	-0.290	0.664				
High	0.360	0.193	1.866	0.062	-0.018	0.738				
Socio-demographics										
Age	-0.012	0.014	-0.849	0.396	-0.040	0.016	0.721	0.396	95.34	
Country							2.627	0.105	95.04	

America	0.645	0.090	7.211	<.0001	0.470	0.821			
Europe	-0.325	0.201	-1.621	0.105	-0.718	0.068			
Education							2.792	0.248	95.90
Low	0.255	0.303	0.841	0.400	-0.339	0.849			
Middle	0.434	0.279	1.553	0.120	-0.113	0.981			
High	0.243	0.258	0.941	0.347	-0.263	0.749			
Gender									
Men	1.272	1.058	1.202	0.230	-0.802	3.346	1.444	0.230	95.71
Women	-1.247	1.058	-1.179	0.239	-3.321	0.827	1.389	0.239	95.72
Population							0.102	0.750	95.86
Employees	0.596	0.094	6.325	<.0001	0.412	0.781			
Students	-0.066	0.208	-0.319	0.750	-0.474	0.341			
Sample size	-0.000	0.000	-1.400	0.162	-0.000	0.000	1.960	0.162	95.14