Phishing Susceptibility and the (In-)Effectiveness of Common Anti-Phishing Interventions in a Large University Hospital

Jan Tolsdorf
The George Washington University
Washington, D.C., USA
jan.tolsdorf@gwu.edu

David Langer
H-BRS University of Applied Sciences
Sankt Augustin, Germany
david.langer@h-brs.de

Luigi Lo Iacono
Justus Liebig University Giessen
Giessen, Germany
luigi.lo_iacono@uni-giessen.de

Abstract

Phishing attacks via email remain a major entry point for security and privacy breaches in hospitals. In the European Union, faced with both regulatory pressure to act and limited resources for cybersecurity, hospitals may resort to minimal-effort, off-the-shelf anti-phishing interventions such as warning banners in enterprise email systems. However, their effectiveness remains uncertain, particularly given the highly diverse workforce comprising medical, nursing, functional, administrative, IT, and other staff groups. We conducted a large-scale phishing simulation at a German university hospital, targeting 7,044 email accounts, to analyze how phishing susceptibility varies across staff groups, how email characteristics—such as timing, tone, context, and persuasive framing—influence susceptibility, and how 11 common in-situ anti-phishing interventions affect risky staff behavior. We found that susceptibility but also intervention effectiveness differed markedly across staff groups. Even a small number of phishing emails posed a substantial risk that persisted for about three days. The most effective interventions involved robust technical detection, including spam filtering and in-email phishing warnings. Friction-based measures, such as disabling links and active warning pages, showed mixed but promising effects. In contrast, display name suppression and the widely used method of generic `[EXTERNAL]` email tagging had no or inconsistent effects. Surveys revealed that some staff reacted with fear, shame, guilt, and hostility, highlighting the ethical challenges of such simulations. Our findings provide actionable guidance for phishing resilience in healthcare and similarly complex organizations.

CCS Concepts

• **Security and privacy → Human and societal aspects of security and privacy**.

Keywords

Hospitals, Phishing, Susceptibility, Interventions, Resilience

ACM Reference Format:
Jan Tolsdorf, David Langer, and Luigi Lo Iacono. 2025. Phishing Susceptibility and the (In-)Effectiveness of Common Anti-Phishing Interventions in a Large University Hospital. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25), October 13–17, 2025, Taipei, Taiwan.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3719027.3765164

CCS '25, Taipei, Taiwan

ACM ISBN 979-8-4007-1525-9/2025/10
https://doi.org/10.1145/3719027.3765164

1 Introduction

Hospitals are increasingly targeted by cyberattacks [28, 38]. Phishing, in particular, remains a major entry point for cybersecurity and privacy incidents [28], often resulting in the compromise of sensitive data such as electronic health records [89]. 2–3% of incoming email and web traffic in a UK NHS trust was flagged as suspicious, demonstrating a high baseline risk [62]. At the same time, hospitals in the European Union (EU) rank in the lower midfield in cybersecurity readiness and funding compared to other sectors [29]. Key challenges include fragmented IT systems, a heterogeneous workforce, work processes that conflict with common security measures, and persistent shortages in financial resources and skilled IT and cybersecurity staff [18, 23, 44, 76]. This reflects a long-standing underregulation and underfunding of cybersecurity in the healthcare sector. A shift began in 2016 with the introduction of the Directive on Security of Network and Information Systems (NIS1) [25], the first legally binding EU-wide cybersecurity directive, which promoted stronger cyber resilience in critical infrastructures, such as healthcare. However, NIS1 had a limited impact due to delays and inconsistent implementation across member states. To address this, the EU introduced the NIS2 Directive [27], which came into force in October 2024. NIS2 introduced stronger enforcement mechanisms, including audits, fines, and management liability. NIS2 is also the first EU directive to explicitly mandate robust, human-centered cybersecurity measures. To support this shift, the EU has recently launched dedicated funding programs for hospitals [26] and backed research into security measures and awareness programs that fit hospital workflows and constraints [58].

Improving resilience requires time and significant changes in IT systems, processes, and staff awareness. Yet under NIS2, hospitals and other organizations face immediate pressure to act. Phishing is an area where a quick and seemingly low-cost response is to utilize readily available anti-phishing interventions already included in widely adopted systems such as Microsoft 365, Google Workspace, or Barracuda, with minimal impact on infrastructure. These tools often include, or can be extended with, in-situ features such as warning banners or external sender tags, which vendors actively promote [35, 54, 57]. This raises a central question: If hospitals are vulnerable to phishing and required to act, how effective are these common and readily available anti-phishing interventions?

Answering this question for hospitals in the EU seems particularly challenging due to strict labor and privacy regulations that—other than in the U.S. [36, 37, 40]—often prohibit individual tracking, making it difficult to assess phishing simulations, trainings, and interventions effectively [76]. Previous phishing simulations in the EU have been halted after backlash over deceptive emails, showing that such exercises can cause serious organizational disruptions [64].

To explore this question, we conducted two phishing simulation studies in a large German university hospital, targeting 7,044 email accounts. In Study I, we compared 12 phishing emails using a Plackett-Burman design [61] to assess which email characteristics influence susceptibility. In Study II, we used a between-subjects experimental design to test 11 anti-phishing interventions. All were drawn from commercial products and recent research, and required minimal or no changes to the hospital's existing infrastructure. Given the ongoing strain in EU hospital work environments following COVID-19 [3], we also examined how staff emotionally respond to phishing simulations. Past efforts resulted in negative staff reactions and complications due to strict labor and data protection laws [64]. Our study answers four research questions:

RQ1: How likely are phishing emails to trigger risky behavior among hospital staff? In a phishing campaign involving 12 emails of varying yet overall high effectiveness, we found a 6.5% probability that a staff member would engage in risky behavior—i.e., interact with the login prompt of the phishing site, likely submitting login credentials—within the first 12 hours. Time-based modeling shows that just 69 successful deliveries are enough to reach a 99% probability that at least one employee will do so within 12 to 24 hours. The likelihood of falling for a phishing email drops sharply after three days of its delivery. Staff in nursing and functional services were more likely to engage in risky behavior than those in medical service, administration, IT, or other areas. We also found that click rates—often the sole measure in phishing simulations [32, 72, 84]—overestimate risky behavior. Across 12 phishing emails, click rates were 9 to 23 percentage points higher than actual risky behavior (i.e., login interaction).

RQ2: Which features of a phishing email increase the likelihood of risky behavior by hospital staff? We found that email features, including timing, context, presentation, tone, and persuasive framing, had heterogeneous effects on phishing susceptibility across professional roles. Phishing emails that exploit loss aversion, urgency, authority, and liking—especially when timed before the weekend or tied to sensitive topics like payroll—are more likely to succeed. But impact varies by role, with staff showing different sensitivities. No single feature increased the probability of risky behavior by more than 6.7 percentage points across all roles. However, susceptibility varied substantially between professional groups, with differences exceeding 10 percentage points in some cases. These findings suggest that different types of phishing emails pose varying levels of risk depending on the recipient's role in a hospital. Treating an organization's staff as a homogeneous group in phishing simulations—as is common in organizational phishing literature [36, 37, 40, 43, 48, 49, 62, 64]—, susceptibility can be dramatically over- or underestimated for various professional roles. Moreover, the presence or absence of specific phishing features is strongly associated with potentially risky behavior. Even within the same email context, login interaction rates ranged from 18% to 40%, demonstrating that seemingly minor differences in email design or delivery timing can produce substantial shifts in behavior.

RQ3: To what extent do common in-situ anti-phishing interventions reduce risky behavior among hospital staff? In-situ interventions using robust technical phishing detection achieved the largest reduction in risky behavior. Routing phishing emails to spam folders and embedding explicit warning banners reduced login interaction rates by 81% to 94%.

Other interventions showed mixed results. Friction-inducing interventions such as disabling links in emails reduced login interaction rates by 61%, and an active warning page using nudges and URL highlighting led to a 44% reduction. The widely used practice of external email tagging (e.g., `[EXTERNAL]`) only showed effects when implemented via banners or combined tagging of the subject line and `From` field, resulting in a 53% and 58% reduction in login interactions. In contrast, isolated tagging in either the subject line or the `From` field yielded weak, non-significant effects. Suppressing display names had a similarly negligible impact. Nonetheless, some interventions may prove more effective within specific staff groups.

RQ4: What negative emotional responses should hospitals expect from staff after falling for a phishing simulation? To better understand hospital staff emotional reactions, we collected emotional affect from 530 voluntary survey participants who fell for the phishing attempt and interacted with the fake login prompt. Sixty percent reported being moderately to extremely scared in response to the phishing simulation; 41% felt shame, 36% guilt, and 25% hostility. These emotions were strongly correlated. Our findings highlight the psychological impact of phishing simulations on hospital staff and underscore the ethical importance of carefully designed simulations. Some employees also contacted the CISO and research team with questions and (ethical) concerns.

Contributions. To our knowledge, this study presents the first large-scale phishing simulation in an EU hospital and the most comprehensive investigation to date of phishing susceptibility and in-situ anti-phishing interventions in organizational settings, involving over 7,000 employees and 11 widely used, real-world interventions. Rather than relying on reporting click rates [36, 37, 40, 43, 62, 64], we model hazard rates over time using login prompt interaction. Our modeling provides a more realistic and operational understanding of how phishing emails unfold within the time window after delivery. Even modest phishing campaigns with relatively low numbers of emails pose a significant risk to hospitals.

We provide empirical evidence on how strategic timing (e.g., sending emails on Friday afternoons) affects phishing susceptibility, and also demonstrate how persuasion strategies influence phishing success in hospitals. Our results support the effectiveness of authority, scarcity, and liking. By contrast, we find no evidence for strong affect, despite its frequent use in real-world phishing [31].

The study demonstrates that technical measures that filter or flag risky emails—especially those marking senders as untrustworthy—prove consistently more effective. In contrast, the widely used generic tagging of external emails is unreliable. More aggressive interventions, such as hyperlink disabling or dynamic warnings, show further potential but may introduce usability trade-offs.

Unlike prior work, we account for workforce heterogeneity and demonstrate that phishing susceptibility and intervention effectiveness differ across four major staff roles. This insight likely generalizes beyond healthcare [12].

Our results provide actionable guidance for organizations facing regulatory pressure and tight resource constraints. While especially relevant for hospitals in the EU, the findings apply more broadly to institutions implementing similar interventions.

2 Background and Related Work

2.1 Phishing Susceptibility

2.1.1 *Phishing susceptibility in hospitals.*

Research on phishing in hospitals has primarily relied on aggregated click rates (i.e., users clicking on malicious links) as a behavioral proxy for potentially harmful actions and as a measure of phishing susceptibility, but without distinguishing between different types of hospital staff. Reported click rates across studies in hospitals range from 2% to 55% [36, 40, 64]. A retroperspective analysis of phishing simulations in six U.S.-based hospitals over eight years revealed [36] a median click rate of 16.7%. The authors observed a decline over time. In contrast, a recent U.S.-based phishing simulation study with 19,000 healthcare workers over eight months [40] found click rates increased over time rather than decreasing. Meta-analytic estimates from the broader phishing research, including various industries and contexts, indicate an average click rate of 24% [70].

Our study complements prior work by showing that the login interaction rate is a more reliable behavioral proxy for hospital phishing simulations and by demonstrating differences in staff behavior, highlighting the importance of accounting for personnel diversity. Additionally, we model the hazard rate over time to estimate the probability that an employee will engage in a harmful action in the hours or days following receipt of a phishing email.

2.1.2 *Drivers of phishing susceptibility.*

Phishing susceptibility is the result of a mix of (1) a person's *long-term stable* traits (e.g., knowledge, demographic attributes), (2) *situational* factors (e.g., workload, stress), and (3) *in-the-moment* states (e.g., cognitive and emotional states) [11, 21, 70, 71, 79, 86, 90].

For *long-term stable* factors, higher phishing knowledge is generally linked to lower susceptibility to phishing [90], while findings on age and gender remain inconsistent [11, 66, 90].

Situational factors such as workload, time pressure, stress, and fatigue are linked to higher phishing susceptibility [21, 70, 90]. These conditions likely impair decision-making and prompt users to rely on heuristic processing, i.e., focusing on surface-level email features such as layout and branding instead of verifying links or sender details [45, 59, 71, 79]. Susceptibility also varies across organizations and industries [12]. Clear security policies and norms may improve phishing resistance [21, 66, 90]. Few studies have investigated situational factors in hospital settings. Jalali et al. [43] found that workload, not intention, was the strongest predictor of phishing susceptibility, indicating that even motivated hospital staff may fall for phishing attacks under work-induced stress. Additionally, a retrospective study across six U.S. hospitals (2011–2018) found lower click rates in spring and summer than in fall, likely reflecting seasonal workload patterns [37]. While attackers could exploit email timing, it remains an underexplored factor [90].

In-the-moment reactions are commonly manipulated by phishing emails through social engineering tactics [31]. Frequently used principles include *authority* [31, 77], *distraction* [77, 91] (especially *scarcity* [91] and *strong affect* [31]), as well as *reciprocity*, *integrity*, and *consistency* [31, 91], which have large overlap in practice. *Liking* and *social proof* seem less common, likely because they require more personalized or contextual information [77]. For example, these latter tactics often rely on personalization by using the recipient's name or referencing job-specific information [66, 70]. Evidence on the real-world effectiveness of persuasion tactics in organizational phishing research is mixed, likely due to variations in email content, organizational context, and research methods [73, 74, 77, 86]. Common email manipulations exploit heuristic processing. Standard techniques include HTML formatting, spoofed sender fields [42], and misleading or obfuscated URLs [32, 80].

We extend prior work by systematically analyzing how the timing of emails and the presence of strong affect impact risky behavior. We also examine how the most common and context-relevant persuasion tactics and standard email manipulation techniques affect response rates and how effects differ by hospital staff group.

2.1.3 *Employee Reactions to Phishing Simulations.*

Simulated phishing attacks can trigger a wide range of emotional responses, both positive and negative. In sectors such as manufacturing, logistics, finance, and IT, phishing simulations are generally well-received, especially when clearly explained and integrated into a training strategy [48, 67]. At the same time, participants in these sectors have reported negative reactions, including shock, fear, shame, embarrassment, mistrust, confusion, annoyance, and self-directed disappointment or anger [67, 68]. These negative perceptions were more common among employees working under high time pressure [67]. Simulated phishing campaigns can also reduce self-efficacy and contribute to stress, particularly when poorly managed [68].

These effects are likely more severe in hospitals, where staff well-being is under greater strain than in any other industry in the EU [24]. EU healthcare workers—especially medical, nursing, and functional services—face high levels of burnout, mental health strain, and job dissatisfaction, often linked to chronic understaffing, time pressure, and friction with management [3, 24, 51]. Strikes and labor actions are frequent [17]; our own simulation had to be rescheduled due to a strike. The only large-scale phishing simulation in an EU hospital was halted after staff backlash and union complaints. Employees reported confusion and disappointment, and the campaign was discontinued [64]. This reinforces concerns that poorly designed simulations can harm staff well-being and erode trust—especially when perceived as manipulative or punitive [82].

To better understand the dynamics of adverse reactions, we extend prior work by quantifying the emotional responses of hospital staff who fall for phishing emails during simulations.

2.2 Human-Centric Anti-Phishing Interventions

Human-centered anti-phishing efforts fall into two categories: (i) ex-ante measures (e.g., training) and (ii) in-situ interventions that activate during user interaction [32]. Given their widespread use and assumed potential to reduce phishing susceptibility, most organizational phishing research has focused on ex-ante interventions [21, 32, 40, 48, 49, 90]—including studies in hospital settings [36, 40]. However, ultimately, findings suggest that embedded training approaches, in particular, offer limited effectiveness [32, 45], while consuming substantial financial and human resources [14].

In contrast, widely used in-situ interventions have undergone less rigorous study, especially in real-world organizational contexts. Widely available in-situ anti-phishing mechanisms in off-the-shelf email infrastructure include sender identity indicators [50], [EXTERNAL] tagging to mark emails originating from outside an organization [35, 55, 57], dedicated phishing warning banners added as visual warnings to emails [35, 42, 54, 57], and friction-inducing mechanisms like delivering emails to spam folders, disabling links for emails from certain domains or spam, and showing warnings after link clicks [53, 63]. Technically, email delivery and in-situ defenses depend on a mix of authentication protocols (SPF [47], DKIM [5], DMARC [52]) and phishing classifiers [2, 41].

Sender identity indicators display technical source information (e.g., "via" tags), but studies show users often misinterpret them, limiting effectiveness [50]. External tags are better received—especially when users work with few outside contacts—but preferences for tag placement and actual impact remain unclear [67, 86].

Unlike sender identity indicators or external tags, phishing warning banners are triggered only when technical checks flag an email as a potential phishing attempt. Their effectiveness has been moderately studied in organizational contexts. A crowdsourced study found that banners warning of spoofing reduced clicks by about 10 percentage points, suggesting limited impact [42]. In contrast, a large field study reported a 64–67% reduction in click rates, with no difference between brief and detailed warnings [49]. Other work shows that effectiveness depends heavily on placement, timing, and content [60]. Warnings are more effective when displayed near the phishing link or triggered by user interaction (e.g., hovering, clicking), although this requires support from the email client. Some research has explored friction mechanisms, such as click delays for unverifiable sources paired with tooltips, which can help users better assess a link's legitimacy [81].

A key concern is long-term efficacy. Users habituate quickly, leading to alert fatigue and lower responsiveness. Even well-placed warnings are often ignored due to cognitive overload, weak threat comprehension, or misplaced trust in polished emails [45, 50, 90].

Our study provides the first systematic evaluation of 11 common and readily available in-situ anti-phishing interventions in a large organizational context, spanning four strategies: (1) sender identifier manipulation, (2) external tagging, (3) embedded phishing warnings, and (4) friction-inducing tactics including link disabling, active warning pages, and spam folder delivery.

3 Method

3.1 Study Organization

The study was conducted at a large university hospital in Germany with around 8,500 staff. The researchers connected with the hospital's Chief Information Security Officer (CISO) team at an industry forum in early 2023 and started the collaboration a few months later. The hospital had independently planned to run a phishing simulation and awareness training, for which it commissioned an external provider through a formal tender. The researchers worked only as scientific advisors. They helped design the phishing experiment, crafting the phishing emails, and created the study protocol and debriefing page, which included a short, voluntary questionnaire. The hospital utilized a standard mail user agent across all departments, ensuring consistent email rendering.

As a typical organizational measure, the phishing simulation had to go through the hospital's standard review and approval process. The CISO team secured the necessary approvals from hospital management, employee councils, and the data protection officer. The researchers supported each step and also coordinated technical matters with the external service provider. The technical setup was completed in December 2023. The hospital ensured that the provider had access to employee email addresses and that phishing emails were not blocked by internal systems. This involved setting up allowlisting rules in the hospital's email infrastructure.

The first simulation took place in January 2024, the second in May 2024. After each, the hospital shared anonymized data with the researchers for analysis. The results were reported back to the CISO. No awareness training (e.g., embedded training [48]) was conducted before or during the simulations.

3.2 Ethical Considerations

Our university does not have a formal Institutional Review Board (IRB) process. Nevertheless, we took comprehensive steps to minimize potential harm to our participants and ensure ethical integrity. While the simulation was initiated by the hospital independently of our research, we are aware that phishing simulations can inadvertently cause stress for employees [68], placing them in uncomfortable situations. We hence incorporated this into the design, monitoring, and support structure of the study.

3.2.1 Study clearance and data handling. To obtain approval from the hospital, we engaged in detailed discussions with the CISO and submitted a full study protocol. This included details on study procedures, data handling, privacy safeguards, and consent forms for the optional exit survey. The protocol was reviewed by the hospital's Data Protection Officer, as well as both the scientific and non-scientific works councils. We also held a workshop with the works councils to address their concerns and refine the study design based on their feedback. Following this process, formal approval was granted by the works councils and the Data Protection Officer, and final authorization was obtained from hospital management. Additionally, our own institution's Data Protection Officer reviewed and approved the protocol and data handling.

The phishing simulation involved sending hospital staff simulated phishing emails that directed them to a phishing website with a login prompt. We refer to "login interaction" as entering text into the username and password fields and clicking "Login." To protect employee safety and anonymity, no submitted credentials were collected, stored, or verified for authenticity by the researchers or administrators. All data were analyzed in aggregated form (i.e., click and login interaction rates), with a targeted minimum group size of 100 employees, in compliance with the requirements of the hospital's works councils. Researchers never had access to employees' personal data. Employee email addresses and names were managed by the hospital and the external contractor. The phishing website and the online survey were hosted on researcher-controlled infrastructure. Server logs containing IP addresses were automatically deleted, and no data that could identify individuals was stored.

Finally, we acknowledge that publishing real-world phishing susceptibility metrics could, in theory, inform malicious actors. However, we disclose no information that a determined attacker could not uncover through simple trial and error. Conversely, transparency serves a broader public interest. Our study provides CISOs, IT teams, and institutional decision-makers with grounded data on how users respond to phishing, how risks evolve over time, and when vigilance tends to decline. These insights can inform evidence-based defenses and justify targeted investments, particularly in resource-constrained environments such as hospitals.

3.2.2 Consent Waiver. As is common in organizational phishing research [40, 49], the hospital scheduled the phishing simulations independent of the researchers as part of its routine security measures, which staff are contractually required to follow. Neither the simulation nor the researchers' involvement created additional risk to employees and did not infringe on their rights. Phishing is a real and ongoing threat, and employees may receive malicious emails at any time, regardless of this study. The use of deception was essential to preserve ecological validity. Individual informed consent, even post hoc, was not feasible due to strict anonymization. To this end, the study adhered to the ethical guidelines of the German Society for Psychology [33] and the Ethics Code of the German Sociological Association [34] for covered research: No employee identification was possible, the study aimed to improve organizational processes, and no legal, financial, or professional risk resulted for the hospital staff. Following recommendations by the American Political Science Association [9] for covert research without individual consent, we sought institutional forms of consent: The hospital's legally elected works council was fully informed, involved in the study's planning, and approved the study on behalf of the workforce. In addition, employees were debriefed after the study (see supplementary materials [75] for details). Ultimately, the study offers meaningful societal value by deepening our understanding of phishing susceptibility in hospital settings and evaluating the impact of widely used anti-phishing measures. The findings can help the participating hospital and similar organizations improve their defenses against phishing threats.

3.3 Procedure and Data Elicitation

To answer our research questions, we conducted two independent phishing simulation studies. In Study I, we focused on understanding the susceptibility of hospital staff to phishing. In Study II, we tested 11 low-cost, human-centered phishing interventions.

3.3.1 Participant selection. The phishing simulation targeted 7,044 employees. For technical reasons, we could only target hospital employees listed with active roles in the organization's Active Directory. Because hospitals employ a highly heterogeneous workforce, we controlled for role-specific differences in susceptibility to phishing. This approach is guided by prior research showing that staff roles, particularly medical and nursing staff, differ in security awareness and behavior [4, 30], and by evidence that phishing susceptibility varies across industries and organizational contexts [12].

In coordination with the hospital's CISO team—and to ensure at least 100 participants per group—we divided the workforce into four groups: (1) *Medical Service*, (2) *Nursing & Functional Services*, (3) *Administration & IT*, and (4) *Other Personnel.* The grouping reflects the practical needs of the CISO team and clear differences in digital work routines: *Medical* staff use personal computers and handle a higher volume of sensitive patient data and emails. *Nursing & Functional* staff work more at the bedside, often share devices, and check email less often. *Admin & IT* staff have office-based roles. The *Other* group is heterogeneous and includes smaller staff groups, such as technical services, cleaning, kitchen, laundry, childcare, pastoral care, and other support roles.

Each of the four groups was randomly split into 12 subgroups, yielding 48 total subgroups. Random assignment improves causal inference by reducing confounding variables [69]. This allows observed differences to be linked to the interventions themselves, enabling valid counterfactual reasoning [8]. To meet the minimum of 100 participants per condition, we reallocated participants from the *Other* group to the smaller *Medical* (1,174) and *Admin & IT* (746) groups, which initially fell short of the 1,200 required.

3.3.2 Study I: Phishing susceptibility. In Study I, we employed a fractional factorial screening experiment using a Plackett–Burman design [61]. This design efficiently estimates main effects [56], allowing us to test a relatively large set of factors within the constraints of a limited sample size. To control for variability in phishing susceptibility across job functions, we incorporated job function as a blocking factor into the Plackett–Burman design. This approach enabled us to account for systematic group-level differences while isolating the main effects of individual email features. Each subgroup received one of 12 distinct phishing emails. These emails varied across a set of binary design factors (see Sec. 3.4) and were constructed according to a Plackett–Burman design. The full experimental layout is provided in the supplementary materials [75].

3.3.3 Study II: Anti-phishing intervention effectiveness. In Study II, a separate between-subjects design was implemented to evaluate 11 anti-phishing interventions (see Sec. 3.5). Eleven subgroups per personnel category received emails containing one intervention each, while the twelfth group served as the control condition, receiving a phishing email without any intervention. For the topic of the email, we chose a voucher offered by the hospital's cafeteria.

3.3.4 Email and website design. All phishing emails included obfuscated links with arbitrary-looking domains, including the-pas.de, your-pas.de, service.vacations, and attack.promo, as this was the only strategy supported by the phishing contractor. In Study I, URLs were hidden behind the anchor text "click here." In Study II, URLs were shown in full for consistency, as the "disabling links" intervention required users to copy visible URLs. Upon clicking, the contractor recorded the "click" and redirected participants to a spoofed login page at HOSPITALNAME.multi-health.net, a domain registered for this study. The website featured a deep HTML clone of the hospital's official external-facing login portal. If participants interacted with the login prompt, i.e., entered text into the login fields and clicked "Login," the contractor recorded the click and then redirected them to a debriefing page. No credential data was ever stored or analyzed to ensure participant safety and anonymity. The debriefing page explained that this was a simulation, that no credentials were exposed, and that participants remained anonymous with no professional consequences. It also provided information about the phishing simulation as well as contact details for the research team and the hospital's CISO office. The debriefing page in Study I did not include any particular form of training.

3.3.5 Exit-questionnaire design. Participants were invited to complete a short, voluntary exit survey that collected informed consent and assessed emotional reactions to the simulation. The survey

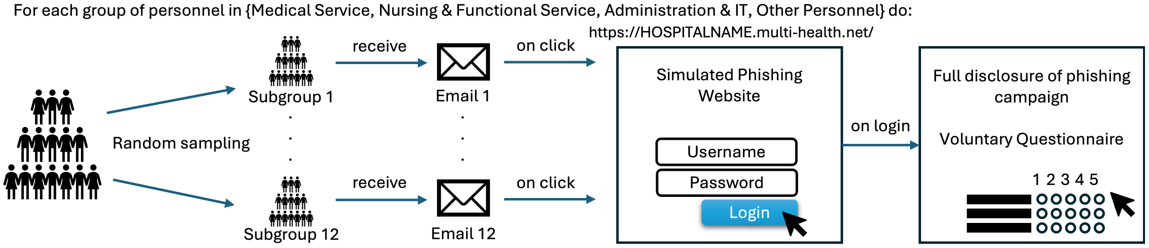

For each group of personnel in {Medical Service, Nursing & Functional Service, Administration & IT, Other Personnel} do:
Random sampling
Subgroup 1
Subgroup 12
receive
Email 1
Email 12
on click
https://HOSPITALNAME.multi-health.net/
Simulated Phishing Website
Username
Password
Login
on login
Full disclosure of phishing campaign
Voluntary Questionnaire
1 2 3 4 5

Figure 1: Phishing Campaign Method: Each of the four personnel groups was randomly divided into 12 subgroups (totaling 48 subgroups), with each subgroup receiving one of 12 emails. In Study I, emails followed a Plackett–Burman experimental design [61]. In Study II, one email served as a control, while the remaining implemented individual phishing interventions.

was intentionally limited to a few closed question items, following explicit requirements from hospital management to avoid overburdening staff. This compromise ensured organizational support while still allowing us to capture employee emotional responses. Emotional responses were measured using a subset of the negative affect subscale from the German version of the PANAS-X questionnaire [13, 85], specifically focusing on four items: *scared*, *ashamed*, *guilty*, and *hostile*. While the full negative affect subscale includes 10 items, the chosen four-item set correlates strongly with the excluded six-item subset ($r = .6$). Participants were asked to indicate the extent to which they felt each emotion in response to the simulation, using a five-point scale from "not at all" to "extremely."

3.4 Selection of Email Phishing Features

Considering sample size constraints, we implemented ten email features using a fractional factorial design. The selection of manipulations was informed by the NIST PhishScale [20, 71], prior research on phishing susceptibility (cf. Sec. 2.1.2), widely used social engineering tactics [31], actual phishing emails reported by the hospital, and established behavioral patterns in healthcare. Also, factors such as power hierarchies, professional subcultures, and perceived authority [6, 65] are known to influence staff responses. In addition, staff in hospital environments often operate under cognitive overload and time pressure [18, 38], and frequently experience alert fatigue due to a constant stream of urgent messages [7, 19]. These conditions may increase susceptibility to phishing, particularly when emails exploit emotional or social triggers.

Based on these considerations, we implemented six *in-the-moment* features tailored to the hospital context. To manipulate *authority*, we varied the sender type (internal vs. external) and the presence or absence of a formal closing that included authentic contact details taken from the hospital's website and the hospital's official legal email sub-script to enhance credibility.

To assess the effects of *distraction* and *liking*, we varied three elements: (1) urgency, using language like "immediately" and "soon"; (2) salutation type, comparing generic greetings vs. personalized greetings with the recipient's name; and (3) strong affect, adding "ACHTUNG" (caution/ attention) to the subject line and "ACHTUNG – WICHTIG" (important) in the body. These terms were picked from the Berlin Affective Word List [83], which rates German words by valence (pleasant to unpleasant) and arousal (calm to excited). While strong affect is a common tactic in phishing campaigns [31], its effectiveness in organizational settings remains largely unexplored.

Moreover, we manipulated message framing (gain vs. loss) to reflect *reciprocation* versus *distraction* appeals. Messages using *gain* emphasized access to a new, improved system requiring employee action (account activation), while *loss* messages warned of discontinuation or deactivation (account closure).

We further implemented two *situational* features focused on timing to capture the effects of fluctuating workload and attention [43]: day of the week (Monday vs. Friday) and time of day (morning vs. afternoon). These variables are particularly relevant in hospitals, where 24/7 clinical operations contrast with limited administrative and IT support outside core hours [76]. To our knowledge, this is the first empirical test of how email timing influences phishing susceptibility in a high-stress, real-world organization [90].

Finally, we manipulated contextual relevance by comparing a generic Outlook upgrade message—commonly used in phishing simulations and encountered in practice [40]—with a payroll-related message about a system update, which the CISO team identified as realistic and high-stakes. We also varied the email format (HTML vs. plain text), as HTML emails are generally perceived as more credible and deceptive due to their polished appearance [90]. Details are provided in the supplementary materials [75].

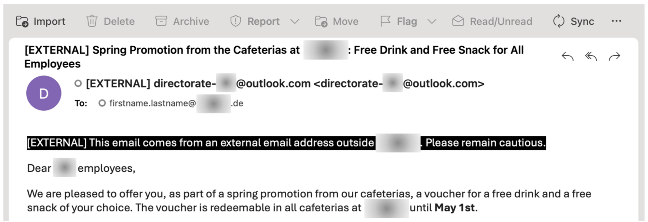
3.5 Selection of Anti-Phishing Interventions

The sample size requirement enabled us to test 11 in-situ anti-phishing interventions. We selected these interventions based on their practical feasibility in hospital environments, specifically those that are supported by existing enterprise email infrastructure and require minimal technical or organizational changes. Each was designed to be deployable using the hospital's existing email systems without any changes to mail user agents.

Sender Identifier Manipulation. We tested display name suppression, i.e., using `From: <john.smith@hospital.org>` instead of `From: John Smith <john.smith@hospital.org>`, which eliminates a visual shortcut users often rely on when assessing email legitimacy [59, 86, 90], and which is frequently spoofed [86].

External Tagging. We tested four `EXTERNAL` tagging variants: (1) subject line, (2) `From` header, (3) email body, and (4) a multi-tag version combining all three (cf. Fig. 2).

Phishing Warning Banners. We tested three phishing warning banners: (1) a plain text banner already used by the hospital, (2) an HTML version with a warning color, and (3) an HTML banner



[EXTERNAL] Spring Promotion from the Cafeterias at [redacted]: Free Drink and Free Snack for All Employees

○ [EXTERNAL] directorate-[redacted]@outlook.com <directorate-[redacted]@outlook.com>

To: ○ firstname.lastname@[redacted].de

[EXTERNAL] This email comes from an external email address outside [redacted]. Please remain cautious.

Dear [redacted] employees,

We are pleased to offer you, as part of a spring promotion from our cafeterias, a voucher for a free drink and a free snack of your choice. The voucher is redeemable in all cafeterias at [redacted] until **May 1st**.

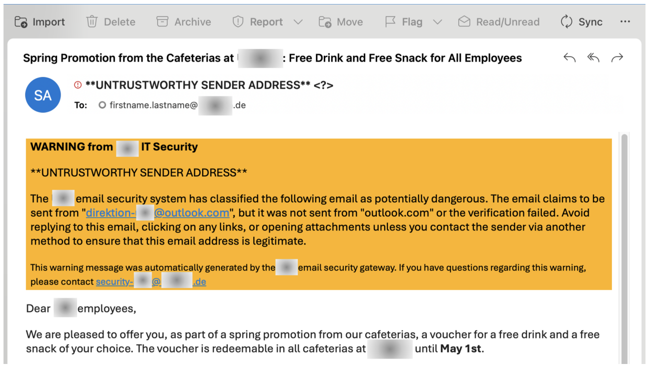
Figure 2: Multi-Tag external tagging.

combining a warning color with sender manipulation, by replacing the From field with "UNTRUSTWORTHY SENDER ADDRESS <?>", moving the actual address into the banner, and providing an explanation that the sender address could not be verified (cf. Fig. 3). This design aims to disrupt employees' heuristic processing of the potentially spoofed sender information when authentication checks, such as DKIM or SPF, fail.

Friction. We investigated three friction-inducing mechanisms that aim to introduce deliberate interruptions in the user's interaction flow with phishing content, aiming to increase cognitive engagement and reduce the likelihood of impulsive, unsafe actions [32, 45, 81]. The three mechanisms were (1) email delivery to the spam folder, (2) link disabling to force users to copy and paste the URL manually, and (3) an active warning shown in the browser after clicking the phishing link. Our implementation of an active warning is shown in Fig. 4. Its design is informed by previous studies and applies a combination of interactive elements, nudges, and URL highlighting [32].

3.6 Study Limitations

As with all organizational phishing simulations, our study has limitations. First, it was conducted in a single hospital. Although the institution is large and includes a range of departments and job roles, the findings may not generalize to settings with different organizational or technical contexts [12]. Further, the group *Other Personnel* was highly heterogeneous, limiting the precision of subgroup analyses. Additionally, to meet group size requirements, we reallocated some staff from *Other Personnel* into *Medical Service* and *Admin & IT*, introducing minor overlap and potential bias. This compromise preserved the Plackett–Burman design and avoided reducing the experimental coverage of email features and intervention strategies.



Spring Promotion from the Cafeterias at [redacted]: Free Drink and Free Snack for All Employees

****UNTRUSTWORTHY SENDER ADDRESS** <?>**

To: ○ firstname.lastname@[redacted].de

WARNING from [redacted] IT Security

UNTRUSTWORTHY SENDER ADDRESS

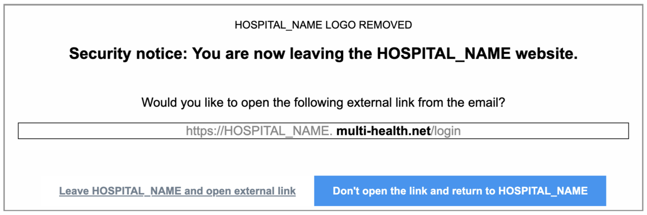
The [redacted] email security system has classified the following email as potentially dangerous. The email claims to be sent from "direktion-[redacted]@outlook.com", but it was not sent from "outlook.com" or the verification failed. Avoid replying to this email, clicking on any links, or opening attachments unless you contact the sender via another method to ensure that this email address is legitimate.

This warning message was automatically generated by the [redacted] email security gateway. If you have questions regarding this warning, please contact security-[redacted]@[redacted].de

Dear [redacted] employees,

We are pleased to offer you, as part of a spring promotion from our cafeterias, a voucher for a free drink and a free snack of your choice. The voucher is redeemable in all cafeterias at [redacted] until **May 1st**.

Figure 3: Warning banner using an HTML banner in the email body and additional From header manipulation for an untrustworthy sender.



HOSPITAL_NAME LOGO REMOVED

Security notice: You are now leaving the HOSPITAL_NAME website.

Would you like to open the following external link from the email?

https://HOSPITAL_NAME. **multi-health.net**/login

Leave HOSPITAL_NAME and open external link
Don't open the link and return to HOSPITAL_NAME

Figure 4: Active warning shown after link click.

We further note that the study did not test all persuasive techniques and specifically excluded *social proof*. Instead, it prioritized those most relevant to the hospital context and commonly used in phishing attacks. Given that *social proof* has been shown to influence user behavior in organizational settings [21, 31, 71], it remains a strong candidate for inclusion in future hospital-based studies.

Due to technical limitations of the contractor's simulation platform, links in phishing emails varied across messages. Although randomized, the inconsistency may have influenced click rates, particularly in Study II, where visible links likely appeared more suspicious. Additionally, we could not determine the proportion of login interactions that involved actual credential entry, due to ethical and privacy constraints. Therefore, login interactions are treated as a proxy for potentially risky behavior. While this may reduce accuracy, it aligns with common practice [48, 49]. If actual credential entry were lower, it would strengthen our broader point: click rates likely overestimate real behavioral risk.

Moreover, we did not track false positives, i.e., legitimate emails mistakenly flagged as phishing. In deployment, this could lead to alert fatigue and reduce trust in banners or external tags, lowering long-term effectiveness [78]. Consequently, our results likely represent an upper bound of intervention performance, reinforcing that common anti-phishing measures cannot be fully relied upon. We provide further discussion in Sec. 5.2.

Technical issues further affected three emails in Study I, which failed to send. This slightly reduced the sample size to n = 99 for two participant groups. The deviation was approved by the hospital and does not impact the overall conclusions. Finally, the study included two measurement points: one to assess baseline phishing susceptibility and another to evaluate intervention effectiveness. While learning effects between phases are possible, no formal training occurred, and the four-month gap makes this unlikely.

In sum, while our study is limited in scope, it is—to our knowledge—the first to conduct high-resolution phishing simulations in a hospital setting within the EU, under the constraints of complex legal and organizational structures that do not allow for individual tracking as is common in research outside the EU [36, 49]. Still, the experimental design enabled a detailed comparison across personnel groups and message characteristics. This approach represents a promising model for balancing experimental control with real-world applicability in sensitive and high-stakes environments.

4 Results

4.1 Phishing Susceptibility

To answer RQ1, "*How likely are phishing emails to trigger risky behavior among hospital staff?*", we analyzed click and login interaction rates from the phishing simulation in Study I. All data is available in the supplementary materials [75].

4.1.1 Click and Login Interaction Rates. Across all 12 simulated phishing emails sent in Study I, approximately 31% (sd = 6.6, min = 21.3%, max = 43.8%) of hospital staff clicked on the link, and 26% (sd = 6.4, min = 18.2%, max = 39.7%) interacted with the login prompt. Notably, both click and login interaction rates varied by a factor of two across emails, underscoring the impact of email-level differences (e.g., visual appearance or tone) on user behavior. Conversely, the overall click-to-login ratio of 85.4% (sd = 4.3, min = 79%, max = 91%) was consistently high. Furthermore, we find that click and login interaction rates for the staff groups *Admin & IT*, *Medical*, and *Other* were below average, whereas *Nursing & Functional* was above average (cf. Table 1). The click-to-login conversion rates were similarly high across all groups (84.2% – 87.5%).

4.1.2 Group Differences in Login Interaction Rates. To assess differences in login interaction rates across occupational groups, we conducted a Pearson's Chi-squared test, which revealed a significant effect ($\chi^2(3) = 20.47$, $p < .001$). Pairwise proportion comparisons with Holm correction for multiple testing [1] indicated that the *Nursing & Functional* group differed significantly from both the *Other* ($p = .0015$) and *Admin & IT* ($p = .0015$) groups.

To better understand the magnitude and direction of group differences, we ran a logistic regression with login behavior as the outcome and occupational group as the predictor. The *Nursing & Functional* group was set as the reference category. We computed Average Marginal Effects (AMEs) [10] to provide an intuitive interpretation of the results. AMEs indicate the average change in probability, expressed in percentage points (p.p.), associated with a one-unit change in the predictor variable. We find that *Nursing & Functional* had significantly higher login probabilities compared to the remaining groups. Specifically, the probability of login interaction was 5.7 p.p. lower for *Admin & IT* ($p < .001$), 4.8 p.p. lower for *Other* ($p < .001$), and 3.2 p.p. lower for *Medical* ($p = .036$).

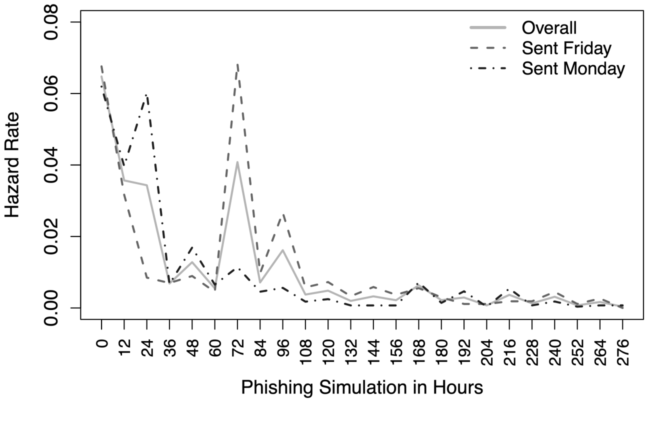


Figure 5: Estimated hazard rates in Study I.

Table 1: Rates of potentially risky behavior in Study I.

Group	*n*	Click-%	Login-%	Click-to-Login
Medical Service	1199	29.4%	25.8%	87.5%
Nursing & Func	2700	33.9%	29.0%	85.6%
Admin & IT	1199	27.4%	23.3%	84.8%
Other	1943	28.7%	24.2%	84.2%

Note. "Login" refers to interaction with the login prompt on the phishing website; the authenticity of any submitted credentials was not verified.

4.1.3 Hazard Rate. We use the hazard rate [87] to model how the risk of falling for a phishing attempt changes over time. The hazard rate $h(t)$ describes the instantaneous risk that an employee will interact with the login prompt on the phishing website at a specific point in time t, given that they have not done so before [87]. The hazard rates, modeled in 12-hour intervals and including emails sent on Monday and Friday, are reported in Fig. 5.

At the start of the simulation, the overall hazard rate was 0.065, indicating a 6.5% probability that a hospital staff member would interact with the login prompt within the first 12-hour window. Notably, phishing emails appeared to receive a "second chance" on the following working day, with hazard rates rising to levels comparable to those observed in the initial interval. For Monday emails, a peak occurred after 24 hours; for Friday emails, the peak appeared after 72 hours, immediately following the weekend. Smaller secondary peaks were also observed on the second working day. Hazard rates began to stabilize by the third day, and for both conditions, they remained below 0.7% after 4.5 days.

To illustrate the implications of the observed hazard rates, we modeled the probability that at least one employee would interact with the phishing login prompt under three hazard rates (cf. Fig. 6). With a 6.5% hazard rate in the first 12 hours, sending just 45 phishing emails results in a 95% probability of at least one login prompt interaction. This corresponds to phishing emails reaching only 0.64% of the 7,041 hospital email accounts. With 69 emails (0.98% of accounts), the probability rises to 99%, and with 138 emails (1.96% of accounts), it reaches 99.99%. However, a lower hazard rate of

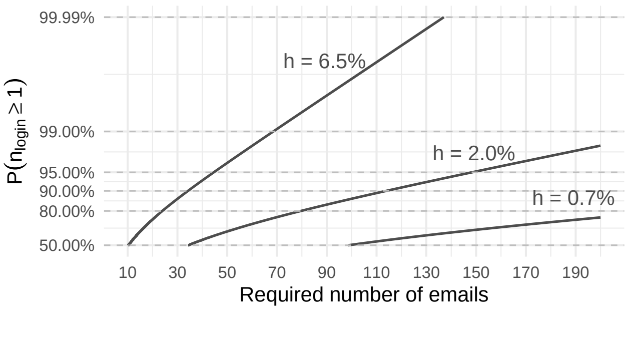


Figure 6: Estimated number of phishing emails (x-axis) needed to achieve a probability P of at least one employee engaging with a phishing site's login (y-axis) for various hazard rates (h), with a logit-like stretched transformation on the y-axis to highlight higher probabilities.

Table 2: Average Marginal Effects (AMEs) of various email features (login prompt interaction).

Category	Factor	All	Medical	Nursing & Func	Admin & IT	Other
Timing	Friday vs. Monday	0.027^{*}	-0.002	0.060^{**}	0.037	0.004
	Afternoon vs. Morning	-0.056^{***}	-0.135^{***}	-0.009	-0.024	-0.103^{***}
Context and Presentation	Payroll vs. Email Account	0.055^{***}	0.098^{***}	0.037	0.023	0.081^{***}
	HTML vs. Plain Text	-0.049^{***}	-0.092^{**}	-0.034	-0.036	-0.064^{**}
	Personalized vs. Generic Salutations	0.022	0.018	0.062^{**}	-0.027	-0.008
	Internal vs. External Sender	0.002	0.029	-0.017	0.029	0.003
Tone and Appeal	Gain vs. Loss	-0.067^{***}	-0.014	-0.068^{***}	-0.086^{**}	-0.074^{**}
	Urgency: Yes vs. No	0.029^{*}	0.078^{*}	0.005	0.036	0.044
	Formal closing + postscript: Yes vs. No	0.029^{*}	0.048	0.000	0.011	0.074^{**}
	Strong Affect: Yes vs. No	-0.016	-0.038	0.028	-0.063	-0.035
Further Statistics	*Nagelkerge‘s Pseudo-R^2 (Cragg & Uhler)*	0.032	0.079	0.026	0.039	0.072
	AUC	0.594	0.649	0.583	0.610	0.645
	n	7041	1199	2700	1199	1943

Note. *p < .05, **p < .01, ***p < .001, p-values are Holm-corrected.
Example interpretation: A phishing email sent in the afternoon significantly reduces the predicted probability of an employee interacting with the login prompt by, on average, 5.6 p.p. compared to sending it in the morning. I.e., phishing susceptibility is higher for emails sent in the morning.

approximately 2.0%, as observed on the third day, means that the probability of at least one login interaction with 138 emails drops to 94%. After 4.5 days, the probability further decreases to 62% and the hazard rate is 0.7%.

4.1.4 Summary. Study I revealed substantial variability in both click and login interaction rates across the 12 phishing emails. Despite this variation, the average click-to-login conversion rate remained high, indicating that users who clicked were very likely to engage with the login prompt. Significant differences across occupational groups suggest that susceptibility is unevenly distributed within the organization. In particular, the *Nursing & Functional* group showed significantly higher susceptibility compared to other groups. Time-based analysis revealed that phishing emails maintained their effectiveness beyond initial delivery, with hazard rates spiking again on the following working day, highlighting a delayed risk pattern missed by single-timepoint assessments. Overall, these findings demonstrate that even small-scale phishing campaigns targeting as few as 0.64% of staff can trigger risky behavior and likely credential compromise.

4.2 Feature Effects

To answer RQ2, *“Which features of a phishing email increase the likelihood of risky behavior by hospital staff?”*, we conducted a logistic regression analysis to extract the main effects of 10 different cues in phishing emails from Study I. In Table 2, we present the Average Marginal Effects (AMEs) of how email cues impact login interaction rates and thus employees’ susceptibility to phishing.

4.2.1 Timing. The time of day at which phishing emails were delivered had a significant effect on hospital staff susceptibility. Emails sent in the morning had a 5.6 p.p. higher likelihood of login interaction compared to those sent in the afternoon ($p < .001$). These effects were particularly pronounced among specific occupational groups: morning delivery increased login interaction by 13.5 p.p. among *Medical* staff ($p < .001$) and by 10.3 p.p. among *Other Personnel* ($p < .001$). In contrast, the *Nursing & Functional* and *Admin & IT* groups showed no significant difference based on time of day.

Day-of-week effects were more modest overall. Emails sent on Fridays increased login interaction by 2.7 p.p. across all staff ($p < .05$), with the effect primarily driven by the *Nursing & Functional* group, who showed a 6.0 p.p. increase compared to Monday deliveries ($p < .01$). No significant weekday effects were observed for other groups. Taken together, time-of-day effects had a much larger overall impact and within specific subgroups than the more modest effects observed across weekdays.

4.2.2 Context and Presentation. Emails referencing the payroll system were associated with a significantly higher overall AME of 5.5 p.p. in the probability of interacting with the login prompt ($p < .001$), compared to emails concerning email account (de-)activation. While all occupational groups showed at least a slight preference for payroll-related content, this effect was only statistically significant among *Medical* staff (9.8 p.p., $p < .001$) and *Other Personnel* (8.1 p.p., $p < .001$). A similar pattern was observed for email formatting. Emails in plain-text format led to significantly higher login interaction rates than HTML-formatted emails, with an overall AME of 4.9 p.p. ($p < .001$). However, this effect was again only significant in the *Medical* (9.2 p.p, $p < .01$) and *Other Personnel* (6.4 p.p., $p < .01$) groups.

A personalized salutation using first and last name, compared to a generic salutation, had no overall effect. However, the group *Nursing & Functional* was significantly more susceptible to emails using a personalized salutation (6.2 p.p., $p < .001$).

Finally, whether the email address had an internal or external domain did not affect the probability of disclosing login interaction.

4.2.3 Tone and Appeal. Compared to gain framing, loss framing raised the predicted probability of login interaction by 6.7 percentage points overall ($p < .001$). The effect was strongest among

Table 3: Effectiveness of various anti-phishing interventions measured as relative reduction of login interaction rates.

Category	Condition	n_{All}	All	Medical	Nursing & Func	Admin & IT	Other
				Login interaction rates			
NA	Control group	587	0.165	0.170	0.129	0.120	0.241
				Relative reductions: $1 - \frac{\text{Login Rate Treatment Group}}{\text{Login Rate Control Group}}$			
Heuristic	No Display Name	587	−0.18 (.04)	−0.12 (.03)	0.03 (.01)	0.33 (.06)	−0.51 (.16)**
External Tag	From	587	−0.24 (.06)	−0.24 (.06)	−0.21 (.04)	0.50 (.08)	−0.49 (.15)**
	Subject Line	587	−0.17 (.04)	0.00 (.00)	−0.10 (.02)	0.25 (.04)	−0.41 (.13)*
	Banner	587	−0.53 (.13)***	−0.35 (.09)	−0.41 (.09)	−0.58 (.13)	−0.67 (.22)***
	Combined	587	−0.58 (.15)***	−0.59 (.15)	−0.62 (.14)**	−0.25 (.05)	−0.64 (.21)***
Warning Banners	Plain Text	587	−0.80 (.22)***	−0.94 (.28)***	−0.76 (.18)***	−0.75 (.17)*	−0.80 (.27)***
	HTML	587	−0.83 (.23)***	−0.77 (.21)**	−0.76 (.18)***	−0.92 (.22)**	−0.87 (.31)***
	HTML + Unver. Sender	587	−0.94 (.27)***	−1.00 (.30)***	−0.83 (.20)***	−1.00 (.25)**	−0.97 (.36)***
Friction Inducing Approaches	Spam Folder	587	−0.90 (.26)***	−0.88 (.26)***	−0.86 (.21)***	−1.00 (.25)**	−0.90 (.32)***
	Link Disabled	587	−0.61 (.16)***	−0.77 (.21)**	−0.62 (.14)**	−0.08 (.02)	−0.69 (.23)***
	Active Warning Site	587	−0.44 (.11)***	−0.12 (.03)	−0.38 (.08)	−0.25 (.05)	−0.69 (.23)***
	N total	7044		1200	2700	1200	1944

Note. *p < .05, **p < .01, ***p < .001. Statistical comparisons are based on Pearson's Chi-squared tests with Yates' continuity correction, comparing each treatment group to the respective control group within each column. In parentheses are the Cramer's V values.

non-medical groups, with increases ranging from 6.8 to 8.6 percentage points. *Medical* staff was the only group that showed a weak and non-significant difference between loss and gain framing.

The presence of urgency had a minor overall effect of 2.9 p.p. ($p < .05$). The strongest impact was observed among *Medical* staff, where it increased the predicted probability of login interaction by 7.8 p.p. ($p < .05$). Conversely, the remaining personnel groups did not demonstrate any significant effects.

Conveying credibility had a minor overall effect, increasing login interaction probability by 2.9 percentage points ($p < .05$). This effect was only significant within the *Other Personnel* group, where it increased susceptibility by 7.4 p.p. ($p < .01$). In contrast, the use of strong affect in the subject line and email body did not result in any significant change in predicted probabilities across any group.

4.2.4 Summary. Our findings highlight that both the timing and framing of phishing emails critically influence hospital staff susceptibility. The features most associated with increased susceptibility include emails sent in the morning, loss framing, plain-text formatting, and payroll-related content. Importantly, no single email feature produced uniform, significant effects across all groups, underscoring heterogeneity in staff susceptibility; however, individual groups exhibited susceptibility differences as large as 12.4 p.p.

4.3 Phishing Interventions Effects

To investigate RQ3, *"To what extent do common in-situ anti-phishing interventions reduce risky behavior among hospital staff?"*, we compared login interaction rates between the treatment and control groups in Study II. The login interaction rate was 16.5% for the control group. For the 11 treatment groups with interventions, we observed an average login interaction rate of 8% (sd = 4, min = 1%, max = 14%). The average click-to-login interaction rate was 68% (sd = 12, min = 46%, max = 88%). All data is available in the supplementary materials [75]. To compare the login interaction rate, we conducted Pearson's Chi-squared tests with Yates' continuity correction. These comparisons were conducted both overall and within each occupational group. We then calculated the relative reduction in login interactions and quantified the effect size using Cramer's V (cf. Table 3). We interpret values of $V < .10$ as negligible, $.10 \leq V < .30$ as small, and $V \geq .30$ as medium effects [16].

4.3.1 Friendly Name Suppression. Suppressing the friendly name in the email sender field had an overall negligible, non-significant negative effect on login interaction rates. However, it showed a small, significant effect for the group *Other Personnel*, for whom this intervention effectively halved login interaction rates (−51%, $p < .01$). This suggests that the effect of sender display is non-uniform among different groups of hospital staff.

4.3.2 External Tagging. External tagging had mixed and generally small effectiveness. The banner tag reduced login interactions by −53% ($p < .001$), while the combined multi-tag strategy showed a reduction by −58% ($p < .001$). Although all subgroups showed similarly high relative reductions, statistically significant effects were limited to *Other Personnel* and *Nursing & Functional* . In contrast, no significant deterrent effects were found among *Medical* or *Admin & IT* staff. Notably, we observed a non-significant increase in login interaction for *Admin & IT* when using external tags in the sender field and subject line.

4.3.3 Warning Banners. Both HTML and plain-text warning banners significantly reduced login attempts across all personnel groups, with a relative reduction of −80% ($p < .001$). When combined with an "unverified sender" warning, the deterrent effect further increased, resulting in a relative reduction of −94% ($p < .001$). The most pronounced impact was observed among *Medical* and *Admin & IT* staff, where login interactions dropped to zero (−100%, $p < .001$). Overall, these interventions produced small to medium effect sizes, indicating that explicit phishing warnings are broadly effective—especially when they strip away visual trust cues often exploited in phishing attacks.

4.3.4 Friction-Inducing Approaches. The tested friction-based interventions had small to medium effects in reducing phishing susceptibility. Filing phishing emails in the spam folder significantly reduced login interaction rates (−90%, $p < .001$) and showed reliable, significant effects among all groups of employees, ranging from small to medium effect sizes. Disabling hyperlinks significantly reduced login interaction rates by −61% overall ($p < .001$). While the effect size was small, the reduction was consistent across all hospital staff groups—except for *Admin & IT*, where the effect was negligible and non-significant. Notably, among groups where the intervention was effective, the reductions were substantially greater than the overall mean.

The active warning site produced a significant overall reduction in login interaction of −44% ($p < .01$), reflecting a small effect size. However, subgroup analyses showed that this effect was almost entirely driven by *Other Personnel*, who exhibited a 70% reduction—comparable to link disabling and even surpassing phishing banners. For all other staff groups, the intervention had negligible and non-significant effects.

4.3.5 Summary. Explicit phishing warning banners and delivering emails into the spam folder produced the strongest overall reductions in login interaction rates. Disabling links and tagging external emails using banners were somewhat less effective and less reliable. Active warning pages had even smaller, less reliable effects. However, the impact of these interventions varied dramatically across occupational groups. For example, interventions in the categories heuristic and external tagging were up to 40% more effective for Other Personnel than for other groups. This suggests that some interventions are more or less effective for specific staff segments.

4.4 Hospital Staff Emotional Response

To answer RQ4, *What negative emotional responses should hospitals expect from staff after falling for a phishing simulation?*, we analyzed survey data on emotional reactions from Study I and summarized additional observations from the phishing simulation.

4.4.1 Emotional Responses. Five hundred thirty employees participated in the survey on the debriefing page in Study I, which is a response rate of 28.8%. 60% of the respondents reported feeling moderately to extremely scared in response to the phishing simulation. 40% experienced shame or guilt to a similar extent, and 25% felt hostile (cf. Fig. 7). Polychoric correlation analysis further revealed significant positive relationships between these emotional responses. To test for differences between employee groups, we conducted Kruskal-Wallis rank sum tests for feeling guilty ($\chi^2(3) = 5.86$, $p = .119$), scared ($\chi^2(3) = 4.61$, $p = .203$), hostile ($\chi^2(3) = 5.61$, $p = .132$), and ashamed ($\chi^2(3) = 7.00$, $p = .072$). None of the tests reached conventional statistical significance ($p < .05$), indicating that there is no strong evidence to suggest that participants' emotional responses differed systematically across the groups.

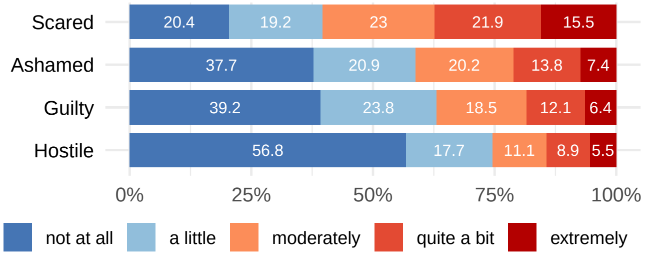


Figure 7: Participants' self-reported emotional responses to the phishing simulation (n = 530 responses). Respondents are hospital staff who entered login credentials on the phishing website <u>and</u> voluntarily participated in the survey.

4.4.2 Additional Employee Responses. We noted several unanticipated incidents during the phishing simulation. Notably, a small number of employees actively responded to the campaign in ways that revealed operational and ethical implications.

One employee reported the phishing website to the hosting provider, who then contacted us directly, disclosing the employee's name and email by forwarding the unredacted ticket. Another employee reached out with ethical concerns, to which we responded with a pre-prepared document outlining the study's rationale, ethical justification, and contact information. A third employee reported technical issues accessing the spoofed login page, unaware it was part of a simulation. The hospital also received several internal inquiries related to this matter. While these interactions revealed some identities, none were linked to behavioral or survey data.

5 Discussion

5.1 Phishing Susceptibility

Our analysis of Study I shows that a relatively small number of phishing emails are highly likely to result in at least one employee in a large organization interacting with a phishing login prompt within 12 to 24 hours. We also found a "second wave" of employee engagement on the next workday and still some engagement on the third day, showing that phishing emails remain a threat for several days. Removing or flagging them within the first day seems critical.

We cautiously estimate that our modeling of phishing susceptibility likely represents an upper bound. In particular, the observed click rates in Study I are at the higher end of the spectrum reported in previous literature on phishing in hospitals [36, 37, 40, 64] as well as the broader cybersecurity literature [70]. Click-to-login-prompt interaction rates, however, are more difficult to classify due to the lack of systematic reporting in existing research.

Furthermore, our systematic comparison of ten email features reveals substantial variation in phishing susceptibility. We provide empirical evidence that the *situational* factors *day-of-week* and *time-of-day* can be strategically exploited by attackers to increase phishing success in organizations. Notably, slightly higher susceptibility on Fridays suggests that hospitals should bolster support staffing at the end of the week and on weekends. Mainly, because phishing emails gain a "second chance" on Monday morning.

We also found that some *in-the-moment* features influenced behavior—particularly *distraction*-based techniques. The strongest effect emerged for loss framing compared to gain framing. In contrast, urgency cues had weaker effects, and strong affective language had no significant impact—despite being common in phishing [31]—with trends even pointing negatively. Likely, strong affective language already serves as a familiar warning signal to employees. Features related to *authority* and *liking* (e.g., personalized salutations) showed weaker or negligible effects. For *authority* cues, this may be because employees tend to disregard the domain part of email addresses (`local@domain.org`) [50] and focus on the local part instead, or they may skip reading closing lines altogether. These findings open up promising directions for future research.

We further observed that context influenced response: phishing emails framed as payroll-related yielded higher login rates than general Outlook notifications. Surprisingly, plain-text emails resulted in significantly more risky behavior than HTML-formatted ones. This may reflect a false sense of security, as hospital staff might associate plain-text formatting with legitimate internal communication [45].

Most importantly, our study shows that phishing susceptibility varies substantially across occupational groups, exposing the flaw in treating staff as a uniform population—a common approach in organizational phishing research [74, 86], including in the hospital context [36, 37, 40, 64]. For example, susceptibility among the *Medical* group increased by nearly twice the average when exposed to plain-text emails, payroll-related content, and urgency cues (cf. Table 2). Moreover, *Medical* and *Nursing & Functional* staff responded to the same email features in divergent, non-overlapping ways—each differing again from the *Admin & IT* group, which represents a typical job segment that is common in phishing research [74]. This suggests that findings based on one staff group cannot be easily generalized to others [12]. Finally, the consistently significant effects observed in the heterogeneous *Other* group likely reflect internal variability, reinforcing the need for more fine-grained analysis.

While differences between hospital staff or email characteristics in the range of 3 to 13 p.p. may appear small, they have substantial practical implications in organizational settings. For example, switching from an HTML to a plain-text phishing email that increases risky behavior by five p.p. in a hospital with 8,500 employees could result in 425 additional login interactions. Given the issues we observed with internal phishing reporting, our findings underscore the need for hospitals to implement robust internal communication and reporting mechanisms [45, 49, 66, 67]. This can indeed be challenging in hospitals, especially on weekends, when only first-level support—often without security training—is available [76].

5.2 Efficacy of Anti-Phishing Interventions

Overall, eight of eleven common in-situ anti-phishing interventions significantly reduced login rates by −38% to −93%. In line with previous findings [42, 49], embedded explicit phishing warnings were especially effective, as was filtering phishing emails directly into spam folders. The most effective single intervention combined sender name invalidation with embedded phishing warnings. Although not formally tested for significance, it lowered login rates by an additional 13 p.p. compared to the default phishing banner, suggesting that suppressing sender information disrupts heuristic processing and promotes more careful evaluation [45, 66]. Notably, display name suppression alone had a negligible overall effect but significantly cut login rates by half among *Other Personnel*. We therefore recommend further investigation of sender name manipulations, particularly when combined with warning banners.

The interventions of active warning pages and link disabling showed mixed results. Active warning pages reduced clicks by 44%, and link disabling by 61%, indicating potential. However, the practical utility and acceptance of link disabling remain unclear. The hospital's CISO team raised valid concerns about usability: indiscriminate link blocking could disrupt workflows, increase time spent on tasks, and offer poor cost-benefit tradeoffs for users [32, 39]. Selective deployment, such as disabling only external links, may be a more viable option. Long-term use could also lead to behavioral workarounds—for example, users habitually copying and pasting URLs. The effectiveness of friction-based interventions likely depends on context: in low-stakes cases like cafeteria vouchers, users may abandon the attempt; in high-stakes cases such as payroll or email access, they may bypass the friction regardless.

Moreover, the use of `[EXTERNAL]` tags showed limited efficacy. It significantly reduced login rates only when displayed as a banner or applied across multiple locations (banner, subject line, and `From` field). One possible explanation is that staff were not trained to recognize or interpret the meaning of external tagging, whereas the banner provided additional context. As a result, employees may not have perceived the tag as a security cue [88].

Furthermore, we cannot recommend using plain text emails as a defense strategy to hinder users from relying on weak visual cues [32, 45] as they resulted in higher susceptibility (cf. Sec. 5.1).

Lastly, we observed substantial differences in the effectiveness of interventions across occupational groups. This suggests that some staff groups may benefit more or less from specific interventions. Especially *Other Personnel* consistently showed significant effects in login rates across all interventions, with reductions ranging from 43% to 98%. This group had both higher baseline susceptibility and a larger sample size, which likely contributed to greater statistical power. Interestingly, the *Nursing & Functional* group, despite being similarly large, exhibited weaker and less consistent effects.

In conclusion, several widely available in-situ anti-phishing interventions show potential to reduce risky behavior. However, our results should be interpreted with caution. The tests were conducted under controlled conditions and likely overestimate real-world, long-term effectiveness. They do not account for repeated exposure and false positives—i.e., legitimate emails incorrectly flagged as phishing—which can lead to habituation [11, 78]. Continuous exposure to false alarms may erode the protective effect over time. This concern is especially relevant for indicators that operate independently of technical detection, such as `EXTERNAL` tags. Given users' well-documented tendency to ignore passive indicators [22] and habituate to visual warnings [46], the long-term efficacy of these signals remains questionable. Especially long-term and group-specific effectiveness requires careful evaluation in the future [11].

This leads to a central insight: while in-situ anti-phishing measures can help delay compromise in small-scale campaigns, they are not sufficient on their own. These interventions may buy time and increase attacker effort, but they must not be seen as a substitute for robust technical controls. Our findings reinforce that employees are an inconsistent line of defense—and that effective phishing protection depends on strong, system-level safeguards.

5.3 Measurements in Phishing Simulations

The range of observed click and login rates between 21.5% to 43.8% in Study I and 18.2% to 39.7% in Study II, respectively, reveals nearly twofold differences [75]. This variability underscores the importance of systematically assessing phishing email difficulty, as emphasized by frameworks like the NIST Phish Scale [20, 71].

However, our results suggest that heuristic approaches like cue-counting, as used in the Phish Scale, have strong limitations. Even subtle cue variations produced large shifts in behavior, indicating the need for empirically grounded risk models [72]. Controlled experiments like ours offer a foundation for such models, enabling more accurate email profiling and risk assessment.

Our findings highlight the need for caution when interpreting changes in click or login rates as evidence of improved phishing awareness—a common practice in both research and real-world assessments [36, 37]. While prior studies acknowledge topic-based variation [36, 40], our results show that even within a single topic (e.g., payroll or Outlook), small changes in presentation, timing, or affective framing can shift behavior by up to 21 percentage points. Even seemingly minor factors, like time of day, had measurable effects. This makes direct comparisons across emails over time unreliable unless email properties are carefully controlled.

Furthermore, our results challenge the validity of using click rates as a valid proxy for risky behavior, an issue noted in prior research [40, 70]. In particular, while we observed high click-to-login conversion rates (min = 79%, max = 91%) with a small standard deviation (4%) in Study I, we, in contrast, observed much smaller click-to-login rates (min = 46%, max = 88%) with a much larger standard deviation (12%) in Study II. This means that even under "perfect" conditions, as in Study I, click rates only provide a very rough estimator at best. Meanwhile, click rates can also dramatically overestimate any harmful action, as seen in Study II. This discrepancy is especially important when evaluating the efficacy of anti-phishing interventions. For example, a click-to-login rate of 46% suggests that users may still open phishing links but choose not to log in, possibly because the intervention did not prevent the click but did raise enough suspicion to stop further engagement. While drive-by downloads are often cited to justify measuring clicks [32, 72, 84], their actual risk is negligible in modern, well-managed IT environments compared to human error [15].

Moreover, our findings highlight the risks of relying on aggregated metrics. For example, *Nursing & Functional* and *Medical* staff showed divergent susceptibility to the same email features, yet the larger size of the *Nursing & Functional* group disproportionately influenced overall effects: *Day-of-week* timing showed overall significant effects, although the *Medical* group responded with a null effect. This kind of variation exposes a key limitation in common phishing research practices [36, 37, 40, 64]: they obscure group-specific vulnerabilities. If we wish to uncover risks and develop effective, targeted interventions, researchers and practitioners must move beyond aggregate analyses, especially in environments with a heterogeneous workforce.

In conclusion, we strongly recommend that future studies—and any organization conducting phishing simulations—collect more granular behavioral outcomes to enable accurate comparisons and improve the robustness of susceptibility assessments. Consistent with recent work [40, 49], our findings also highlight the importance of using methods that support causal inference, such as randomized controlled designs, to validly track changes over time. However, these approaches require statistical and methodological expertise and may be difficult to implement in resource-constrained settings.

5.4 Emotional Response to the Simulation

Our findings suggest that phishing simulations can provoke substantial emotional responses among hospital staff. While participation in the affective response survey was voluntary and not representative of the entire workforce, the absolute numbers are substantial: between 100 and 300 employees reported moderate to extreme feelings of fear, shame, guilt, or hostility in response to the simulation. These results must be viewed in light of the broader working conditions in healthcare, where EU health and social care workers report the highest levels of psychosocial stress—driven by chronic time pressure and workload [24]. In this context, any additional emotional strain introduced by phishing simulations deserves serious attention. Prior research suggests that negative emotions such as fear or shame can impair security behavior [68].

That said, our data also show a tendency for employees to report feeling scared more often than hostile. Since we did not ask about positive emotions, it remains possible that some staff viewed the simulation as neutral or even beneficial—a pattern observed in other sectors [48, 67, 68]. Additionally, prior work suggests that employee sentiment often improves over time with repeated exposure [67].

Still, even in relative terms, a 25% hostility rate in large organizations implies that CISOs and IT must be prepared to manage a potentially substantial volume of complaints and concerns. In our case, multiple employees contacted both researchers and the CISO team with ethical objections and resistance to the simulation. These responses require time, communication, and support—resources that must be planned for alongside technical implementation and training [14]. Running the simulation itself required extensive coordination with internal stakeholders, including the CISO, staff councils, data protection officers, and hospital leadership. Researchers and practitioners aiming to study this domain must be prepared not only to justify their efforts to hospital staff but also to commit to what is often a lengthy and demanding process.

6 Conclusions

This study presents the first large-scale phishing simulation in a European Union hospital, showing that even small campaigns have a high likelihood of success and remain a risk for several days. Susceptibility varied across staff groups, with notable differences in response to both email features and anti-phishing interventions. Of the 11 human-centered interventions tested, explicit warning banners and automated spam filtering were most effective. Disabling links and active warning pages showed mixed but promising effects. In contrast, `[EXTERNAL]` tagging showed less reliable or no effects. Overall, our results underscore the need for strong system-level safeguards rather than relying on the anti-phishing interventions commonly available in enterprise email products. Finally, many hospital staff reacted to the phishing simulation with fear, shame, guilt, and even hostility, underscoring the need for careful balancing of the supposed benefits against potential psychological costs.

Acknowledgments

We would like to express our sincere gratitude to the participating hospital, particularly the CISO team, for making this study possible. We thank the IT department, staff councils, and data protection officers for resolving technical, legal, and ethical issues. Above all, we are deeply grateful to the hospital staff, whose participation enabled us to gain a deeper understanding of phishing susceptibility. We also want to thank the reviewers and our shepherd for their incredibly valuable feedback, which helped us improve this paper. This research was supported by the German Federal Ministry of Health with grant number ZMI1-2521FSB801.

References

[1] Hervé Abdi. 2010. Holm's sequential Bonferroni procedure. *Encyclopedia of research design* 1, 8 (2010), 1–8.

[2] Saeed Abu-Nimeh, Dario Nappa, Xinlei Wang, and Suku Nair. 2007. A Comparison of Machine Learning Techniques for Phishing Detection. In *Proceedings of the Anti-phishing Working Groups 2nd Annual Ecrime Researchers Summit.* 60–69.

[3] Linda H. Aiken, Walter Sermeus, Martin McKee, Karen B. Lasater, Douglas Sloane, Colleen A. Pogue, Dorothea Kohnen, Simon Dello, Claudia B. Bettina Maier, Jonathan Drennan, Matthew D. McHugh, and Magnet4Europe Consortium. 2024. Physician and Nurse Well-Being, Patient Safety and Recommendations for Interventions: Cross-Sectional Survey in Hospitals in Six European Countries. *BMJ Open* 14, 2 (2024), e079931.

[4] Dari Alhuwail, Eiman Al-Jafar, Yousef Abdulsalam, and Shaikha AlDuaij. 2021. Information Security Awareness and Behaviors of Health Care Professionals at Public Health Care Facilities. *Applied Clinical Informatics* 12, 04 (2021), 924–932.

[5] Eric P. Allman, Jon Callas, Jim Fenton, Miles Libbey, Michael Thomas, and Mark Delany. 2007. DomainKeys Identified Mail (DKIM) Signatures. RFC 4871. https://rfc-editor.org/rfc/rfc4871.txt

[6] Saad Altamimi, Karen Renaud, and Timothy Storer. 2020. "I Do It Because They Do It": Social-Neutralisation in Information Security Practices of Saudi Medical Interns. In *Risks and Security of Internet and Systems.* 227–243.

[7] Jessica S. Ancker, Alison Edwards, Sarah Nosal, Diane Hauser, Elizabeth Mauer, and Rainu Kaushal. 2017. Effects of Workload, Work Complexity, and Repeated Alerts on Alert Fatigue in a Clinical Decision Support System. *BMC Medical Informatics and Decision Making* 17, 1 (2017).

[8] John Antonakis, Samuel Bendahan, Philippe Jacquart, and Rafael Lalive. 2010. On making causal claims: A review and recommendations. *The Leadership Quarterly* 21, 6 (2010), 1086–1120.

[9] American Political Science Association (APSA). 2020. Principles and Guidance for Human Subjects Research. https://connect.apsanet.org/hsr/principles-and-guidance/

[10] Vincent Arel-Bundock, Noah Greifer, and Andrew Heiss. 2024. How to Interpret Statistical Models Using marginaleffects for R and Python. *Journal of Statistical Software* 111, 9 (2024), 1–32.

[11] Shahryar Baki and Rakesh M. Verma. 2023. Sixteen Years of Phishing User Studies: What Have We Learned? *IEEE Transactions on Dependable and Secure Computing* 20, 2 (2023), 1200–1212.

[12] Oskar Braun, Jan Hörnmann, Norbert Pohlmann, Tobias Urban, and Matteo Große-Kampmann. 2025. Different Seas, Different Phishes – Large-Scale Analysis of Phishing Simulations across Different Industries. In *Proceedings of the 20th ACM Asia Conference on Computer and Communications Security.*

[13] Bianka Breyer and Matthias Bluemke. 2016. *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel).*

[14] Lina Brunken, Annalina Buckmann, Jonas Hielscher, and M Angela Sasse. 2023. "To Do This Properly, You Need More Resources": The Hidden Costs of Introducing Simulated Phishing Campaigns. In *Proceedings of the 32nd USENIX Security Symposium.*

[15] Kang Leng Chiew, Kelvin Sheng Chek Yong, and Choon Lin Tan. 2018. A Survey of Phishing Attacks: Their Types, Vectors and Technical Approaches. *Expert Systems with Applications* 106 (2018), 1–20.

[16] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2 ed.). Academic Press.

[17] Nicholas Counter. 2025. Health Care Workers Across Germany Go on Strike. https://www.dw.com/en/health-care-workers-across-germany-go-on-strike/a-71844915. Accessed: 2025-07-27.

[18] Lynne Coventry, Dawn Branley-Bell, Elizabeth Sillence, Sabina Magalini, Pasquale Mari, Aimilia Magkanaraki, and Kalliopi Anastasopoulou. 2020. Cyber-Risk in Healthcare: Exploring Facilitators and Barriers to Secure Behaviour. In *Proceedings of the 2nd International Conference on HCI for Cybersecurity, Privacy and Trust.* 105–122.

[19] Maria Cvach. 2012. Monitor Alarm Fatigue: An Integrative Review. *Biomedical Instrumentation and Technology* 46, 4 (2012), 268–277.

[20] Shanee Dawkins and Jody Jacobs. 2023. *NIST Phish Scale User Guide.* Technical Report NIST TN 2276. National Institute of Standards and Technology (U.S.). NIST TN 2276 pages.

[21] Giuseppe Desolda, Lauren S. Ferro, Andrea Marrella, Tiziana Catarci, and Maria Francesca Costabile. 2021. Human Factors in Phishing Attacks: A Systematic Literature Review. *Comput. Surveys* 54, 8 (2021), 173:1–173:35.

[22] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 2008. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 1065–1074.

[23] Elizabeth V. Eikey, Alison R. Murphy, Madhu C. Reddy, and Heng Xu. 2015. Designing for Privacy Management in Hospitals: Understanding the Gap between User Activities and IT Staff's Understandings. *International Journal of Medical Informatics* 84, 12 (2015), 1065–1075.

[24] Flash Eurobarometer. 2022. *OSH Pulse - Occupational Safety and Health in Post-Pandemic Workplaces.* Report. European Agency for Safety and Health at Work.

[25] European Commission & European Parliament. 2016. Directive (EU) 2016/1148 on security of network and information systems.

[26] European Commission, Directorate-General for Communication. 2025. European Action Plan on the Cybersecurity of Hospitals and Healthcare Providers. https://commission.europa.eu/cybersecurity-healthcare_en.

[27] European Parliament & Council of the European Union. 2022. Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2).

[28] European Union Agency for Cybersecurity. 2024. ENISA Threat Landscape 2024. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024.

[29] European Union Agency for Cybersecurity (ENISA). 2024. *NIS Investments 2024: Cybersecurity Policy Assessment.* Technical Report TP-01-24-001-EN-NN. ENISA.

[30] Luiza Fabisiak and Tomasz Hyla. 2020. Measuring Cyber Security Awareness within Groups of Medical Professionals in Poland. In *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS).* 3871–3880.

[31] Ana Ferreira and Soraia Teles. 2019. Persuasion: How Phishing Emails Can Influence Users and Bypass Security Measures. *International Journal of Human-Computer Studies* 125 (2019), 19–31.

[32] Anjuli Franz, Verena Zimmermann, Gregor Albrecht, Katrin Hartwig, Christian Reuter, Alexander Benlian, and Joachim Vogt. 2021. SoK: Still Plenty of Phish in the Sea — A Taxonomy of User-Oriented Phishing Interventions and Avenues for Future Research. In *Proceedings of the 17th Symposium on Usable Privacy and Security (SOUPS).* 339–358.

[33] Deutsche Gesellschaft für Psychologie. 2018. *Ethisches Handeln in der psychologischen Forschung: Empfehlungen der Deutschen Gesellschaft für Psychologie für Forschende und Ethikkommissionen* (1. Auflage ed.). Hogrefe.

[34] Deutschen Gesellschaft für Soziologie and Berufsverbandes Deutscher Soziologinnen und Soziologen. 2017. Ethik-Kodex der Deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologinnen und Soziologen (BDS). https://soziologie.de/dgs/ethik/ethik-kodex

[35] Google. 2025. Advanced phishing and malware protection. https://web.archive.org/web/20250321123725/https://support.google.com/a/answer/9157861 Accessed: 2025-03-23.

[36] William J. Gordon, Adam Wright, Ranjit Aiyagari, Leslie Corbo, Robert J. Glynn, Jigar Kadakia, Jack Kufahl, Christina Mazzone, James Noga, Mark Parkulo, Brad Sanford, Paul Scheib, and Adam B. Landman. 2019. Assessment of Employee Susceptibility to Phishing Attacks at US Health Care Institutions. *JAMA Network Open* 2, 3 (2019), e190393.

[37] William J Gordon, Adam Wright, Robert J Glynn, Jigar Kadakia, Christina Mazzone, Elizabeth Leinbach, and Adam Landman. 2019. Evaluation of a Mandatory Phishing Training Program for High-Risk Employees at a US Healthcare System. *Journal of the American Medical Informatics Association* 26, 6 (2019), 547–552.

[38] Ying He, Aliyu Aliyu, Mark Evans, and Cunjin Luo. 2021. Health Care Cybersecurity Challenges and Solutions Under the Climate of COVID-19: Scoping Review. *Journal of Medical Internet Research* 23, 4 (2021), e21747.

[39] Cormac Herley. 2009. So long, and no thanks for the externalities: the rational rejection of security advice by users. In *Proceedings of the 2009 Workshop on New Security Paradigms Workshop.* 133–144.

[40] Grant Ho, Ariana Mirian, Elisa Luo, Khang Tong, Euyhyun Lee, Lin Liu, Christopher A. Longhurst, Christian Dameff, Stefan Savage, and Geoffrey M. Voelker. 2025. Understanding the Efficacy of Phishing Training in Practice. In *Proceedings of the 2025 IEEE Symposium on Security and Privacy (SP).* 76–76.

[41] Grant Ho, Aashish Sharma, Mobin Javed, Vern Paxson, and David Wagner. 2017. Detecting Credential Spearphishing Attacks in Enterprise Settings. In *Proceedings of the 26th USENIX Conference on Security Symposium.* 469–485.

[42] Hang Hu and Gang Wang. 2018. End-to-End Measurements of Email Spoofing Attacks. In *Proceedings of the 27th USENIX Security Symposium.* 1095–1112.

[43] Mohammad S Jalali, Maike Bruckes, Daniel Westmattelmann, and Gerhard Schewe. 2020. Why Employees (Still) Click on Phishing Links: Investigation in Hospitals. *Journal of Medical Internet Research* 22, 1 (2020), e16775.

[44] Mohammad S. Jalali and Jessica P. Kaiser. 2018. Cybersecurity in Hospitals: A Systematic, Organizational Perspective. *Journal of Medical Internet Research* 20, 5 (2018), e10059.

[45] Daniel Jampen, Gürkan Gür, Thomas Sutter, and Bernhard Tellenbach. 2020. Don't Click: Towards an Effective Anti-Phishing Training. A Comparative Literature Review. *Human-centric Computing and Information Sciences* 10, 1 (2020), 1–41.

[46] Soyun Kim and Michael S. Wogalter. 2009. Habituation, Dishabituation, and Recovery Effects in Visual Warnings. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 53, 20 (2009), 1612–1616.

[47] D. Scott Kitterman. 2014. Sender Policy Framework (SPF) for Authorizing Use of Domains in Email, Version 1. RFC 7208. https://rfc-editor.org/rfc/rfc7208.txt

[48] Daniele Lain, Tarek Jost, Sinisa Matetic, Kari Kostiainen, and Srdjan Capkun. 2024. Content, Nudges and Incentives: A Study on the Effectiveness and Perception of Embedded Phishing Training. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*. 4182–4196.

[49] Daniele Lain, Kari Kostiainen, and Srdjan Čapkun. 2022. Phishing in Organizations: Findings from a Large-Scale and Long-Term Study. In *2022 IEEE Symposium on Security and Privacy (SP)*. 842–859.

[50] Enze Liu, Lu Sun, Alex Bellon, Grant Ho, Geoffrey M. Voelker, Stefan Savage, and Imani N. S. Munyaka. 2023. Understanding the Viability of Gmail's Origin Indicator for Identifying the Sender. In *Proceedings of the Nineteenth USENIX Conference on Usable Privacy and Security*. 77–96.

[51] L. Maniscalco, M. Enea, N. de Vries, W. Mazzucco, A. Boone, O. Lavreysen, K. Baranski, S. Miceli, A. Savatteri, S. Fruscione, M. Kowalska, P. de Winter, S. Szemik, L. Godderis, and D. Matranga. 2024. Intention to Leave, Depersonalisation and Job Satisfaction in Physicians and Nurses: A Cross-Sectional Study in Europe. *Scientific Reports* 14, 1 (2024), 2312.

[52] Franck Martin, Eliot Lear, Tim Draegen, Elizabeth Zwicky, and Kurt Andersen. 2016. Interoperability Issues between Domain-Based Message Authentication, Reporting, and Conformance (DMARC) and Indirect Email Flows. RFC 7960. https://rfc-editor.org/rfc/rfc7960.txt

[53] Microsoft. 2024. Safe Links in Microsoft Defender for Office 365. https://web.archive.org/web/20240528173533/https://learn.microsoft.com/en-us/defender-office-365/safe-links-about Accessed: 2025-04-03.

[54] Microsoft. 2025. Anti-phishing policies in Microsoft 365. https://web.archive.org/web/20250323024553/https://learn.microsoft.com/en-us/defender-office-365/anti-phishing-policies-about Accessed: 2025-03-23.

[55] Microsoft. 2025. Set-ExternalInOutlook. https://web.archive.org/web/20250304002443/https://learn.microsoft.com/en-us/powershell/module/exchange/set-externalinoutlook?view=exchange-ps Accessed: 2025-04-03.

[56] Douglas C. Montgomery. 2017. *Design and analysis of experiments* (9th ed.).

[57] Barracuda Networks. 2025. Email Warning Banner Messages - Barracuda Email Gateway Defense. https://web.archive.org/web/20250117190326/https://campus.barracuda.com/product/emailgatewaydefense/doc/167977270/email-warning-banner-messages/ Accessed: 2025-01-17.

[58] PANACEA Consortium. 2019–2022. PANACEA Research: People-centric Cybersecurity Toolkit for Healthcare Institutions. https://panacearesearch.eu/.

[59] Kathryn Parsons, Marcus A. Butavicius, Malcolm Robert Pattinson, Dragana Calic, Agata McCormac, and Cate Jerram. 2015. Do Users Focus on the Correct Cues to Differentiate Between Phishing and Genuine Emails?. In *Proceedings of the 2015 Australasian Conference on Information Systems (ACIS)*.

[60] Justin Petelka, Yixin Zou, and Florian Schaub. 2019. Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.

[61] R. L. Plackett and J. P. Burman. 1946. The design of optimum multifactorial experiments. *Biometrika* 33, 4 (1946), 305–325.

[62] Ward Priestman, Tony Anstis, Isabel G. Sebire, Shankar Sridharan, and Neil J. Sebire. 2019. Phishing in Healthcare Organisations: Threats, Mitigation and Approaches. *BMJ Health & Care Informatics* 26, 1 (2019), e100031.

[63] Proofpoint. 2025. URL Defense FAQ's. https://help.proofpoint.com/Proofpoint_Essentials/Email_Security/User_Topics/Targeted_Attack_Protection/URL_Defense_FAQ's Accessed: 2025-04-03.

[64] Fabio Rizzoni, Sabina Magalini, Alessandra Casaroli, Pasquale Mari, Matt Dixon, and Lynne Coventry. 2022. Phishing Simulation Exercise in a Large Hospital: A Case Study. *DIGITAL HEALTH* 8 (2022), 20552076221081716.

[65] Sumantra Sarkar, Anthony Vance, Balasubramaniam Ramesh, Menelaos Demestihas, and Daniel Thomas Wu. 2020. The Influence of Professional Subculture on Information Security Policy Violations: A Field Study in a Healthcare Context. *Information Systems Research* 31, 4 (2020), 1240–1259.

[66] Orvila Sarker, Asangi Jayatilaka, Sherif Haggag, Chelsea Liu, and M. Ali Babar. 2023. A Multi-vocal Literature Review on Challenges and Critical Success Factors of Phishing Education, Training and Awareness. *Journal of Systems and Software* (2023), 111899.

[67] Katharina Schiller, Florian Adamsky, Christian Eichenmüller, Matthias Reimert, and Zinaida Benenson. 2024. Employees' Attitudes towards Phishing Simulations: "It's like When a Child Reaches onto the Hot Hob". In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*. 4167–4181.

[68] Markus Schöps, Marco Gutfleisch, Eric Wolter, and M. Angela Sasse. 2024. Simulated Stress: A Case Study of the Effects of a Simulated Phishing Campaign on Employees' Perception, Stress and Self-Efficacy. In *Proceedings of the 33rd USENIX Security Symposium*. 4589–4606.

[69] William R. Shadish, Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

[70] Teodor Sommestad and Henrik Karlzén. 2019. A Meta-Analysis of Field Experiments on Phishing Susceptibility. In *Proceedings of the 2019 APWG Symposium on Electronic Crime Research (eCrime)*. 1–14.

[71] Michelle P. Steves, Kristen K. Greene, and Mary F. Theofanos. 2019. A Phish Scale: Rating Human Phishing Message Detection Difficulty. In *Proceedings of the Workshop on Usable Security (USEC)*.

[72] Thomas Sutter, Ahmet Selman Bozkir, Benjamin Gehring, and Peter Berlich. 2022. Avoiding the Hook: Influential Factors of Phishing Awareness Training on Click-Rates and a Data-Driven Approach to Predict Email Difficulty Perception. *IEEE Access* 10 (2022), 100540–100565.

[73] Ronnie Taib, Kun Yu, Shlomo Berkovsky, Mark Wiggins, and Piers Bayl-Smith. 2019. Social Engineering and Organisational Dependencies in Phishing Attacks. In *Human-Computer Interaction – INTERACT 2019*. 564–584.

[74] Chuan (Annie) Tian, Matthew L. Jensen, and Alexandra Durcikova. 2023. Phishing Susceptibility across Industries: The Differential Impact of Influence Techniques. *Computers & Security* 135 (2023), 103487.

[75] Jan Tolsdorf, David Langer, and Luigi Lo Iacono. 2025. *Supplementary Materials for the Paper "Phishing Susceptibility and the (In-)Effectiveness of Common Anti-Phishing Interventions in a Large University Hospital"*. doi:10.5281/zenodo.17014954

[76] Jan Tolsdorf and Luigi Lo Iacono. 2024. Expert Perspectives on Information Security Awareness Programs in Medical Care Institutions in Germany. In *Proceedings of the 6th International Conference on HCI for Cybersecurity, Privacy and Trust*. 98–117.

[77] Amber Van Der Heijden and Luca Allodi. 2019. Cognitive Triaging of Phishing Attacks. In *Proceedings of the 28th USENIX Conference on Security Symposium*. 1309–1326.

[78] Anthony Vance, Brock Kirwan, Daniel Bjornn, Jeffrey Jenkins, and Bonnie Brinton Anderson. 2017. What Do We Really Know about How Habituation to Warnings Occurs over Time?: A Longitudinal fMRI Study of Habituation and Polymorphic Warnings. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2215–2227.

[79] Arun Vishwanath, Tejaswini Herath, Rui Chen, Jingguo Wang, and H. Raghav Rao. 2011. Why Do People Get Phished? Testing Individual Differences in Phishing Vulnerability within an Integrated, Information Processing Model. *Decision Support Systems* 51, 3 (2011), 576–586.

[80] Melanie Volkamer, Karen Renaud, and Paul Gerber. 2016. Spot the Phish by Checking the Pruned URL. *Information & Computer Security* 24, 4 (2016), 372–385.

[81] Melanie Volkamer, Karen Renaud, Benjamin Reinheimer, and Alexandra Kunz. 2017. User Experiences of TORPEDO: TOoltip-poweRed Phishing Email DetectiOn. *Computers & Security* 71 (2017), 100–113.

[82] Melanie Volkamer, Martina Angela Sasse, and Franziska Boehm. 2020. Analysing Simulated Phishing Campaigns for Staff. In *Computer Security*. 312–328.

[83] Melissa L. H. Võ, Markus Conrad, Lars Kuchinke, Karolina Urton, Markus J. Hofmann, and Arthur M. Jacobs. 2009. The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods* 41, 2 (2009), 534–538.

[84] Rick Wash and Molly M. Cooper. 2018. Who Provides Phishing Training? Facts, Stories, and People Like Me. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[85] David Watson, Lee Anna Clark, and Auke Tellegen. 1988. Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology* 54, 6 (1988), 1063–1070.

[86] Emma J. Williams, Joanne Hinds, and Adam N. Joinson. 2018. Exploring Susceptibility to Phishing in the Workplace. *International Journal of Human-Computer Studies* 120 (2018), 1–13.

[87] Kazuo Yamaguchi. 1991. *Event history analysis*. SAGE Publications, Inc.

[88] Weining Yang, Aiping Xiong, Jing Chen, Robert W. Proctor, and Ninghui Li. 2017. Use of Phishing Training to Improve Security Warning Compliance: Evidence from a Field Experiment. In *Proceedings of the Hot Topics in Science of Security: Symposium and Bootcamp*. 52–61.

[89] Liu Hua Yeo and James Banfield. 2022. Human Factors in Electronic Health Records Cybersecurity Breach: An Exploratory Analysis. *Perspectives in Health Information Management* 19, Spring (2022), 1i.

[90] Sijie Zhuo, Robert Biddle, Yun Sing Koh, Danielle Lottridge, and Giovanni Russello. 2023. SoK: Human-centered Phishing Susceptibility. *ACM Transactions on Privacy and Security* 26, 3 (2023), 1–27.

[91] Olga A. Zielinska, Allaire K. Welk, Christopher B. Mayhorn, and Emerson Murphy-Hill. 2016. A Temporal Analysis of Persuasion Principles in Phishing Emails. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60, 1 (2016), 765–769.