A Lot Less Likely Than I Thought: Introducing Evidence-Based Security Risk Assessment for Healthcare Software

Charles Weir
Computing & Communications
Lancaster University
Lancaster, UK
c.weir1@lancaster.ac.uk

Anna Dyson
Computing & Communications
Lancaster University
Lancaster, UK
a.dyson@lancaster.ac.uk

Dan Prince
Computing & Communications
Lancaster University
Lancaster, UK
d.prince@lancaster.ac.uk

Abstract— **Security and privacy are particularly important for health applications and health-related devices. So, it is vital that health software developers, especially in small to medium companies, devote their time and resources only to the security and privacy activities that will be most effective for them. Accordingly, this paper describes the creation and development of a facilitated workshop to help developers create risk assessments, using a structured series of activities based on a healthcare industry risk model. The authors found little publicly available information on risk probabilities, requiring our own calculations. The results of six workshop trials showed that cards with stories and probabilities promoted effective risk analysis, and that this was valuable to less experienced development teams. This workshop approach provides a powerful lightweight approach to calculating evidence-based security and privacy loss expectations, allowing better decision making to improve the security of the many healthcare software systems we all depend upon.**

Keywords— Software security, health software, IoT, software developer, cybersecurity, workshop, Design Based Research, developer centered security.

I. Introduction

Software security and privacy are critical in health-related devices and applications [1], [2]. While there are many aspects to security and privacy for such Health Internet of Things (HIoT) devices, the development process used to create the software clearly must have a significant impact. Three industry trends contribute to this impact: the increasing connection of smart healthcare devices directly via the internet; the DevOps movement; and the increasing use of microservices and Software as a Service (SaaS) components in the cloud-based services that communicate with such devices. All three trends require security to be 'in the code' rather than being the responsibility of separate operations or security teams. So, development teams—product owners, managers, developers, testers—must themselves be effective at creating secure software.

To achieve this effectiveness, the teams need to be able to *focus* their time and expenditure on the *most important* aspects of security and privacy for their projects: the aspects that matter for their clients and stakeholders. Currently their focus is often misplaced: many development teams spend effort addressing threats that are either too unlikely to be important, or irrelevant to their clients and stakeholders; they neglect more important problems [3]. Even cybersecurity experts and threat modelling may not be helpful in identifying an effective focus: cybersecurity experts can be less successful than managers at identifying the important security and privacy problems to address [4].

Yet the techniques to identify the appropriate problems to address are already well-known. Risk management is a professional discipline [5], with widely used practices to compare different kinds of risk by calculating 'loss expectation' values for each one [6]. Threat assessment for cybersecurity has also received a great deal of attention [7], [8]. Indeed, risk assessment *for safety* is already an essential part of medical device development and is mandated by many health safety standards [9].

Two problems prevent health development teams, especially in Small to Medium Enterprises (SMEs), from using risk assessment techniques to prioritize their security and privacy work. First, development teams may be unaware of the techniques, or not know how to carry them out without professional support. Second, there is very little publicly available information about the security and privacy risks relevant to the health industry. Nor is statistical data available about the probability of such risks; such information is treated as commercially sensitive, and public sharing of it is rare—in contrast to widespread public sharing of *vulnerability* data. This means that a health SME development team is unlikely to have access to such information.

But what if such information could be made available? What if we could find a way to work with SME HIoT development teams, to use it to improve the security and privacy of their development? The authors of this paper undertook a project to address that possibility. The main research question explored by that project was therefore:

RQ1 *How can industry-based cybersecurity data improve security and privacy aspects of the development of Health IoT systems within resource constrained development teams?*

This paper, therefore, describes the creation of a facilitated workshop, for health-based development teams, to use such data to improve their security and privacy. In the workshops, the participants identify security and privacy risks, then estimate their organization's loss expectation for each, using banding techniques and objective industry likelihood figures. From the results, the development teams can help stakeholders make evidence-based decisions where and how much effort to

This work has been supported by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has been funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/S035362/1.

spend addressing different security and privacy threats. Workshop participants can also take the materials and repeat the workshop independently as part of their normal development.

Using those loss expectation figures, developers—especially product managers—can compare the financial benefit of security and privacy work with the financial benefits of other product enhancements, giving them a solid basis for their decisions.

The facilitated workshop was trialled on a range of development teams working on health-related—where possible, HIoT—products. Using the Design Based Research methodology (Section IV.A), we carried out two cycles of trials, using findings from the first cycle to improve the workshop for the second cycle. We surveyed all participants after each workshop.

The novelty of this approach is in the provision of a specific (health) industry risk model, including information about the probability of such risks based on objective statistical data. Indeed, assessing such high-level risks is itself rare, in contrast to the popularity of threat modelling techniques to find technical vulnerabilities.

The rest of this paper is as follows. Section II discusses the health devices domain and relevant past research; Section III describes the requirements for the workshop package and how they were implemented. Section IV describes the research method and introduces research sub-questions. Section V details and discusses the results from using the workshop; and Section VI provides a summary conclusion.

II. Background

Research related to development teams and risk-based analysis for secure and health software has taken a variety of different approaches. This section introduces the domain of Health IoT software, then explores existing research into risk-oriented and security-oriented activities.

A. Health IoT Software

Health IoT software powers or aids the managing and monitoring of IoT devices related to human health. Two surveys explore the topic [10], [11]. Although IoT devices include both sensors that collect data, and actuators that control physical devices [12], many typical HIoT applications relate only to sensor devices [13]. These might monitor aspects of individual health or safety, ranging from tracking running activity to heart function and wheelchair management [13]. Actuators, such as implantable cardiac devices, are also deployed [13]. A range of large suppliers offer infrastructure to support both communication and data analysis, and the medical aspects are heavily legislated, with notable differences between different jurisdictions [13].

B. Risk Management

Cyber Risk Management in the software lifecycle is the process of identifying, analyzing, evaluating and addressing an organization's cyber security threats [14]. The processes required to do this at a corporate level are now mature, and a variety of competing standards, such as ISO2001, the Payment Card Industry Data Security Standard (PCI-DSS) and, in the UK, Cyber Essentials each provide extensive prescriptions of checks and activities to carry out [14]. Academic work on the subject has included the quantification of cyber risks [6], and the evaluation of the effectiveness of particular standards, e.g. [15]. Surveys highlight overconfidence and lack of resources as particular problems in SMEs [16].

In this research, we are interested in the aspects of risk management that relate to software development product decisions—the processes of identifying threats and problems, and of estimating impact and likelihood. We shall refer to that as 'cyber risk assessment'. Unfortunately, surprisingly little information is available publicly about the probability of cyber risks. In addition, a major challenge in the use of risk-driven security metrics is the lack of evidence for their effectiveness at improving security in the early phases of product development and risk analysis, when the needs for it are at their greatest [17].

C. Development Team Activities

Surprisingly, given academia's domination of the education space, there is relatively little literature on activities to help improve the security behavior of software developers.

In two early case studies, first a single penetration testing session and workshop failed to have much effect on a distributed development team [18]. Second, working with a team to *challenge and teach [developers] about security issues of their product* also proved unsatisfactory, due to the pressure to add functionality [19].

'Security Patterns' offered another approach, though the benefits proved inconclusive [20]. A recent book by Bell et al. [21] provides support for developers and tool recommendations, containing much valuable practitioner experience, but little objective assessment of the advice it provides.

One promising approach is to *raise developers' security awareness*, such as by using discussions about security [22]. Another is to use structured workshops to teach the importance of effective decision making, threat assessment, and suitable presentation of the results [3]. Some researchers have had success with threat assessment workshops using less conventional approaches, such as design fiction [23].

D. Implications

Section B suggests that it will be valuable for development teams to include risk management and security requirements techniques in their development processes. But Section C indicates that so far there has been little work on approaches to help them do so. This implies that an approach to teach practical, inexpensive, and accurate risk-based threat assessment for development teams will offer a valuable contribution.

The remainder of this paper describes the design of such an approach, some ways used to get objective HIoT industry information, approaches to present it effectively to HIoT developers, and the development, trials and assessment of a facilitated workshop to do so.

III. Design of the Facilitated Workshops

To create and trial the workshop, we needed: first 'objective HIoT industry risk information', next a structure and design for the workshop, and finally trials with a range of health development teams. We explore these in the next sections.

A. Estimation of Health Industry Risk Likelihoods

In gathering 'objective HIoT industry information', we wanted to establish the likelihood of each kind of security and

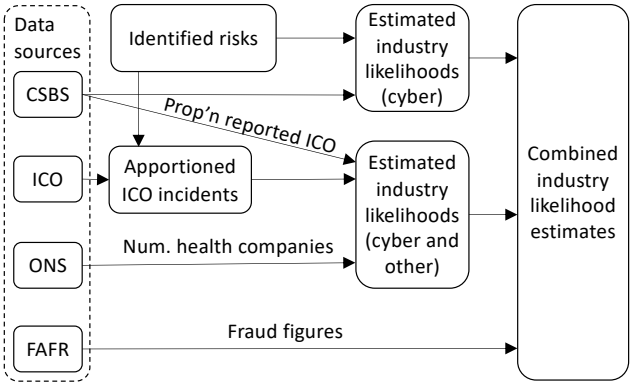


FIGURE 1: HIOT INDUSTRY INFORMATION CALCULATION

privacy risks happening to an SME organization—using the term 'likelihood' to mean the probability of an incident in a given period. We used two main data sources. First, the UK Cyber Security Breaches Survey 2022 (CSBS) provides a dataset including over 400 interviews with randomly selected health sector companies [24]. Second, the UK Information Commissioner's Office (ICO) publishes data, including health-sector figures, about the major security and privacy incidents that companies must by law report to them [25]. Further data came from the Office of National Statistics (ONS) report on UK Business [26], and from the UK Finance Annual Fraud Report [27].

Figure 1 shows an overview of the calculation. From the four datasets and the experience of the researchers and colleagues, we constructed a set of 'identified risks': types of cyber-related problems, such as 'Random Ransomware' and 'Leaked Encryption Keys'.

From the CSBS, we extracted Health organisation reports of security and privacy breaches for each of the survey's range of eleven cyber incident types (such as *Computers becoming infected with ransomware*), considering only the reports where the company reported suffering financial loss. We then mapped each identified risk to one of these types, and estimated the likelihood of each risk on the assumption that all identified risks in each cyber incident type were equally likely and the only possibilities in their category. From that survey, we also extracted the proportion of all companies and charities who suffered losses who reported an incident to the ICO (around 2%).

The ICO data includes both cyber (i.e. criminal adversarial) incidents and non-cyber incidents, and categorizes them separately into 8 and 15 categories respectively. Using dual coding, we mapped our identified risks into these ICO categories, and estimated the number of health industry incidents of each risk assuming that all identified risks in a given category were equally represented. We then estimated the likelihood of each identified risk causing loss in the wider population, by scaling these numbers according to the proportion of companies who reported an incident to the ICO (from the CSBS report) and the number of health companies in the UK (from the ONS report).

This gave us two estimates of likelihood for cyber risks: one based primarily on ICO and one based on CSBS. Given that, as shown in Figure 2, the distribution of the calculated likelihoods was closer to logarithmic than linear, we combined

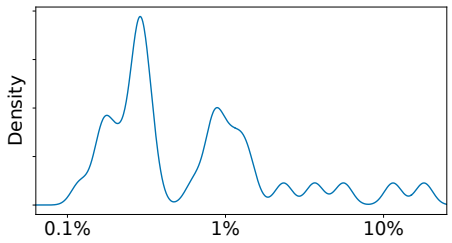


FIGURE 2: DENSITY LOG PLOT: ANNUAL LIKELIHOOD OF RISK TYPES

them using their logarithmic mean[1]. For non-cyber risks the process gave us only one estimate of likelihood: based on ICO data scaled using the CSBS's reporting figure above.

There are limitations to this approach: the granularity of the classifications in each report made assigning identified risks to them difficult: Cohen's Kappa was 0.5 between two coders for the initial ICO categorization. The assumption that all identified risks of each incident type are equally likely and the only possibilities in their category is questionable, as is the assumption that the proportion of all incidents reported to the ICO is constant across ICO categories. The assumption that different organizations are sufficiently similar for average likelihood figures to be meaningful can also be challenged. The only claim we make for this approach is that the results provide objective risk data.

Given these limitations and the huge differences we expected in likelihoods for a given identified risk between different companies and products, we mitigated against these issues leading to erroneous conclusions by avoiding suggesting an inappropriate precision. Figure 2, which plots relative density against (logarithm of) annual likelihood, shows that the calculated likelihoods were distributed exponentially. What appears important, therefore, is only their *order of magnitude*. We therefore logarithmically rounded[2] the results to the nearest order of magnitude (10%, 1%, 0.1% per year), to give an indication of the *scale* of each likelihood, and categorized the resulting values, as shown in Table 1.

We carried out the above analysis using a combination of Microsoft Excel spreadsheets and Python on Jupyter notebooks [28].

TABLE 1: EXAMPLE STORIES AND LIKELIHOODS

Category	L'hood	Story
Common	10%	Embarrassing Journalist
Common	10%	Misdirected communication
Common	10%	Shared Login
Common	10%	Random Ransomware
Common	10%	Helpful Employee
Infrequent	1%	Toe-rag hacker
Infrequent	1%	Targeted Ransomware
Infrequent	1%	Leaked Encryption Keys
Infrequent	1%	Disgruntled Employee
Rare	0.1%	Cyber Hijack
Rare	0.1%	WiFi Wireless Intercept
Rare	0.1%	Denial of Service Malware
Rare	0.1%	Crypto Mining
Rare	0.1%	Professional Fraudsters

[1] The logarithmic mean of x and y is $10^{\frac{\log(x)+\log(y)}{2}}$.

[2] Logarithmic rounding for x is $10^{\lfloor \log(x)+\frac{1}{2} \rfloor}$.

B. Workshop Design

To design the workshop, we first identified a set of requirements. First, a recent study of SMEs in the Health and HIoT arena provided evidence how we might best communicate with HIoT developers [29], specifically:

Requirement 1. Assume a working understanding of the terms 'security' and (usually) 'privacy';

Requirement 2. Expect, but not rely on, some knowledge of risk-based threat assessment;

Requirement 3. Assume a need for security or privacy from their customers (but not that either may necessarily be a sales point);

Requirement 4. Motivate security in terms of compliance with existing safety and privacy standards;

Requirement 5. Avoid, or take great care with, terms such as 'threat', 'threat actor' and 'victim';

Requirement 6. Use stories to express cyber 'threats' in an easily understandable way; and

Requirement 7. Avoid approaches that over-simplify the complexity of decision-making, such as providing 'how to decide' instructions.

As already described, a key element of the workshop was the introduction of industry risk and probability data around threats (Section III.A), leading to:

Requirement 8. Convey to participants both the set of cyber-related threats to consider, and their relative probabilities.

We also identified general requirements for any activity to help improve developer security, specifically that it should:

Requirement 9. Take less than one working day for a development team to carry out, to keep costs acceptable;

Requirement 10. Work with development teams, since a majority of developers work in teams [30];

Requirement 11. Work without security specialists, since many teams do not have access to them [31];

Requirement 12. Work without product managers present in the workshops, since while it is obviously a benefit to include them, in many cases they may not be available or persuaded to attend;

Requirement 13. Support developers currently using few or no assurance techniques, since many teams do not currently use them [31]; and

Requirement 14. Be leadable by non-researchers, to permit the use of the workshop where the researchers are unavailable.

With the increase in demands on companies following the end of COVID lockdown, we found a new reluctance from companies to devote the time to the workshops; we therefore further tightened Requirement 9 to aim for a half-day workshop.

We addressed Requirement 1 and Requirement 2 with a short introduction teaching session for the workshop. We used developer language, rather than security specialist language, in the workshop materials (Requirement 5); in particular, we talked of 'risks' rather than 'threats'.

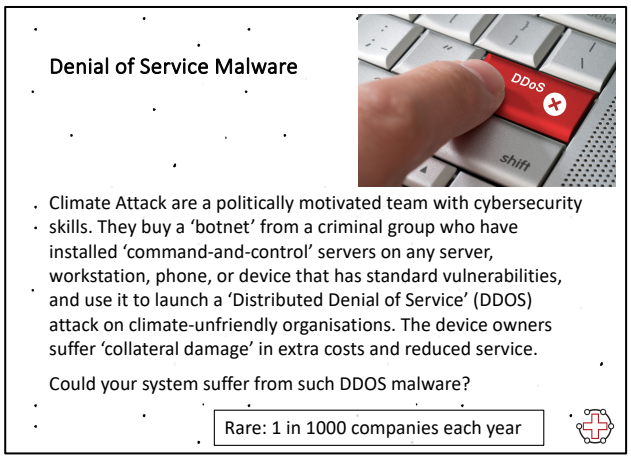


FIGURE 3: EXAMPLE RISK CARD

To communicate the identified risks to developers (Requirement 8) we used 'risk cards' with incidents expressed as named stories (Requirement 6), including stories around compliance (Requirement 4), and the impact on customers (Requirement 3). We illustrated the risk cards and added likelihood information both as text and as a density of dots on the card margins. To encourage participants to treat the stories as exemplars to encourage their own ideation, we added an open question to each card. Table 1 shows likelihoods and story names, with shading to distinguish likelihood values (order of magnitude probability of an incident of that kind within a year). Figure 3 shows an example card.

To prompt long-term benefits from the workshop, as a final step we asked participants an open question to discuss: how could they incorporate the approach into their own projects (Requirement 7).

The remaining requirements we addressed as follows. To address the revised Requirement 9 (less than half a working day), we 'time boxed' the work identifying and quantifying risks and created both online and in-person versions of the materials. For Requirement 10 (working with teams) in some workshops we had teams of developers attend the workshops to discuss their own projects. In workshops where the attendees were from different teams, we used a case study project based on an industry exemplar [32]. For Requirement 11 (avoiding security specialists), Requirement 12 (no product managers) and Requirement 13 (for developers using few assurance techniques) we kept discussions and outputs away from technical security knowledge and activities. To help address Requirement 14 (leadable by non-researchers), we provided the materials, with full instructions, for participants to use themselves in follow-up sessions.

The workshops, therefore, had three stages, each with a different activity:

1. Agreement on impact thresholds—between low, medium and high losses—for their project, or for the case study (about 30 minutes);
2. Examining each risk card in turn, and ideating corresponding risks for the project or case study, along with impact levels to create a 'risk landscape' document (1-2 hours); and
3. Discussing the integration of such a risk assessment process into their own development (30 minutes)

Participants recorded the results of the first two activities on a shared 'Risk Landscape' document. More details of the

workshop structure, and instructions how to carry them out, are available [33].

C. Workshop Trials

For the trials, we wanted workshops with SME health IoT software developers. We used two kinds of trial: private workshops with single teams working on particular HIoT projects; and public workshops aimed at health SMEs. To recruit participants, we used three sources: healthcare contacts established during earlier research; the university business network; and open advertisement through Eventbrite and social media.

The motivation for delegates was to learn better ways improve their development process and the security and privacy of their products. The public workshops also included an initial speaker and food to encourage attendance; no other compensation was provided. Some of the trials were online, some in person, requiring us to use both physical and virtual versions of the workshop materials. All participants signed research consent forms.

Each workshop had at least three participants in addition to a leader and ethnographer from the research team (see Section IV.C), and included a break at about half time. The leader introduced each activity with a short slide presentation. For the second activity, ideation of risks, we provided post-it notes and pens (or online equivalents) and encouraged participants to write down possible risks individually, then to discuss them before adding some to the 'risk landscape'.

The research was approved by the Lancaster University Faculty of Science and Technology Ethics Committee.

IV. Research Methodology

To evaluate the workshops, we used a pragmatic, rather than interpretive, approach, focusing on the knowledge we gained and the possibilities for resulting action [32]. Our main purpose was to support a potentially large number of developers in improving security, and therefore to build on previous research and discoveries [34]. We chose Design-Based Research (DBR) as our methodology for the project [35]. We had used it to develop workshops in earlier projects, and therefore this project provided further cycles of improvement within the same context. DBR focusses on designing an artifact, accepts the involvement of researchers in trials, develops both academic theory and practical outcomes, and expects different users for the artifact in each cycle [36].

A. Introduction to Design-Based Research

DBR comes from education research. It started with the 'design experiments' of Brown [37], and Collins [38] working with teachers as co-experimenters. It emphasizes the development of theory in parallel with the practical creation of innovations. DBR is now an accepted research approach, and has a recent guidebook for practitioners [35]. Figure 4, based on Ejersbo et al. [39], shows the two parallel cycles of DBR research: creating theory and creating the artefact. The bold, colored, arrows are additions based on the authors' own experience of the DBR process.

The practical aspects of carrying out DBR are defined by the 'integrative' nature of DBR: both design and assessment techniques must come from other research methodologies [40]. In this research, we used the *techniques* of the Canonical Action Research method [41], though not that method's

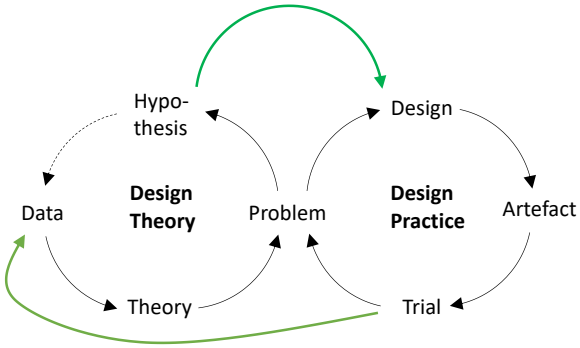


Figure 4: Activities in Practical Design-Based Research

overriding *paradigm*. Specifically, we participated in an activity to help the participants change their behavior. We recorded the discussions involved, transcribed them, and analyzed them in detail. And we used the research findings to inform changes to the workshop to incorporate into a further cycle of development; then repeated the process.

B. Research Questions

DBR requires separate research questions for the Design Practice cycle and the Design Theory cycle. Accordingly, we needed to break down the primary research question RQ1 into sub-questions covering each of Design Practice and Design Theory. For the Design Practice, first, we wanted to understand how to improve the workshops:

RQ 1.1 *What aspects of the workshop worked well, and what changes might help?*

Second, we wanted to know what learning impact the workshops may have had:

RQ 1.2 *What changed in the participant understanding as a result?*

In terms of Design Theory, most important was to question the assumptions of the workshop, that this is a worthwhile and practical set of tasks for the participants:

RQ 1.3 *Can teams of developers produce risk-impact assessments, and integration plans effectively in this way?*

To provide insight to help improve this and similar workshops in future, we also wanted to find out where it was effective, and where less so:

RQ 1.4 *In what ways do the workshop results vary with different participant contexts?*

C. Data Collection

We collected three types of data to address the research questions. First, we recorded the audio of each workshop and created transcripts of the discussions for analysis. For face-to-face workshops we used audio recording devices; for online workshops we used Microsoft Teams' video recording, then extracted the audio and deleted the video. For both, we also used the Microsoft automated transcription service to create a text version of the discussion.

Second, during the workshops one of the researchers listened to the discussions, making observational notes on participant interactions, engagement, perceptions and conclusions—a form of ethnographic analysis [42]. The observational approach was primarily structured to capture data relevant to helping us answer RQ 1.1-RQ 1.4. For example, the observer noted which features of the workshop

TABLE 2: CYCLES, WORKSHOPS, PARTICIPANTS AND NUMBER OF SURVEY RESPONSES (SR)

	ID		Workshop overview	Participants	Study	SR
Cycle 1	W1	Online	Small start-up creating a mobile phone app to support patients.	Entrepreneur and 2 postgraduate cybersecurity students	Own	3
	W2	In person	University team creating proof-of-concept applications for SMEs	4 software developers and 2 project managers.	Own	6
Cycle 2	W3	In person	Public workshop	Entrepreneur running software development company; experienced software developer; cybersecurity masters student.	CS	3
	W4	Online	Start-up creating app to help with mental health issues.	Entrepreneur, leader of outsourced development team, User Interface specialist.	Own	3
	W5	In person	Short public workshop	8 delegates, including 7 with more than 20 years' industry experience, of which 2 were SME Chief Executive Officers (CEOs)	CS	3
	W6	Online	Small business supplying sensor-related apps to the NHS.	2 senior staff, responsible for quality control and 1 product management.	Own	1

were working well, and how participants were engaging with the activities. This structured approach to observation requires the observer to 'stand apart' from the subject under observation [43]. Accordingly, the observing researcher did not take part in discussions and interaction with participants and facilitation of the workshop was led by a different researcher.

Finally, we used an online exit questionnaire implemented using the Qualtrics service[3], and completed by the participants on their mobile devices or computers at the end of the workshop. The questions were designed to address RQ 1.1 *What aspects of the workshop worked well, and what changes might help?* and RQ 1.2 *What changed in the participant understanding as a result?*

D. Analysis Approach

The workshop data analysis was as follows. For the closed questions in the questionnaire, we used graphical statistical analysis [44] to summarize responses addressing aspects of RQ 1.2, RQ 1.3, and RQ 1.4.

The remaining data was unstructured text:

- answers to open questions in the questionnaire,
- ethnographic observations, and
- transcripts (and audio) of the workshops

To analyze this, two researchers dual coded [45] all the open material according to the RQs it addressed. Coding was line-by-line for the questionnaire results and ethnographic observations. We calculated inter-rater reliability figures (Cohen's Kappa) for both [46].

For the coding process we used the tool NVivo. For the workshop content we open coded the automated transcript, referring to the original audio where the transcript was unclear; we created a public domain tool to reformat the Microsoft Teams transcripts for NVivo use[4].

Following an initial coding pass, the two coders met and discussed their findings, then one researcher extracted appropriate findings and quotes to inform the next version of the workshop materials and the results write-up in this paper.

For graphical statistical analysis, we used Python, with Pandas and Seaborn libraries in Jupyter notebooks [28], supported by ChatGPT.

In the first analysis step, we focused particularly on answers to RQ 1.1 *What aspects of the workshop worked well, and what changes might help?*. Based on those answers, we modified the structure of the remaining workshops.

In the second analysis step, we then used Thematic Analysis [47]. A single researcher 'open coded' the text already coded to each separate RQ, then clustered the codes into themes for each RQ. We then gave names to the themes, and wrote up the results.

V. RESULTS AND DISCUSSION

We carried out a total of six trial workshops, with a first cycle of two workshops, followed by analysis and improvements, then a second cycle of four workshops using the improved materials. Table 2 summarizes the participants in each. As shown, three of the workshops were online; the other three were in person. In workshop W1, with only a single entrepreneur from the company, we enlisted two cybersecurity postgraduate students to act as their team. Workshop W5 was a public taster session with a workshop of 1.5 hours rather than 3 hours, that nevertheless included all the elements of the workshop, though planning issues meant that no discussions were recorded. All participants discussed projects related to health IoT apps, their own or a case study, as indicated in the 'Study' column. Most participants filled in the online survey either during or following their workshop; the final column in Table 2 shows the number of valid survey responses received for each. As shown, only a minority of participants in workshops W5 and W6 responded to the survey.

The following sections describe the results first from Cycle 1, then Cycle 2. Sections D to G then give the results of the analysis method described in Section IV.D, each describing results related to a research sub-question, then discussing each one. Tables highlight some of the themes we identified, giving example quotations from workshop participants. Lastly, Sections H, I explore validity and possible next steps.

[3] https://www.qualtrics.com

[4] https://securityessentials.github.io/Teams2NVivo/

TABLE 3: CYCLE 1 IMPROVEMENTS

Problem	Example Quotes	Change Implemented
Cards were taken literally rather than as inspiration.	*But what about risks that are unique to the particular projects? (W2)*	Changed process so all participants write risk ideas on 'post its' related to each risk card.
Participants did not consider impacts beyond immediate losses.	*The worst that could happen re impact is just like you failed to secure your data, here's a fine (W2)*	Added initial explanation of 'financial loss' as a simplification to cover all possible losses.
Participants did not understand likelihoods represented as 'once in N years'.	*Once in 500 years, that's a good scope, isn't it? Like if I live to 500 years' time and I get attacked on the year 499, I know who to complain to. (W1)*	Represented likelihoods on the risk cards as 'one in every N companies in a year'
Participants were daunted by numeric calculations in the impact band instructions.	*A lot of the maths on the instructions made the workshop appear more complicated than it actually was (W2 S)*	Removed calculations from the impact band instructions.
Some participants tended not to contribute to the discussion.	-	Encouraged participants to take it in turn to read out the risk stories.

A. Workshop Costs

The workshops took between 2 and 3 hours. Allowing a further 2–4 person-hours for setup and preparation and given that the workshop materials are available for free, this means that the cost of running the workshop was less than half a day's effort for a development team.

B. Cycle 1 Results

Our initial analysis of workshops W1 and W2 focused on RQ 1.1 *What aspects of the workshop worked well, and what changes might help?,* in order to identify useful improvements to the workshop for cycle 2. The corresponding findings, quotations and improvements are given in Table 3. Statements identify the workshop and source: 'W1' is from the workshop discussion or observations; 'W1 S' is from survey comments. We also implemented a range of smaller improvements, such more initial explanation about risk, encouragement to participants to create further impact bands if appropriate, and clarifications in the risk card stories.

C. Results Including Cycle 2

This section's analysis covers results from the Cycle 1 and Cycle 2 workshops combined. Figure 5 shows density charts of the results from closed questions in the 19 survey results after the workshops.

The Cohen's Kappa values for agreement on coding to the four RQ codes were as shown in Table 4. We can interpret these figures as showing reasonable agreement on questionnaire results and observational notes. We used open coding in the transcripts, identifying text relevant to the research questions, and combined both sets of coding together for the analysis.

D. What Worked Well; What to Change?

Results: The analysis for RQ 1.1 *What aspects of the workshop worked well, and what changes might help?* found several aspects of the workshop reported as working well, as shown in Table 6, as well as three aspects that caused difficulties to the participants, in Table 5. Both tables name and describe relevant themes, with example statements and the supporting statements found in the dataset.

Where workshops involved participants from a mixture of teams and companies (W2, W3, and W5), a case study was used (Section III.B). We observed that this presented challenges around the ability of such groups to form a consensus on details. In W3 and W5, this led to the discussion often going off track due to disagreements over details in the case study.

Discussion: The results in Figure 5 also contribute to answering this question. From the charts we can see that most

TABLE 4: INTER-RATER RELIABILITY

Item Type	Cohen's Kappa
Questionnaire results	0.55
Workshop observational notes	0.32

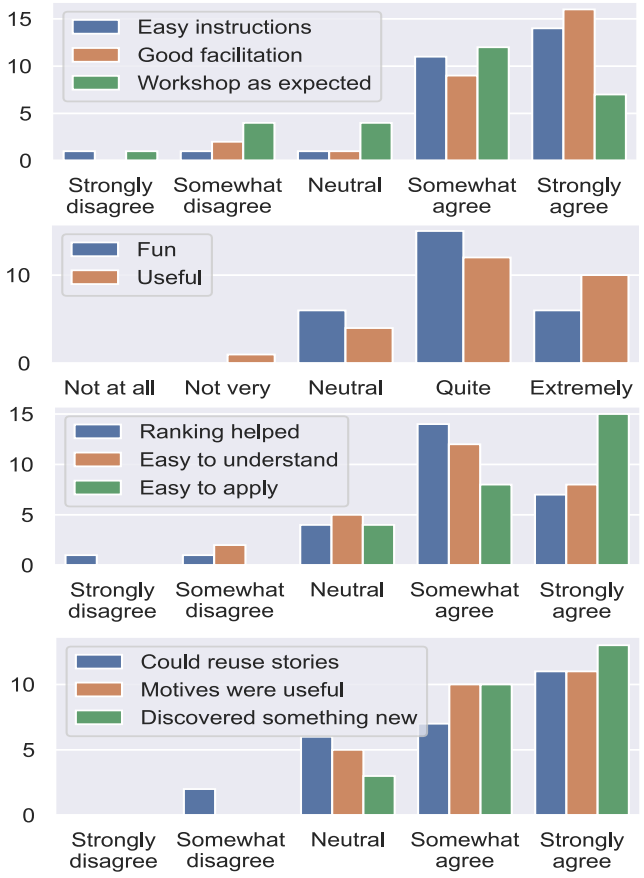


FIGURE 5: PARTICIPANT SURVEY RESPONSES

participants felt they had 'discovered something new' and that the workshops are 'easy to apply'.

Exploring what further changes might help in Cycle 2, we can therefore identify several further possible improvements to the instructions, such as giving an example entry for the 'risk landscape' document and improving the documentation around the case study. However, an issue that gave trouble in several workshops was the difficulty assigning financial numbers to losses, which has no easy solution.

E. Changes in Participant Understanding

Results: Exploring RQ 1.2 *What changed in the participant understanding as a result?* we found the biggest change among participants was a deeper understanding of the problems their products might face, and therefore a better understanding of the risk landscape, as shown in Table 7.

One ethnographic observation was that the use of story-based risk cards allowed participants to consider a broader range of risks than they had previously come across. For some, the process presented risks they would have otherwise missed or overlooked in their normal practices of risk assessment. By encouraging interactivity between participants, where individuals could bounce ideas and 'what if' scenarios between each other, the process established a creative thought process around risk exploration and assessment.

Discussion: From Figure 5, we see that most participants 'discovered something new', and we conclude that many participants did benefit from the workshops.

F. Ability to Produce Assessment and Plans

Results: Considering RQ 1.3 *Can teams of developers produce risk-impact assessments, and integration plans effectively in this way?,* all of the workshops did produce risk landscape documents with the exception of W6, who found no new risks to add.

We observed that each group also proved able to use this process effectively to decide which risks to prioritize and which to disregard. Working through the impact thresholds at the beginning of the process allowed teams to establish criteria for aiding their decision-making alongside technical knowledge of their product. Figure 5 *further* shows that a majority of participants agreed to the propositions that the method was easy to understand and easy to apply, though there was less agreement on 'reusing stories'. Addressing the 'integration plans' part of RQ 1.3, we noted that while each group did try to think about potential integration of workshops into their development process, most did not engage substantially in discussion, and most made no commitment to plans. The exception was in W3, which included a focused discussion around the practicalities of integration; highlighting some of the broader challenges in prioritizing cybersecurity with their clients.

Discussion: We concluded the integration discussion part of the workshop was not very effective, at least in its current format. Possible ways forward include creating a more structured activity for scoping integration plans; or creating a stand-alone follow up session, targeted at groups actively taking the process further.

TABLE 6: ASPECTS THAT WORKED WELL

Theme	N	Description	Example Quotes
Story cards	18	The story cards provided an effective learning tool.	[*I discovered something new]: a range of threats that I had not considered before (W2 S)* *[What worked well was] the idea of using a consistent set of cue cards to assess risk (W2 S)* *For me, it was an eye opener. Probably 90 percent of the threats we went through, I hadn't considered. It was really important for me to take those learnings back (W1 S)*
Group discussion	8	The discussion process produced valuable results	*[What worked best was] the input and the insights from everybody (W1 S)* *[It was not what I was expecting]: … the workshop led to open discussion where I definitely learnt a lot (W3 S)*
Risk method	2	Participants benefited from learning the risk method.	*Having a methodology around risk and around risk we need to mitigate is a new process for me and that's definitely something that's been helpful (W4 S)*
Process fit	2	Participants stated the process would be easy to use in future.	*This would slot in quite nicely [to our development] I think. (W2)* *I will take what we've done today and give that to the development teams and ensure that's at the front of their mind. (W1)*

TABLE 5: AREAS FOR IMPROVEMENT

Theme	N	Description	Example Quotes
Difficult to apply financial numbers	3	Some participants struggled with the impact threshold activity and found assigning financial figures difficult.	*Because of the type of software we work on, the impact is not so much financial….we deal in the medical sector which means the impact is actually the loss of life* (W6)
Case study led discussion off topic	2	Information in the case study being either too little or too much led to conversation going off-topic.	*There's not enough information in the case study to make a determination on quite how deep a loss there needs to be for it to be completely terminal.* (W3)
Need for examples	2	Some participants suggested the addition of examples.	*[Please provide] some examples in the Word doc to reference before compiling* (W4) *Going through worked examples from OWASP could be useful to identify potential problems or risks* (W6)

TABLE 7: CHANGES IN PARTICIPANT UNDERSTANDING.

Theme	N	Description	Example Quotes
Better understanding of threats	8	Participants gained a more detailed understanding of possible threats	*[I discovered] a range of threats I had not considered before (W2 S)* *[I discovered] a list of things to mitigate in the future (W4 S)*
Nuanced understanding of risk	8	Participants gained a better understanding of the risk landscape	*There are some risks that I often overlook when assessing security risks in applications (W2 S)* *The actual rate of risk was a lot lower than I imagined (W1 S)*
New perspectives on cyber	2	The process made participants reflect on cyber security	*How important cyber security is regardless of how technical you are (W2 S)* *I realized most programmers or cyber specialists are in the business of sales…selling comfort for customers. (W3 S)*

G. Variations with Context

Results: Exploring RQ 1.4 *In what ways do the workshop results vary with different participant contexts?* we found that results varied most depending on three aspects: the company maturity, whether the UK National Health Service (NHS) was a customer, and whether more than one company was involved. Table 8 details these findings.

We noted a major variation in results between those already working with the NHS and those not doing so. W6 and W1, both of which involved companies that have gone through such processes already, discounted many of the risks presented and were confident their existing mitigations were sufficient. This stood in contrast to W4, a start-up company, where most risks from the story cards were added to their risk landscape and were deemed something that could potentially happen in the context of their product.

Finally, we observed the power dynamics of the group may also influence the results of the workshop. This happened in W4, where the developers took a backseat in discussing the risk cards and let the product owner take a clear lead. This led to some technical aspects of the discussion not being fully explored as those with the technical expertise were not as forthcoming.

Discussion: The difference in results between NHS-compliant companies and start-ups means that start-ups are likely to gain more from this workshop. For companies already adhering to high standards of compliance, the workshop process was perceived as one more of reassurance than necessity.

The difficulties with keeping discussion on track with mixed groups using the case study suggests a further scope for improvement (for RQ 1.1). It points to a need for more detailed and specific case study information.

Lastly, an assumption we made was that it would be good to bring product owners (such as CEOs) and developers together in the workshop. While this approach meant product owners took insights back to their company, and it supported multiple perspectives at the table, an unanticipated byproduct of this was a power imbalance between participants. We conclude that future workshops need to ensure everyone is empowered to take part.

H. Trustworthiness and Limitations

Table 9 explores five quality criteria for qualitative research of this kind [48], [49] and highlights ways in which this paper satisfies those criteria. We can, however, identify limitations in our deductions from the analysis:

- The calculations behind the likelihood figures are doubtful (Section III.A)
- Not all participants did the exit survey (Table 2).

TABLE 8: VARIATIONS IN RESULTS ACROSS DIFFERENT PARTICIPANT CONTEXTS

Theme	N	Description	Example Quotes
Mixed groups presented challenges	3	Mixed groups struggled to form consensus using the case study.	*There was confusion around the case study examples …trying to map those to a project …was a bit difficult. (W2 S)* *There's not enough information about the company to put a figure on it. (W3)*
NHS-compliant companies identified fewer vulnerabilities	2	Companies already NHS-compliant disregarded the most threats when compiling the risk landscape. They already had many controls in place. The process was perceived more as a reassurance than a necessity.	*We've ended up with absolutely zero [new risks identified]…You should have seen the last two years of being audited continuously and trying to resolve these issues. (W6)* *I thought I've been overthinking it, I'm pleased these thoughts keeping me up at night is paying off finally. (W1)*
Risk landscape assessment held most impact for start-ups	2	Participants from start-ups seemed to get the most from the process, particularly in relation to broadening their understanding of possible threats.	*I probably wouldn't have thought of a bunch of those…having something to flick through as inspiration is quite useful. (W2)* *Just having a list of things that we need to mitigate, that is a new process for me…that's definitely something which has been helpful. (W4)*

TABLE 9: QUALITY CRITERIA

Criteria	What It Means	How Addressed in This Paper
Credibility	The research findings are plausible and trustworthy	Basis in extensive previous work (Section IV); explicit focus and answers to multiple research questions (Section IV.D); detailed and documented analysis (Section V)
Dependability	The extent to which the research could be replicated in similar conditions	Workshop materials publicly available with full instructions (Section III.B); risk analysis explained (Section III.B)
Confirmability	There is a clear link or relationship between the data and the findings	Mapping of the results to research questions (Sections IV.D, V); use of quotes to substantiate results (Section V.A, Table 3–Table 8)
Transferability	Findings may be transferred to another setting, context or group	Effectiveness in a range of situations (Table 2); possibility of further industry sector support (Section V.I)
Reflexivity	A continual process of articulating the place of the researcher and the context of the research	Explicit descriptions of the researcher roles in the research (Section III.B)

- There is likely to be a bias to the positive in the results in in Figure 5, given the situation (a free workshop, and researchers in the room during survey completion), even though all results were anonymous.
- While it seems likely that the developers' assessments were sufficient for the purpose, and that the consequences of a wrong risk assessment were less than the consequence of not doing one at all, we have no way of verifying this. This remains an outstanding question for future research.

The findings of this paper, therefore, form an existence proof: yes, a facilitated workshop can teach a lightweight risk assessment process for SMEs. In addition, the range of different types of development involved in the trials prove there is a range of situations in which such a workshop can work.

I. *Next steps*

We have identified several areas for continued work:

Broadening scope: Though the industry statistics we gathered and descriptions on the cards are specific to the health-related SMEs, we know it will be easy to use the same method to generate figures for other industry sectors. Indeed, given the need to round the likelihoods to the nearest power of 10, figures are unlikely to differ much between sectors. With some rewriting to create more generic risk descriptions, this will offer a version of the workshops suitable for a much wider range of users.

Promotion and outreach: The workshops are to be incorporated into our program of academic outreach, with the assistance of two industry partners. We continuing to promote them through training workshops at industry developer conferences.

Statistics improvement: A remarkable finding from the work was the paucity of publicly available probability figures for different security and privacy risks. Working with government, international sources and perhaps insurance companies can help provide better figures.

Proof of impact: We plan interview surveys with these and other workshop participants to determine the long-term impact of the workshops.

A dataset of the anonymized transcripts, survey results and coding is available for other researchers, subject to ethics constraints, on request [50]

VI. SUMMARY

This paper explored the background, design and execution of a facilitated workshop to help SME HIoT software developers to use risk-based assessment in their planning and documentation of security and privacy for software.

An early finding was that objective '*industry-based cybersecurity data*' on risk probabilities not available in any practical form for a SME development team. We therefore used a range of UK government sources (Section III.A) to estimate probabilities. To convey the results to workshop participants, we used prompt cards, with 'stories' describing each threat, and order-of-magnitude probability information.

Using Design Based Research and Ethnography, involving two cycles of trials and six sets of workshop participants, we addressed four specific research sub-questions, as follows:

RQ 1.1 *What aspects of the workshop worked well, and what changes might help?* Following improvements based on answers this question for the first two workshops, the resulting format of the workshop did work well (Section V.D). The main improvement that might help would be improvements in the workshop materials, and in the longer term, further innovation in teaching participants to visualize impact in financial terms (Section V.D).

RQ 1.2 *What changed in the participant understanding as a result?* Participants developed a deeper understanding of the types of problems they might face, and therefore a better understanding of the risk landscape. The workshops that evaluated real products also generated effective risk landscape documents (Section V.E).

RQ 1.3 *Can teams of developers produce risk-impact assessments, and integration plans effectively in this way?* Feedback from the participants, and analysis of the risk landscape documents developed in the workshops, suggest that all the developers could indeed produce risk-impact assessments given the guidance in the workshops (Section V.F). The workshop discussions, however, were not very effective at producing plans to integrate the workshops into the participants' development processes (Section V.F).

Section V.A, however, shows that the time and effort cost of the workshops were small, making them cost effective for the teams involved.

RQ 1.4 *In what ways do the workshop results vary with different participant contexts?* Two main factors affected the results. Companies already supplying to the stringent standards of the UK NHS had little to learn from the workshops (Section V.G). And groups with substantial power dynamics at play produced less effective results (Section V.G) .

Thus, we can answer the main research question, RQ1 *How can industry-based cybersecurity data improve security and privacy aspects of the development of Health IoT systems within resource constrained development teams?* We conclude that a facilitated workshop with discussion groups of developers, using prompt cards with risk 'stories' and probability data, to generate a 'risk landscape' document, provides an effective and inexpensive answer.

We conclude that this approach provides a powerful way to help healthcare software development teams to make their security and privacy work more effective.

Given the ease of extending the workshop materials to cover a wider range of industries (Section V.I), we look forward to this approach helping *any* software development team improve their decision making around security and privacy.

VII. Acknowledgements

We thank all the teams of developers and companies who contributed to this research, including those who gave permission for their photographs to be used in this paper. We also thank the editors and reviewers who helped us present the work effectively in this paper.

References

[1] D. Zaldivar, L. A. A. Tawalbeh, and F. Muheidat, "Investigating the Security Threats on Networked Medical Devices," presented at the Tenth Annual Computing and Communication Workshop and Conference, 2020.

[2] Y. Sun, F. P. W. Lo, and B. Lo, "Security and Privacy for the Internet of Medical Things Enabled Healthcare Systems: A Survey," *IEEE Access,* vol. 7, pp. 183339-183355, 2019, doi: 10.1109/ACCESS.2019.2960617.

[3] C. Weir, I. Becker, and L. Blair, "A Passion for Security: Intervening to Help Software Developers," in *IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 2021: IEEE, pp. 21-30, doi: 10.1109/ICSE-SEIP52600.2021.00011.

[4] B. Shreeve, J. Hallett, M. Edwards, K. M. Ramokapane, R. Atkins, and A. Rashid, "The Best Laid Plans or Lack Thereof: Security Decision-Making of Different Stakeholder Groups," *IEEE Transactions on Software Engineering,* 2020, doi: 10.1109/TSE.2020.3023735.

[5] ISO, "ISO 31000:2018(en) Risk management - Guidelines," ed, 2018.

[6] D. W. Hubbard and R. Seiersen, *How to Measure Anything in Cybersecurity Risk*. John Wiley & Sons, 2016.

[7] A. Shostack, *Threat Modeling: Designing for Security*. John Wiley & Sons, 2014.

[8] W. Xiong and R. Lagerström, "Threat Modeling – a Systematic Literature Review," *Computers and Security,* vol. 84, 2019, doi: 10.1016/j.cose.2019.03.010.

[9] I. Brass and A. Mkwashi, "Risk Assessment and Classification of Medical Device Software for the Internet of Medical Things: Challenges Arising From Connected, Intelligent Medical Devices," in *12th International Conference on the Internet of Things*, 2022: Association for Computing Machinery, pp. 171-178, doi: 10.1145/3567445.3571104.

[10] S. B. Baker, W. Xiang, and I. Atkinson, "Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities," *IEEE Access,* vol. 5, pp. 26521-26544, 2017, doi: 10.1109/ACCESS.2017.2775180.

[11] M. Hassanalieragh *et al.*, "Health Monitoring and Management Using Internet-of-Things (IoT) Sensing with Cloud-Based Processing: Opportunities and Challenges," in *International Conference on Services Computing*, 2015 2015: Institute of Electrical and Electronics Engineers Inc., pp. 285-292, doi: 10.1109/SCC.2015.47.

[12] S. H. Shah and I. Yaqoob, "A Survey: Internet of Things (IOT) Technologies, Applications and Challenges," in *4th IEEE International Conference on Smart Energy Grid Engineering, SEGE 2016*, 2016: Institute of Electrical and Electronics Engineers Inc., pp. 381-385, doi: 10.1109/SEGE.2016.7589556.

[13] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K. S. Kwak, "The Internet of Things for Health Care: A Comprehensive Survey," *IEEE Access,* vol. 3, pp. 678-708, 2015, doi: 10.1109/ACCESS.2015.2437951.

[14] I. T. Governance UK, "Cyber Risk Management," ed.

[15] J. M. Such, J. Vidler, T. Seabrook, and A. Rashid, "Cyber Security Controls Effectiveness: A Qualitative Assessment of Cyber Essentials," 2015. [Online]. Available: http://eprints.lancs.ac.uk/74598/

[16] F. Hoppe, N. Gatzert, and P. Gruner, "Cyber Risk Management in SMEs: Insights from Industry Surveys," *The Journal of Risk Finance,* 2021, doi: 10.1108/JRF-02-2020-0024.

[17] R. M. Savola, C. Frühwirth, and A. Pietikäinen, "Risk-Driven Security Metrics in Agile Software Development-An Industrial Pilot Study," *Journal of Universal Computer Science,* vol. 18, no. 12, pp. 1679-1702, 2012, doi: 10.3217/jucs-018-12-1679.

[18] S. Türpe, L. Kocksch, and A. Poller, "Penetration Tests a Turning Point in Security Practices? Organizational Challenges and Implications in a Software Development Team," in *Workshop on Security Information Workers - SIW*, 2016: USENIX Association.

[19] A. Poller, L. Kocksch, S. Türpe, F. A. Epp, and K. Kinder-Kurlanda, "Can Security Become a Routine? A Study of Organizational Change in an Agile Software Development Group," in *Conference on Computer Supported Cooperative Work - CSCW*, Portland Oregon USA, 2017: ACM, pp. 2489-2503, doi: 10.1145/2998181.2998191. [Online]. Available: https://testlab.sit.fraunhofer.de/downloads/Publications/poller2017routine.pdf

[20] K. Yskout, R. Scandariato, and W. Joosen, "Do Security Patterns Really Help Designers?," in *International Conference on Software Engineering - ICSE*, Firenze, Italy, 2015: IEEE, pp. 292-302, doi: 10.1109/ICSE.2015.49. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7194582

[21] L. Bell, M. Brunton-Spall, R. Smith, and J. Bird, *Agile Application Security: Enabling Security in a Continuous Delivery Pipeline*. Sebastopol, CA: O'Reilly, 2017.

[22] T. Lopez, H. Sharp, T. Tun, A. Bandara, M. Levine, and B. Nuseibeh, "Talking about Security with Professional Developers," in *Workshop on Conducting Empirical Studies in Industry - CESSER-IP*, Montreal, QC, Canada, 2019: IEEE Computer Society, doi: 10.1109/CESSER-IP.2019.00014.

[23] N. Merrill, "Security Fictions: Bridging Speculative Design and Computer Security," in *ACM Designing Interactive Systems Conference*, Eindhoven, 2020: ACM, pp. 1727-1735, doi: 10.1145/3357236.3395451.

[24] Department for Digital Culture Media Sport, "Cyber Security Breaches Survey: Combined Dataset 2016-2022," ed, 2022.

[25] Information Commissioners Office, "Data Security Incident Trends," ed: ICO, 2022.

[26] Office for National Statistics, "UK Business: Activity, Size and Location," ed, 2021.

[27] U. K. Finance, "Annual Fraud Report," 2022. [Online]. Available: https://www.ukfinance.org.uk/policy-and-guidance/reports-and-publications/annual-fraud-report-2022

[28] T. Kluyver *et al.*, "Jupyter Notebooks: A Publishing Format for Reproducible Computational Workflows," IOS Press, 2016, pp. 87-90.

[29] C. Weir, A. Dyson, and D. Prince, "Do You Speak Cyber?: Talking Security with Developers of Health Systems and Devices," *IEEE Security and Privacy Magazine,* 2022, doi: 10.1109/MSEC.2022.3221616.

[30] Stack Overflow, "Annual Developer Survey," 2016. [Online]. Available: https://insights.stackoverflow.com/survey/2016

[31] C. Weir, B. Hermann, and S. Fahl, "From Needs to Actions to Secure Apps? The Effect of Requirements and Developer Practices on App Security," presented at the 29th USENIX Security Symposium, 2020.

[32] M. Mitre, "Playbook for Threat Modeling Medical Devices," 2021. [Online]. Available: https://mdic.org/wp-content/uploads/2021/11/Playbook-for-Threat-Modeling-Medical-Devices.pdf

[33] C. Weir, A. Dyson, and D. Prince. "Hipster Efficient Software Security and Privacy Workshop." https://securityessentials.github.io/HipsterWorkshop/ (accessed 2023).

[34] C. Weir, I. Becker, and L. Blair, "Incorporating Software Security: Using Developer Workshops to Engage Product Managers," *Empirical Software Engineering,* vol. 28, no. 2, pp. 1-33, 2023, doi: 10.1007/S10664-022-10252-0/TABLES/10.

[35] A. Bakker, *Design Research in Education: A Practical Guide for Early Career Researchers*. Abingdon: Routledge, 2018.

[36] A. E. Kelly, R. A. Lesh, and J. Y. Baek, *Handbook of Design Research Methods in Education : Innovations in Science, Technology, Engineering, and Mathematics Learning and Teaching*. Routledge, 2008, pp. 539-539.

[37] A. L. Brown, "Design Experiments : Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings," *Journal of the Learning Sciences,* vol. 2, no. 2, pp. 141-178, 1992, doi: 10.1207/s15327809jls0202_2.

[38] A. Collins, "Toward a Design Science of Education," Springer, 1992, pp. 15-22.

[39] L. R. Ejersbo, R. Engelhardt, L. Frølunde, T. Hanghøj, R. Magnussen, and M. Misfeldt, "Balancing Product Design and Theoretical Insights," Routledge, 2008, pp. 149-164.

[40] F. Wang and M. J. Hannafin, "Design-Based Research and Technology-Enhanced Learning Environments," *Educational Technology Research and Development,* vol. 53, no. 4, pp. 5-23, 2005, doi: 10.1007/BF02504682.

[41] R. M. Davison, M. G. Martinsons, and N. Kock, "Principles of Canonical Action Research," *Information Systems Journal,* vol. 14, no. 1, pp. 65-86, 2004, doi: 10.1111/j.1365-2575.2004.00162.x.

[42] H. Sharp, Y. Dittrich, and C. R. B. De Souza, "The Role of Ethnographic Studies in Empirical Software Engineering," *IEEE Transactions on Software Engineering,* vol. 42, no. 8, pp. 786-804, 2016, doi: 10.1109/TSE.2016.2519887.

[43] A. Mulhall, "In the Field: Notes on Observation in Qualitative Research," *Journal of Advanced Nursing,* vol. 41, no. 3, pp. 306-313, 2003, doi: 10.1046/j.1365-2648.2003.02514.x.

[44] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey, *Graphical Methods for Data Analysis*. CRC Press, 1983, pp. 1-395.

[45] N. McDonald, S. Schoenebeck, and A. Forte, "Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice," *Proceedings of the ACM on Human-Computer Interaction,* vol. 3, no. CSCW, 2019, doi: 10.1145/3359174.

[46] K. L. Gwet, *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters*. Advanced Analytics LLC, 2014.

[47] V. Braun and V. Clarke, "Thematic Analysis," H. Cooper Ed.: American Psychological Association, 2012, pp. 57-71.

[48] T. Stenfors, A. Kajamaa, and D. Bennett, "How to … Assess the Quality of Qualitative Research," *The Clinical Teacher,* vol. 17, no. 6, pp. 596-599, 2020, doi: 10.1111/TCT.13242.

[49] N. K. Denzin and Y. S. Lincoln, *The Sage Handbook of Qualitative Research*. Sage Publications: Thousand Oaks, CA, 2011.

[50] C. Weir, A. Dyson, and D. Prince. *Hipster Workshop Trials Dataset*, doi: 10.17635/lancaster/researchdata/611.

Appendix A Post-Workshop Survey

Welcome our post workshop survey

Thank you for taking part in the workshop, we hope you found it useful and enjoyable. We are looking to capture information which will help us to improve the workshop and also research data which will help us to understand the role of the use of risk and threat intelligence information in product development processes.

The survey should take around 10 mins and is split into two parts: 1) General feedback about the workshop experience 2) In depth questions around the role of the workshop and the impact it might have.

Part 1: General Workshop Feedback

Please provide us with an assessment of your overall experience of participating in the workshop and educational experience.

Q1 Which of these is a substantial part of your normal working role?

(Multiple selection)

- Programming
- Management/Coordination
- Quality Assurance
- Strategy/Product Management

Q2 Workshop Setup and Delivery

(Strongly disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Strongly agree)

- Instructions for the workshop were easy to understand
- The facilitation of the workshop was good.
- The workshop was what I was expecting.

Q3 You have indicated the workshop was not what you were expecting. Please let us know how it was different to what you expected.

Q4 Experience of the Process

(Not at all, Not very, Neutral, Quite, Extremely)

- How much fun did you find the process?
- Did you find it useful to have these discussions with your team?

Q5 What worked best in the workshop?

(Free text)

Q6 How could we improve the workshop in future?

(Free text)

Part 2: The Role of the Workshop

Please answer the following questions to help us research and understand the impact the workshop has had on you and the team.

Q7 What impact did having quantitative likelihood information have?

(Strongly disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Strongly agree)

- The quantitative information was easy to understand
- The quantitative information was easy to apply to assess the risks
- Ranking the threats helped me to understand the severity of the different threats

Q8 Using Stories in the Threat Cards

(Strongly disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Strongly agree)

- I discovered something new or unknown through using stories
- The stories we created for ourselves as a result could be used in future discussions
- Imagining motives was useful for thinking about security and privacy of the system.

Q9 You indicated that you discovered something new or unknown through this process. Please expand on what you discovered

(Free text)

Q10 Was the visual representation of risk data suitable for this process?

(Strongly disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Strongly agree)

- The representation of risk information on the cards was easy to understand
- The representation of risk information was easy to synthesize in relation to risk decisions
- The way the information was presented was accessible to a range of expertise

Q11 Might this process be integrated into your current practices?

(Not at all, Not very, Neither agree nor disagree, Quite, Extremely)

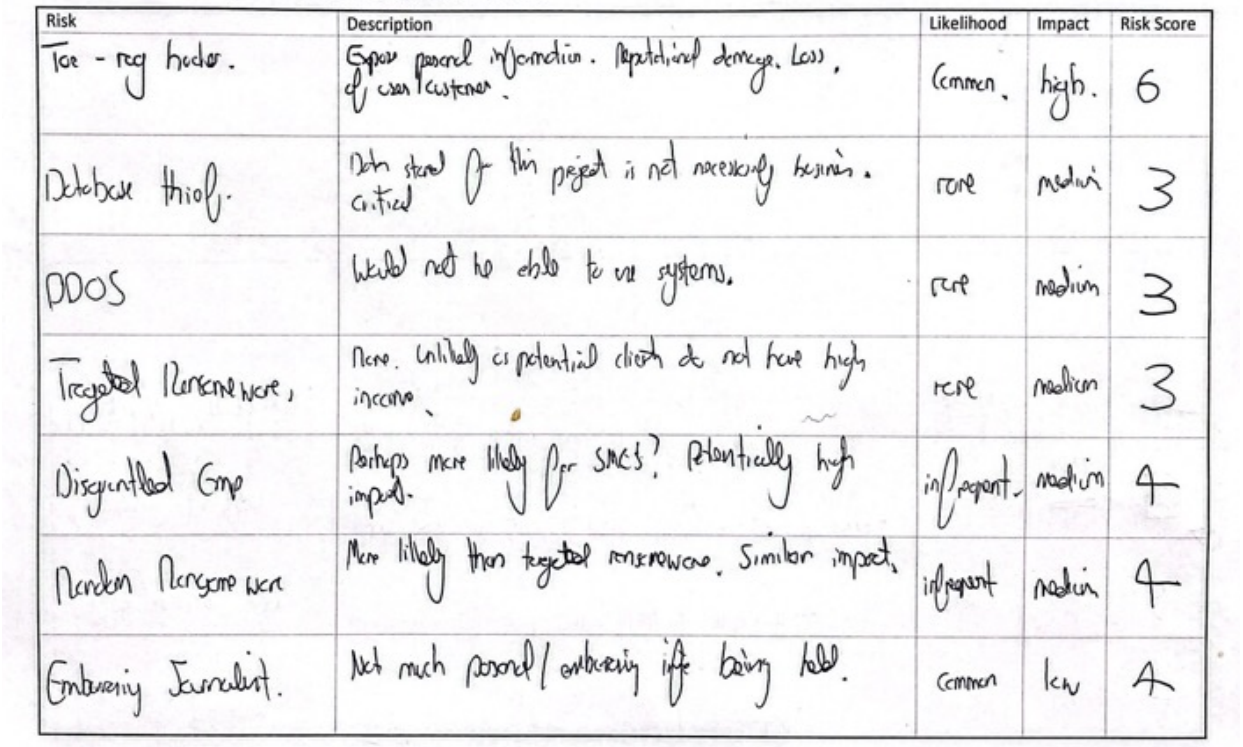
- How convinced are you that you could use this approach in your current development practices?
- How much do you think this might help in discussions with product management?
- How likely do you think it is that you will use this approach in your projects in future?

Q12 What is your key takeaway from the session?

(Free text)

Appendix B Example Risk Landscape Document

This is an extract from a risk landscape document created in a workshop, used by permission.



Risk	Description	Likelihood	Impact	Risk Score
Toe - reg hacker.	Expose personal information. Reputational damage. Loss of user/customer.	Common.	high.	6
Database thief.	Data stored for this project is not necessarily business critical	rare	medium	3
DDOS	Would not be able to use systems.	rare	medium	3
Targeted Ransomware.	None. Unlikely as potential clients do not have high income.	rare	medium	3
Disgruntled Emp	Perhaps more likely for SMEs? Potentially high impact.	infrequent.	medium	4
Random Ransomware	More likely than targeted ransomware. Similar impact.	infrequent	medium	4
Embarrassing Journalist.	Not much personal/embarrassing info being held.	common	low	4

The team had assigned impact thresholds Low-Medium: £100K, Medium-High £1M. So, using the likelihoods in Table 1, we see a risk score of 6 might represent an expectation of loss greater than £100K per year; one of 4, £1K-£10K per year. Note that this was a case study exercise, so the figures were not reviewed.