“There are rabbit holes I want to go down that I’m not allowed to go down”: An Investigation of Security Expert Threat Modeling Practices for Medical Devices

Ronald E. Thompson, Madline McLaughlin, Carson Powers, and Daniel Votipka, *Tufts University*

https://www.usenix.org/conference/usenixsecurity24/presentation/thompson

This paper is included in the Proceedings of the 33rd USENIX Security Symposium.

August 14–16, 2024 • Philadelphia, PA, USA

978-1-939133-44-1

Open access to the Proceedings of the 33rd USENIX Security Symposium is sponsored by USENIX.

"There are rabbit holes I want to go down that I'm not allowed to go down": An Investigation of Security Expert Threat Modeling Practices for Medical Devices

Ronald Thompson, Madeline McLaughlin, Carson Powers, and Daniel Votipka
Tufts University
{rthomp06,mmclau05,cpower04,dvotipka}@cs.tufts.edu

Abstract

Threat modeling is considered an essential first step for "secure by design" development. Significant prior work and industry efforts have created novel methods for this type of threat modeling, and evaluated them in various simulated settings. Because threat modeling is context-specific, we focused on medical device security experts as regulators require it, and "secure by design" medical devices are seen as a critical step to securing healthcare. We conducted 12 semi-structured interviews with medical device security experts, having participants brainstorm threats and mitigations for two medical devices. We saw these experts do not sequentially work through a list of threats or mitigations according to the rigorous processes described in existing methods and, instead, regularly switch strategies. Our work consists of three major contributions. The first is a two-part process model that describes how security experts 1) determine threats and mitigations for a particular component and 2) move between components. Second, we observed participants leveraging use cases, a strategy not addressed in prior work for threat modeling. Third, we found that integrating safety into threat modeling is critical, albeit unclear. We also provide recommendations for future work.

1 Introduction

Threat modeling is a process commonly used for decades in industry and academia during system design to evaluate risk in software [8, 96]. This process is split into two lines of questioning: 1) identifying security objectives, risks, and vulnerabilities, and 2) determining mitigations to limit risk to acceptable levels [17]. This allows security experts—those in charge of securing a system—to abstract away system details, brainstorm what threats might affect their system, and develop a secure architecture [96, pg. 3].

Both industry and academia have developed multiple structured methods for threat modeling [6, 26–28, 58, 88, 92, 106, 109, 118]. Most of these methods, built to help users threat model, have only been evaluated by researchers simulating expected user behavior on a few case-study systems [7, 16, 56, 57, 62, 64, 108, 116]. However, it is unknown if security experts explicitly use these when asked to brainstorm potential threats and mitigations in a real-world system.

Some studies have asked real users to conduct threat modeling tasks [36, 38, 40, 97–99, 110]. These studies focus primarily on participants' threat modeling outcomes (e.g., number, variance, or accuracy of threats identified) for comparison. Of these, four describe security experts' processes [38, 40, 98, 110]. However, these remained at a high level (e.g., categorized design documents as detailed vs. not detailed), limiting the conclusions regarding participants' exact process that can be drawn from their results. Additionally, few participants in their sample populations had experience with threat modeling (i.e., mostly students and non-security experts). Therefore, the question of how security experts threat model in practice remains.

In this paper, we investigate how professional MDM security experts threat model in practice. Threat modeling, by its nature, is context-sensitive. Security experts must understand the design and function of the systems they build to provide a thorough review. As we sought to capture the realistic practice of professional security experts, selecting a single domain of experts was a necessary first step to ensure participants were sufficiently familiar with the threat modeling scenario details. Therefore, we chose to consider professionals involved in medical device design and development, i.e., those who work for and with medical device manufacturers (MDMs). For brevity, we refer to this group as MDM security experts from now on. MDM security experts were ideal for our study as they offer an upper bound for familiarity with threat modeling for several reasons. First, MDM security experts build life-critical technologies [45, 75] that, if disrupted, inhibit essential care to patients [29, 82, 90, 115], making secure design practices essential in this domain. Further, regulators and industry standards recognize the need for "secure by design approaches" [2, 17, 34, 35, 49, 66] and now require threat modeling for medical devices [17, 34, 47, 59, 66, 103]. Note, while most regulatory bodies require threat modeling, no regula-

tory guideline requires a specific method, instead leaving this decision to the MDM security experts.

In this paper, we investigate how professional MDM security experts threat model in practice. Historically, human-computer interaction research has suggested having a model of actual user practice is necessary to establish a foundation for future method and tool development that best supports users, more naturally fitting their needs and providing structured support without getting in the way [95, pg. 19-22]. To this end, we conducted 12 semi-structured interviews with MDM security experts to investigate their threat modeling processes. We chose an exploratory approach because there is limited prior work and a lack of consensus on the most effective processes. While a qualitative study cannot test the effectiveness or prevalence of various practitioner actions and processes, it allows us to enumerate the range of processes followed and study important process characteristics and interactions between parts of their approach and participant backgrounds in depth. Using this qualitative approach, we create a theoretical model of the threat modeling process to guide future tool design and quantitative assessments. We ask the following questions:

- **RQ1:** How do MDM security experts identify specific threats and mitigations?
- **RQ2:** What processes do MDM security experts follow when navigating a system's design to identify threats?

In each interview, we presented participants with two mock medical devices and asked them to identify threats and mitigations. We observed which parts of the system they focused on, how they elicited threats and mitigations in specific system components (RQ1), and if they relied on a specific approach to traverse the system and ensure they covered every part of the potential attack surface (RQ2). These two parts allow us to enumerate the medical device threat modeling process from the practitioners' perspective.

Our contributions. We found MDM security experts do not discuss threats and mitigations in a standard order, e.g., identifying a threat, then the appropriate mitigation. Instead, threat and mitigation identification is very flexible, sometimes starting with a threat or security property, but other times beginning with a mitigation or safety concern. We also found that personal experiences played a significant role in the specific considerations, especially when discussing potential mitigations. MDM security experts tended to navigate the system in an ad-hoc fashion, using multiple approaches during a single threat modeling session, including a specific use case/workflow-focused approach, which has not been investigated deeply before. We also found MDM security experts consider multiple system configurations when threat modeling based on the different ways they expect the systems to be deployed, the system architectures used, and how the device functions throughout its life cycle. This highlights the detail MDM security experts consider during threat modeling.

We outline a two-part process model for medical device threat modeling that captures the range of strategies MDM security experts follow. It enumerates the questions and sub-questions MDM security experts considered for a particular component or subset of components in the system and how they navigate the reviewed system. This model provides a foundation for future work in threat modeling to capture security experts' more varied and flexible approaches. This model can guide empirical evaluation of existing methods and the development of tools aligned closer with actual security expert practice, and we provide recommendations for each.

We also observed safety is a critical aspect of medical device threat modeling, but MDM security experts are uncertain about how this should be integrated into methodologies focusing on security outcomes rather than explicitly on safety.

2 Background & Related Work

In this section, we place our research in the context of the prior literature. We discuss the definition of threat modeling and the various methodologies that have been proposed. We discuss heuristic evaluations of threat modeling, user studies investigating practitioner threat modeling practices, and general secure development user studies.

Threat modeling methods. Because we focused on medical devices, we take a practitioner-focused view and use the definition from the US Food and Drug Administration (FDA): "Threat modeling includes a process for identifying security objectives, risks, and vulnerabilities across the medical device system, and then defining countermeasures to prevent, mitigate, monitor, or respond to the effects of threats to the medical device system throughout its lifecycle" [17, pg. 13]. Methods can vary widely in detail and strictness. At one extreme, these methods can focus on broad questions with no strict order, such as the Four Questions, which was developed by the threat modeling community as a set of guiding questions (What are we building, What can go wrong, What are we going to do about it, and Did we do a good enough job) [118]. Other methods specifically support threat elicitation, without giving a specific order to follow. This includes STRIDE [58], Attack Trees [88], Persona Non-Grata [92], and Security Cards [27]. The specific usage of these methods is open to interpretation and preference, leading to some creating more specific derivatives [91, 101]. Finally, there are methods that cover the entire threat modeling process, provide more specific details, and strict ordering, such as PASTA [109], TRIKE [28], LINDDUN [26], and VAST [106]. It is not known how widely adopted these are and if security experts follow them explicitly in practice or use a variation.

Tools supporting threat modeling. In addition to various threat modeling methods, numerous tools are available to

users to help with threat modeling. These range from creating threat modeling languages [54], creating threat modeling diagrams [61], and providing automation for threat modeling [37, 51, 71, 105]. These tools implicitly guide users to follow a process, some more than others, depending on their interaction design. While prior work has compared the technical aspects of these tools (number of threats and mitigations they can suggest) [93], there has been no investigation of whether their interaction design fits user processes. This motivates our study to identify real-world security experts' practices.

Audits of threat modeling processes. Prior work has investigated various suggested threat modeling approaches or developed their own novel approaches in the context of cyber-physical systems [7, 16, 57, 62, 64, 108]. In these studies, the researchers themselves apply the threat modeling process to a few example systems in a systematic fashion and compare identified threats and mitigations. For example, Khan et al. compared the results of a STRIDE-per-element—investigate each component separately in turn—to STRIDE-per-interaction—investigate components as interacting pairs, finding the former uncovered threats missed by the latter [57]. Khan and the other researchers applied each STRIDE variation to a smart grid in a laboratory testbed and compared their results. Many of these rely only on researcher perspectives and, in most cases, limited case study evaluation. It is unclear how their results generalize to real-world practice. In contrast, we investigate how MDM security experts approach threat modeling in practice to establish a threat modeling process model for more realistic evaluations in future work.

Threat modeling user studies. Turning to investigating how practitioners actually threat model, multiple studies have asked participants to consider and design systems to avoid potential security threats. Shull et al. asked practitioners and novices to produce a threat model for two systems, assigning one of three threat modeling approaches to each participant [36]. They found wide variation among participant outcomes, demonstrating that participants, given the same process, could produce very different results. Similarly, Stevens et al. introduced Center of Gravity (CoG) threat modeling to practitioners at New York City Cyber Command and observed participant behaviors over a period of time [99]. They demonstrated concrete security improvements tied to using CoG and observed participants had more confidence in their ability to secure critical assets and communicate security needs to others. Fulton et al. investigate the relationship between threat model thoroughness and produced code security through a classroom-based, quasi-controlled experiment with 14 student teams asked to develop a secure IoT hub [40]. They found that thorough design and threat modeling produced code with fewer vulnerabilities. Van Landuyt and Joosen looked at what explicit and implicit assumptions were made by users when using STRIDE [110]. While they primarily looked at students, they compared the results to a handful of expert threat models. They found that assumptions were mainly to exclude threats as a way of efficiency, which we also saw to some extent. Unlike these studies, we did not proscribe a specific process, allowing us to capture actual participant processes. We focused explicitly and in detail on participant threat modeling processes rather than outcomes. This allows us to identify natural threat modeling behaviors and develop a model for better understanding why particular outcomes or discrepancies occur by providing a higher-fidelity view of the process.

Most related to our work is Shreeve et al.'s investigation of decision-makers' risk thinking [38, 97, 98]. Shreeve et al. used a tabletop game where twelve teams (four were security experts) were asked to review a cybersecurity risk scenario related to a CPS. The researchers observed how participants identified potential risks and suggested controls to mitigate these risks. They found several participants considered which mitigations to deploy first instead of the more expected risk-first approach or often switched approaches. Our design is similar but differs in several ways. First, we do not restrict the information about the system available to participants. Shreeve et al. included this mechanism to force participants to reveal which information they expected would be more helpful. We do not attempt to capture efficacy but instead, opt for a more naturalistic setting. Additionally, while some of Shreeve et al.'s participants were security experts, they were not experts in the domain investigated. Because we expected threat modeling to be expertise-driven, we focused on individuals in a single field who were experts in security and their domain. Our interviews allow us to dig deeper into participants' reasoning, providing a greater detail to the results.

Developer Security Practices. Considering secure development more broadly, several recent studies have investigated how security is incorporated throughout development. This work has studied organizational factors that impact the extent to which developers think about security in practice [11, 12, 39, 46, 80, 113], and how they think about security (and make mistakes) during implementation [4, 74, 78, 87, 111]. Most relevant to our work has been research studying developer practices during the design phase. For example, Palombo et al. conducted an ethnographic study focused on the secure development lifecycle [80]. In their observations of a software company's design process, they found security vulnerabilities were not only due to developer misunderstandings but also from miscommunication or lack of motivation due to other stakeholder priorities. These studies provide an essential end-to-end view of secure development and highlight the need for thorough design but do not offer insights into how practitioners' actually conduct the task of threat modeling.

3 Methods

We conducted 12 semi-structured interviews with MDM security experts. This includes those directly employed by MDMs and consultants working with MDMs. We investigated how these experts think about threats to medical devices and potential security controls. Interviews were conducted between January 2022 and July 2023. We describe the development of our protocol, recruitment process, analysis approach, ethical considerations, and the study's limitations. Our university's Institutional Review Board (IRB) approved this study.

3.1 Community Engagement

Due to intellectual property concerns, MDM security experts were unwilling to share product designs and directly discuss threat models of their products. To enable engagement with several MDM security experts from various companies, we asked participants to generate threat models for mock scenarios as a mechanism for framing a broader discussion about their threat modeling process. To ensure the scenarios and structure of our study were realistic, we made significant efforts to engage the community over the past two years as we designed our study instruments. We participated in eight conferences and working group meetings in the United States (many with global attendance). During these and subsequent 1-on-1 meetings, we discussed the community's makeup and threat modeling norms with leaders, including current and former FDA regulators, multiple top security personnel at the largest MDMs, and other prominent individuals with combined decades of experience in medical device security. The lead author has been working with MDMs to produce threat models and leading workshops on threat modeling and security for two years in a non-research, professional capacity.

Through our community engagement, we received feedback about and refined our mock scenarios (see Section 3.2) and were provided insights about recruitment to ensure our pool included participants from the roles in MDMs who perform threat modeling (see Section 3.4). We never discussed specific scenarios or interview questions to avoid potentially biasing participants during our community engagement; we focused on high-level concepts to shape the study design.

3.2 Scenario Details

In our interview protocol (see Section 3.3), we presented participants with mock medical devices to observe how they threat model a new device. To ensure these scenarios were realistic and captured a range of participant experiences, we created three scenarios encompassing various connected medical systems requiring a regulatory cybersecurity review. Each device was based on real-world examples with previously discovered vulnerabilities—a common practice in prior threat modeling research [7,16,57,62,64,108]—and was thoroughly vetted by our community contacts (Section 3.1) to ensure it matched a realistic system.

We chose these scenarios to cover patient harm and usage settings. We based these criteria on our experiences, industry guidelines, prior work, and feedback during community engagements. Using these two axes to develop our scenarios allowed us to cover a variety of threats (e.g., internal threats, malicious actors, human errors, and system failures) and the severity of outcomes. As part of our Expert Review (Section 3.3), we validated the classifications we assigned with two former regulators. We further discuss the global regulatory requirements in our Supplemental Materials [1][1].

For patient harm, the FDA and EU's Medical Device Regulation (MDR) classify devices into three tiers for regulatory oversight based partially on potential harm and functionality. Type I devices, such as face masks, syringes, and tongue depressors, are not software-enabled and would not undergo a cybersecurity review. Type II presents moderate or indirect harm, and Type III presents a life-threatening risk; both require threat modeling. We chose devices spanning Types II and III. While the FDA uses the same classification for diagnostic equipment, the EU regulates diagnostic systems under the In Vitro Diagnostic Regulation, which uses a scale from low patient risk (Class A) to high patient and public health risk (Class D). These classifications focus more on the tests being run rather than the medical device itself, and multi-use devices are usually classified as Class A or B [5,48,84,104]. We chose one scenario that would be classified under the IVDR as Class A/B while still requiring threat modeling as it is software-enabled (FDA classifies these devices as Type II).

Varying the usage settings allowed us to cover a range of threat actors, such as different types of insiders and motivations. For example, focusing on insider threats, the at-home insulin pump scenario included the potential for insider threats from the healthcare provider, cloud service provider, or family member. The variation in usage setting also covered diversity of threats, such as power outages, database reconstruction, and novel attacks. There are no well-defined classes of devices. Therefore, we elected to include devices from a wide range of usage settings, including those constantly on the patient (*at-home*), in a clinical setting (*hospital*), or in a non-clinical healthcare (e.g., lab) setting (*lab*).

We chose three scenarios based on our criteria and modeled them after devices with known vulnerabilities that could be caught by threat modeling. We developed the scenarios using the FDA-recommended MITRE Threat Modeling Playbook and ISO 11073 (a medical device standard) as references [30, 35]. We then validated these scenarios through expert reviews, which we discuss in Section 3.3.

Our three mock scenarios were: a connected insulin pump and continuous glucose monitor (*Insulin Pump*), a robotic surgical system (*Robot*), and a high-throughput DNA sequence

[1]Our supplemental materials can be found at https://osf.io/p9xky/?view_only=3aa7a8a4396c4692bcabe59673c12c8f

analyzer (*Sequencer*). We showed participants context diagrams and data flow diagrams (DFDs) for each scenario. We provide the diagrams and system requirements shown to participants, along with relevant potential threat actors (which were not shown to participants) in our Supplemental Materials [1]. We used the same visual style as MITRE's Playbook for Threat Modeling Medical Devices for each diagram [30]. For each scenario, unless we specified, we left the choices of how the system was implemented up to the participants, allowing them to refine the scenario, and we asked them to state any assumptions.

Insulin Pump (implantable, Class III, at-home). Insulin pumps and other implantable devices are commonly subject to cybersecurity attention given their potential for harm of providing an overdose of medication or failing to supply life-critical care in time [81]. Recently reported vulnerabilities have included insulin pumps [20,70] and pacemakers [45]. We chose an insulin pump configuration that is becoming more popular, sometimes called an artificial pancreas [22]. This involves using an insulin pump and continuous glucose monitor (CGM) in a closed loop to provide automatic insulin dosages based on a patient's blood glucose level. The FDA/MDR classifies this type of device as Class III. Our architecture is similar to what is currently on the market [32, 69].

Surgical Robot (surgical, Class II, hospital). There is an increase in the use of connected medical devices while performing surgeries, i.e., perioperative care [41]. This includes robotic systems such as Intuitive's DaVinci [107] or the Auris Monarch [42]. These devices are complex and allow a surgeon to perform a procedure without being physically located at the same facility as the patient [63]. This introduces the opportunity for network-based threats, and attacks have already been demonstrated against robotic surgical systems [15]. The FDA/MDR classifies these devices as Class II, which requires a cybersecurity review.

Next-Generation Sequencer (diagnostic, Class II/A, lab). The final scenario was based on a device that does not directly interact with patients but is classified as a Class II device by the FDA and Class A by IVDR: a DNA sequencer (for brevity, we refer to this as *NGS*). These sequencers were important in combating COVID-19 [14] and are becoming more popular for personalized medicine for cancer treatment [73]. Like the other two scenarios, previous work has shown these devices are exploitable through many channels [76, 79].

3.3 Interview Protocol

We performed semi-structured recorded interviews using the Zoom platform. Each interview was between 60 and 90 minutes. The lead author conducted all interviews to ensure consistency. Each interview was divided into two parts: discussing the participant's background, followed by two threat modeling exercises. A summary of our study protocol is given in Figure 1 and we provide the script we used in Appendix A.

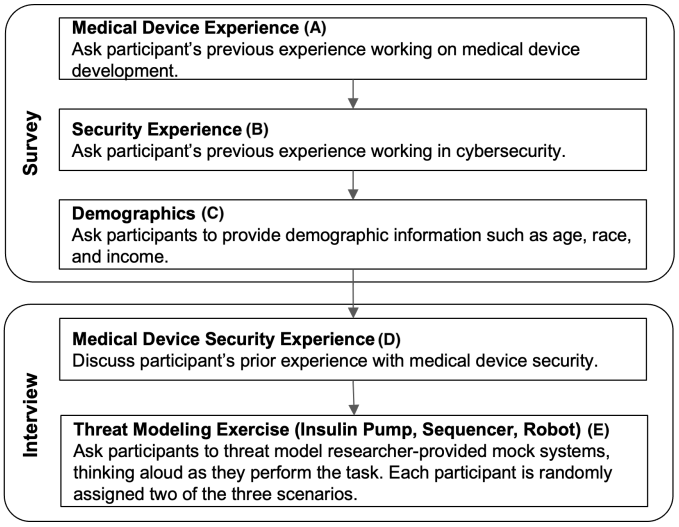

Figure 1: Study protocol diagram.

Medical device security experience (Figure 1.D). After describing the study's goal, we asked participants to discuss how they became involved with medical device security. We began with these questions as an ice breaker to ensure participants were comfortable with the interview before beginning the exercises [13, pg. 94], and also to provide helpful context. Identifying potential vulnerabilities is often driven by prior experience [112], so this question was important as it allowed us to consider how participants' backgrounds may have affected their threat modeling process.

Threat modeling exercises (Figure 1.E). Because we wanted to understand how MDM security experts identify threats and mitigations for specific components (RQ1) and how they navigate the system (RQ2), we chose to have participants threat model two systems under observation. This is preferable to asking participants to answer questions about their process generally as it helps avoid memory bias [72]. Due to proprietary information concerns, it was infeasible to ask participants to recount threat models created for devices in a professional capacity. Therefore, we presented participants with mock scenarios based on real-world devices (see Section 3.2). This had additional benefits as it allowed us to observe participants as they thought about potential threats for a device they had not previously seen, even if they were familiar with that type of device. This situation reflects MDM security experts threat modeling a system for the first time. We used the DFDs from MITRE's Playbook for Threat Modeling Medical Devices as a basis for our DFDs and their context diagrams, which are meant to show top-level functions and inputs/outputs [30, 33]. We gave participants both of these documents and walked them through the mock system design. The presented scenarios are described in Section 3.2, as well as in Appendix A and our Supplemental Materials [1].

We asked participants to review each scenario, identify potential security threats and safety concerns, and discuss what mitigations they might use to address them. We did not prescribe a specific method and instead allowed participants to work through the scenarios however they felt comfortable. This allowed us to observe a more natural representation of their process. At the beginning of each scenario, we asked participants to "think aloud" stating why they were making certain decisions as they walked through the scenario. We occasionally asked participants to restate their thinking or provide more explanation if their description was not clear. If, at any point, participants became stuck and unsure what system components to consider next, we would probe them about vulnerable situations identified during our expert review (described later in this section), such as how they would go about updating the device or what redundancies would they have in place if the hospital's network went down. We asked participants to provide more detail if they described vague threats or mitigations. We continued the threat modeling exercise for 30 minutes per scenario or until the participant had no further thoughts after prompting, whichever came first. We found that 30 minutes was sufficient for all participants. We followed the same approach for the second device. We allowed participants to return to the previous scenario if they had new thoughts while discussing the second scenario. Allowing participants to return to an old threat model represents realistic workflows as MDM security experts generate or update threat models for multiple devices simultaneously, potentially using insights from one threat modeling session to inform concurrent processes. No participants returned during our study.

Participants were randomly assigned two of three scenarios (see Section 3.2 for scenario details). We randomized and balanced the order, and each scenario was shown eight times to avoid ordering effects [85]. After finishing both scenarios, we asked participants how reflective these were of their typical threat modeling process. We asked them to explain the differences if they reported that the process differed.

Expert review and pilots. After generating our scenarios and before the main study, we conducted expert reviews to get additional feedback and ensure the threat modeling exercises were realistic and sufficiently complex to draw thoughtful responses [60, pg. 268]. We piloted our interview script with two threat modeling experts, having them complete the scenarios as participants would. We asked for feedback on questions and scenarios to ensure they matched real-world practice. We only made minor changes to our interview questions based on their feedback. The changes we made were: stating upfront in the interviews and adding text to the diagrams to indicate they were for reference and not exhaustive. The change was made to speed up the interview process and did not change the information provided, as the pilots were eventually given the same information once we realized the confusion. Then, two former regulators who worked on regulation for medical device threat modeling and two senior medical device security experts reviewed the scenarios. They were asked to confirm whether the scenarios matched their real-world experience and covered the most common devices. They added some minor details to the scenarios but indicated they were realistic. These experts who were involved in designing the scenarios were not allowed to participate in the study to avoid bias. Also, due to the small size of the community, we explicitly asked the experts involved in this feedback and our participants not to share details of our study with others.

We piloted the interviews with two MDM security experts for additional face validity review. No changes were needed. Due to the difficulty recruiting in this domain (Section 3.4), we included both pilot participants' responses in our results.

3.4 Recruitment

We used three approaches to recruit participants for our study: contact through organizations and MDM-focused conferences, posts on social media, and personal contacts. These are standard recruitment methods in security expert research [55].

Associated organizations and conferences. Our primary recruitment method was through organizations and conferences associated with medical device security. We were allowed to discuss our study's motivation and goal and solicit participants during a Healthcare Sector Coordinating Council (HSCC) meeting. HSCC is a CISA-established public-private partnership developing cybersecurity recommendations for medical devices [3]. We also publicized our study at several conferences on medical device security, including the Biohacking Village at DefCon [25], CyberMed Summit [100], and the Archimedes Center's Medical Device Security 101 [9]. These conferences are attended by MDM security experts engaged in security for medical systems. We recruited six participants through this method: three from our HSCC presentation, two from Archimedes, and one from DefCon.

Online advertisements. We posted advertisements about our study on LinkedIn using the tags: #medicaldevices, #security, #fda, and #infosec. We had a leader in this domain post the advertisement as well. We did not find this method productive, as no qualified participants responded to these ads.

Direct contact and snowball sampling. Finally, we used the lead author's MDM business contacts to reach out to possible participants directly. For each person we interviewed, irrespective of initial recruitment method, we attempted to snowball sample by asking if any of their peers might be interested in participating—which is often necessary to recruit in small, niche fields where broader recruitment methods are often impossible [52]. We recruited one participant using snowball sampling, and the other five were recruited through the first author's business contacts.

Our recruitment messages described the study's motivation, the value of its findings to the community, and eligibility requirements. Interviewees were given a $40 gift card for par-

ticipating. We interviewed participants until we reached saturation of themes to ensure rigorous qualitative results [18, pg. 113-115]. While we reached saturation quite quickly (after 5 interviews), we continued interviewing participants until we reached the minimum sample size suggested by qualitative research best practices (i.e., 12-20) to provide strong guidance for future quantitative work and develop generalizable recommendations for design [43]. We believe our sample is sufficient as we found through our community engagement that this population is small and closely connected, as evidenced by how quickly we reached thematic saturation.

Recruitment Criteria and Survey. Our only requirement for interview participation was that interviewees worked for or with MDMs conducting threat modeling. To solicit information about their experience and qualifications, potential interviewees were directed to complete a screening survey (the full text is in our Supplemental Materials [1]). The survey included three parts (Figure 1.A-C): their medical device development experience, security experience, and demographic questions. All participants who reported working on medical system design and development for or with an MDM were invited to participate in the interview. Participants were required to provide a resume or their LinkedIn profile for verification.

3.5 Analysis Methods

We analyzed the interviews using iterative open coding [24], splitting each interview into three parts: medical device security experience, scenario one, and scenario two. We split the scenarios to investigate how the different scenarios affected participants' responses and classify the scenario during which a threat was identified. After each interview, the research team used Rev.com's automated service [86] to transcribe the recorded audio. A researcher compared the transcript to the original audio to confirm accuracy.

After all the interviewers, the interviewer and one researcher independently coded four scenarios (i.e., two interviews) and discussed codes among the entire research team to generate the initial codebook. The codebook [1] consisted of four high-level categories discussed by participants: the approach to identify threats (*Approach*), such as considering particular user workflows or the more formal per-component interaction of the STRIDE threat modeling process [91]; the configurations considered for the device (*Configurations*), such as the environment the device would be used in or if a specific part of the design was dependent on business requirements; the security controls or mitigations (*Mitigations*) suggested for system security and safety; and the threats or security properties (*Threats*), including safety, an attacker might violate. Using the initial codebook, the two researchers coded subsequent scenarios in rounds of four, calculating inter-rater reliability using Cohen's Kappa [23], resolving disagreements, and adjusting the codebook after each round. We continued this process until we achieved a $\kappa \geq .8$ for each category, indicating strong agreement and ensuring codebook reliability [23]. We specify the agreement for each category in the codebook [1]. We coded only explicit mentions of threats and mitigations. For example, a participant mentioning encryption to protect data transmitted over the Internet without stating the threat this is intended to mitigate (e.g., data leakage). In this case, we would only code the mitigation of encryption, not the unstated threat. This conservative approach prevented us from introducing our personal bias.

Next, we performed axial coding, identifying connections within and between categories, allowing us to identify common themes and relationships [24, pg. 123-142]. We observed ordering between considerations of each item and theme in how participants moved between them, which allowed us to establish our model of the threat modeling process to cover all the strategies in which participants considered various codes. The codebook [1] contains the detailed definitions.

3.6 Ethical Considerations

Our IRB approved this study. We obtained informed consent before the screening survey, confirming with participants that they understood the consent before the interview. While we designed the study to avoid discussing company propriety information, we informed participants they could skip any questions they might be uncomfortable answering, pause the recording, redact any accidental disclosures after the interview, or terminate the interview at any time. Finally, because the community of MDM security experts is so small, it is possible other community members could identify our participants with minimal information. To prevent deanonymization, we describe participant experience using ranges instead of exact years in Section 4 and select quotes throughout our results to avoid revealing too much information.

3.7 Limitations

Like most interview studies, our sample size is small, and the generalizability of our results has limits. For each finding, we give the number of MDM security experts expressing or demonstrating a theme to indicate prevalence. However, if a participant did not mention or demonstrate a specific theme, that does not necessarily indicate disagreement or lack of applicability to their threat modeling practice. This is especially true as we took a conservative approach, only coding explicitly stated threats and mitigations. Therefore, we do not use statistical hypothesis tests to compare groups. For these reasons, our results may not generalize beyond our sample; however, they suggest directions for future work and provide novel insights for threat modeling.

We specifically focused on medical devices rather than a more generalized system, and so our results may be limited to this domain. However, we chose this approach because we wanted to ensure our scenarios resembled real-world systems

as closely as possible, which requires specific and detailed information. Because this information and the domain expertise necessary to effectively threat model varies greatly between domains, we focus on a single domain as a necessary initial step. Given the need to select a specific domain, medical devices provide the ideal exemplar. Threat modeling has been a common practice in the medical device domain as it is required by regulation [17, 66, 67] and there are many tailored trainings [65, 68, 89] and resources [19, 30, 50, 53] specific to medical device threat modeling. Therefore, it is reasonable to expect MDM security experts to be familiar with threat modeling, and our results likely reflect the ceiling of threat modeling expertise. But as "secure by design" approaches continue to develop [17, 21, 56, 102], we expect other domains to begin to look more like the medical device domain, especially other cyber-physical systems, which share similar characteristics for safety criticality. By focusing on a regulated domain that requires threat modeling, it is possible the regulation could introduce bias toward a specific process. As we show through a detailed discussion of existing regulations in our Regulatory Overview [1], no current regulation prescribes a specific threat modeling method.

Because we generated mock device scenarios, these scenarios may not be realistic. However, we developed our scenarios using actual devices shown to have vulnerabilities, as discussed in Section 3.2. We also conducted expert reviews (see Section 3.3) to verify scenario realism and at the end of each interview, we asked participants if the scenarios were realistic. All the expert reviewers and participants found the scenarios matched their experience, and one asked if they could use one of our scenarios as a template exercise when training others. These scenarios are also not exhaustive. We chose our scenarios to cover a wide range of threat modeling considerations, but some problems were not represented. However, because we use the scenario to prime the discussion, but asked participants to describe their relevant experience broadly, responses were not limited to the scenarios. For example, as we discussion in Section 5.1, some participants described supply chain threats, which were not covered in any scenario.

We also conducted the study in a lab setting, which is not necessarily realistic for the task. When threat modeling their own devices, users might rely on guidance documents, tools, organization-prescribed processes (such as STRIDE), or other members of their teams. We did not require participants to use any tools, processes, or documents that their organizations might require in practice. While this potential setting change does not allow us to capture the exact results of an internal threat modeling process, it allows us to focus on each participant's thoughts independent of workplace structures. Participants knew the focus of the study was security, and there were no strict countervailing forces, such as business interests, beyond the basic scenario functionality. Therefore, our results should be considered a likely best effort, where MDM security experts might be more likely to discount some threats when weighing other organizational factors more heavily. Participants also had limited time to complete the exercise, potentially limiting the depth of their review. The time was sufficient for all participants to complete each exercise, and they indicated it was similar to their experience in practice.

While we did not restrict our recruitment efforts to the US, ten participants were based in the US, and eight worked for US companies. However, all participants had experiences with devices approved for use outside the US, and all MDMs represented have global distribution. This suggests our results likely generalize to participants operating under non-US regulatory regimes. While we did not observe any apparent differences among our two non-US participants, there may be some culture-specific differences in MDM security expert threat modeling in other countries that we did not capture.

Finally, while we did not explicitly limit the selected participants to those reporting security experience and did not state this as a requirement, all our participants reported some security experience. However, this varied as some were more junior than others. This may indicate selection bias [10], making our results representative of more ideal threat modeling considerations when experienced practitioners are available to review. However, based on our community discussions, we believe this is simply characteristic of the population actually threat modeling medical devices. Our work offers a needed first step, but future work should investigate additional perspectives, such as software developers and clinical engineers.

4 Participants

In total, we interviewed 12 MDM security experts. Our participants were all highly educated. All had at least a bachelor's degree and half had a master's degree or doctorate, though none were medically trained. Our participants were mostly men (N=10), white (N=9), and older (45 years old on average). Table 1 shows our participants' backgrounds, current roles, and years of MDM security expert experience. Our participants were generally more experienced (i.e., most have a decade of experience). Therefore, our results present a more expert perspective of the field. However, three participants were more junior, giving us some insights into earlier career MDM security experts. We attempted to recruit clinicians and more junior personnel. However, we found in our community engagement (Section 3.1) that clinicians are rarely involved in development, much less threat modeling. Junior personnel, who we did reach, stated they lacked experience as medical device threat modeling is typically carried out by those in the more senior roles held by our participants. We confirmed this in our discussions with community leaders.

When asking participants how they got into medical device security, they all began focusing on general medical device development (e.g., as electrical or biomedical engineers) or general cybersecurity (typically in enterprise security) and later transitioned to focusing on medical device security specif-

	PID	Role	Exp.	P[1]	R[2]	S[3]
Med. Device	M1C	Dir. of Product Security	30+	X (5)	X (4)	-
	M3C	Security Consultant	30+	-	X (4)	X (3)
	M4S	Quality & Security Program Manager	20-30	X (5)	-	X (4)
	M6L	Biomedical Engineer	1-5	X (4)	X (3)	-
	M11L	Product Security	20-30	X (5)	-	X (1)
	M12L	Security Compliance	5-10	X (5)	X (4)	–
Security	S2S	Dir. of Product Security	20-30	X (1)	X (1)	-
	S5L	Dir. of Product Security	30+	X (5)	-	X (3)
	S7S	Cybersecurity Engineer	1-5	X (2)	-	X (2)
	S8C	Consultant	20-30	-	X (3)	X (2)
	S9C	Security Consultant	10-20	-	X (3)	X (2)
	S10S	Dir. of Product Security	10-20	-	X (4)	X (4)

[1] Shown the Insulin Pump, [2] Surgical Robot, [3] NGS.
The number in parentheses indicates the level of familiarity with the class of device (from Figure 1.C) using a five-point Likert scale from Extremely familiar (5) to Not at all familiar (1).

Table 1: Participants' backgrounds, roles, and years of relevant experience. PID shows the participant's background (S - security; M - medical device), and industry perspective (C - consultant; L - large MDM; S - specialized MDM).

ically. Participants were evenly split between backgrounds in general medical device development (N=6) and general cybersecurity (N=6). Our participants varied in their familiarity with different classes of medical devices, including lab equipment, surgical systems, IMDs, diagnostic equipment, home health systems, life support, and software as a medical device. When asked about scenario familiarity, seven participants were either extremely or moderately familiar with the device for at least one scenario they were shown, and of those, three were extremely or moderately familiar with both.

Participants spanned many organizations, and none worked directly for the same company. Four participants are security consultants for MDMs, four work for specialized MDMs (i.e., they produce devices for a single or small number of specialties), and four work(ed) for very large MDMs (i.e., top ten largest MDMs by revenue). All participants had dedicated security teams within their organizations and were familiar with threat modeling and the FDA cybersecurity guidance.

5 Threat & Mitigation Elicitation Process (RQ1)

We now discuss how MDM security experts threat modeled our scenarios. In this section, we cover how they elicited specific threats and mitigations. In the next section (Section 6), we discuss how MDM security experts worked through the system. Throughout our results, we present counts to denote the prevalence of particular themes. When themes are related to the participants themselves, we use *P* to show how many of the 12 participants expressed that theme. For themes related to a specific scenario, we use *S* to show how many of the 24 scenario responses (i.e., two per participant) mentioned it.

Use of the four questions. To determine threats and mitigations for a particular asset, we observed participants using an implicit structure similar to the Four Questions (see Section 2), focusing mainly on the last three questions: "What can go wrong?", "What are we going to do about it?", and "Did we do a good enough job?" We answered the first question, "What are we building?" with the scenario, and participants occasionally asked clarifying questions.

This implicit structure arose from our data during qualitative analysis. We did not attempt to restrict our coding to any existing framework. However, while participants used this existing structure at a high level, they did not address the questions in sequential order and even showed variation in their order of questions between components. Further, we observed additional detail in the sub-questions in each of the four questions. One of these details we saw considered throughout the process was safety. However, how safety should be treated in the context of threat modeling was unclear for participants. We summarize this process in Figure 2 and will discuss each question below, because safety was seen throughout, we provide a focused discussion in Section 5.4.

5.1 What Can Go Wrong?

Threat modeling's central goal is determining how a malicious actor might influence the system to some negative effect. Participants considered this question across three dimensions: threats, security properties, and safety. Threats were either explicit attacks (e.g., SQL injection) or broader categories found in frameworks like STRIDE (e.g., spoofing). Conversely, security properties were the security guarantees participants sought to ensure (e.g., confidentiality or authenticity). Safety impacts were the potential harms a device might cause a patient (e.g., a malfunctioning device provides an overdose of medication); while not traditional to threat modeling, it is expected for medical devices [57, 64]. There was no strict order participants followed in answering these questions.

We found a natural connection between threats and security properties. When participants began with a specific threat, they would either explicitly or implicitly connect it to a security property. During our analysis, we made this implicit connection by the security property associated with that class of threat. For example, integrity is a security property, but most examples centered on tampering (S=16), a threat classification from STRIDE. However, we did not observe participants mentioning a specific threat if they started with a security property. Common classes of security properties were availability (P=12), confidentiality (P=9), and integrity (P=12). This is what we expected, and all the identified threats matched those found during experts review (See Section 3.3).

The specific threats mentioned covered several different groups of threats and threat actors. Many of these ideas stemmed from personal experiences or anecdotes that par-

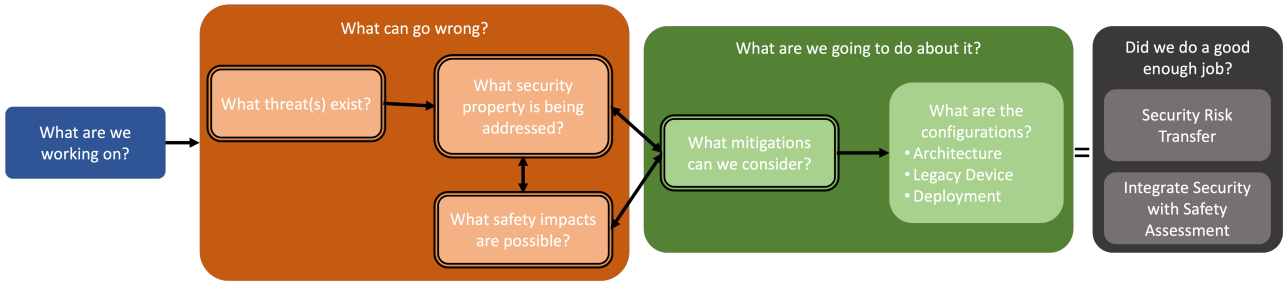

Figure 2: Threat modeling process using the key questions participants considered, following the four-question framework with specific additions. Double-borders indicate questions we observed participants start with. Arrows show a transition from one question to the next by at least one participant.

ticipants referenced. For example, M1C, who is extremely familiar with the IMDs, immediately focused on the potential for weak encryption on the Bluetooth low energy packets in the Insulin Pump. They used that to discuss the trade-offs between encryption algorithms in medical devices, stating they recommend symmetric encryption for ensuring packets are processed quickly to ensure correct dosing of insulin, and to "install those keys during manufacturing put in shared secrets, don't share them necessarily at runtime, make them highly ephemeral, make them very short-lived." Relying on experience was common, even when a participant was less familiar with a class of device. S2S, who was not at all familiar with surgical robots, first looked at the surgical cart for the Surgical Robot and said, "If it is listening for a connection, that means it's potentially open ports...if there are open ports, that means the operating system could be accessible." They then told us they thought about this as they know manufacturers tend to leave some ports open to ship patches. While participants often relied on their experiences to brainstorm relevant mitigations and deployment settings, we did not observe any discernible relationship between participants' processes and scenario familiarity. M3C talked about potential power and/or connectivity loss in the Surgical Robot, which they said could be caused by human error or a malicious actor.

Tools & Brainstorming. While we did not explicitly ask about tool usage during threat modeling, some participants discussed tools organically. The participants who did mention a specific tool only spoke about the Microsoft Threat Modeling Tool (which has a medical device template). During our community engagement, we also heard mainly about the Microsoft Tool, although it appeared to be mainly used for diagramming. Both M1C and S7S mentioned looking over the threats listed, exporting those to a spreadsheet, and manually reviewing, rather than using the tool to conduct the analysis. M4S talked about tools more generally, and the need for tools to integrate more in their processes.

5.2 What Are We Going To Do About It?

Threat modeling is a defensive action, so MDM security experts must also consider how to protect their systems from attack. Again, we observed participants identified mitigations as expected from our expert review, the most common being authorization (P=12), monitoring (P=12), and authentication (P=11). The only exception was language-based mitigations or formal software verification, which only one participant discussed. We followed up with several participants who explained that organizational factors make using these mitigations challenging. M11L enumerated several challenges, including, "availability of current or hirable programmers. . . code comes from an acquired start-up and is too costly to change. . . decision makers that don't understand the issue well." These reasons are similar to what prior work has found to be challenges of adopting languages such as Rust [39].

Mitigations often flowed from what can go wrong. When looking at participants' process, the question of "What are we going to do about it?" naturally followed by defining what "it", i.e., the security property or safety impact, was. For example, S2S said, "Bluetooth is worrying because it's easy to sniff, in some cases." before going on to talk about authentication and confidentiality-related mitigations.

Several participants started by considering mitigations. While many participants followed a linear path through the questions, that was not always the case. Rather than considering "What can go wrong?", we saw several cases where participants' first reaction when reviewing a component was to suggest applicable mitigations. For example, M11L cited their experience with other types of IMDs, considering hardware options with the Insulin Pump, suggesting using hardware limits to ensure safety. "(We) put hardware monitors in there to ensure those safety limits were never exceeded. So it's just defensive design." S7S looked at the connection between the system and analytics server in the NGS scenario, "If the hospitals are in charge of setting it up themselves, ideally I'd put it on a separate VLAN and then have more individual access for that. . . it'd be more role-based access."

Another example we observed was that few participants

mentioned the threat, elevation of privilege (P=5). As we indicated before, all the participants mentioned authorization mitigations such as role-based access control or least-privilege, which are related to elevation of privilege, possibly indicating implicit consideration of elevation of privilege. Another threat not as commonly discussed was non-repudiation (S=10). Again, of those that did not mention non-repudiation (P=5), all five suggested the need for auditing, possibly implicitly considering non-repudiation.

Mitigations' configurations depend on the deployment. After selecting a mitigation, participants considered the mitigation configuration variations that are needed based on the setting. The most prominent consideration (P=11) focused on the clinical setting's effect on mitigation implementation and requirements for interoperability with other technology. S5L described wide variation between hospitals, "When you've seen one hospital, you've seen one hospital. They're so unique." Then, they discussed the challenges hospitals face as "you introduce all of these different manufacturers, bringing in all these different things they want to stick on their network, and how are they supposed to know if they're configured securely or not?" Participants reported the deployment impacting the network set-up (P=10), risk transfer (P=10), integrity (P=9), authorization (P=8), and authentication (P=8). One common example was whether a hospital used directory services, such as Windows Active Directory (AD), for authentication and authorization (P=6). This transfers security risk to the hospital as they are required to manage who has access and ensure sufficient security checks are in place. Alternatively, hospitals with fewer resources might not have AD and rely on MDMs to provide these services in the device design.

The Deployment Setting was explicitly considered for the NGS (S=8/8) and Surgical Robot (S=6/7). This included customers' resources or requirements (Surgical Robot: S=5/7; NGS S=6/8) and the clinical setting, i.e., in a hospital or a clinic (Surgical Robot: S=5/7; NGS S=6/8). For the NGS, the different settings, for both the sequencer and the providers or researchers, included hospital labs, independent labs, forensics labs, and research institutions. Additionally, some participants considered the dynamic nature of the deployment. For example, while the surgeon likely would change locations, the Surgical Robot itself would not.

Architecture impacted mitigation implementation. Different system functionality could tolerate different security overhead without causing clinical impacts, and these differences drove many participants' considerations for mitigation configuration (S=18). Participants relied heavily on their experiences when discussing various configuration issues. For example, S7S discussed needing architectural changes related to the NGS as "I know, like personally, a lot of the radiology systems do not always have encryption, so it's difficult to implement encryption if the other thing is not doing anything." M3C discussed how functional differences between the NGS and Surgical Robot impacted their integrity controls' implementation saying, "there's a lot of devil in the detail. So, for example, in both scenarios, I talked about you signing messages. That's a very different problem in the lab analyzer because data flow latencies are not a problem... with the surgical robot, where latency is extremely sensitive... the technical implementation would probably vary widely because of timing and system resource considerations."

Considering issues with legacy devices. Many medical devices are expensive, requiring a long deployment lifetime, making them hard to replace. Resources are constrained within healthcare [114], making it unreasonable to expect devices to be replaced frequently. Switching devices–even if the original MDM goes out of business–can cause a loss of clinical functionality necessary for life-saving care. Unlike commercial technology, where users and companies regularly upgrade to newer models [44], medical devices face a unique challenge with legacy devices. Regulators explicitly recognize devices "that cannot be reasonably protected against current cybersecurity threats" as a concern [31, 103].

While we did not mention legacy devices in our scenarios, participants (P=5) talked about how their experiences with them impacted their decisions about certain mitigations. Participants mentioned how updating and patching are important to ensuring a device is not legacy but that there are constraints based on the Deployment Setting. S9C said they would "(prefer) a remote update...(and) notify the hospital." However, when considering remote updates, M11L mentioned some hospitals insist on applying these themselves. Additionally, S7S mentioned how they have clients with limited internet access and "service technicians that will go on-site and do the updates like with a thumb drive." This is further complicated when required updates might involve a third-party operating system, such as Windows (P=7).

5.3 Did we do a good enough job?

We did not see any participants evaluate if there were holes in the threat model. Instead, they focused on whether security risk remained that would need to be transferred to another party and prioritization of mitigations concerning safety.

All participants discussed risk transfer. When trying to determine mitigations, all participants, at some point, concluded that a subset of threats could not be removed by implementing changes to the device. Therefore, risk must be transferred to another party, e.g., the user or third-party service. For example, when evaluating the Insulin Pump, some participants mentioned the need to rely on the patient's phone's operating system to protect the device's application from being exposed to other malicious apps (S=4/7). Risk transfer with medical devices has evolved, which M3C described, "traditionally, there has been the tendency to push security responsibility down to the healthcare organization, as in, I can't give you

more secure devices, but I can tell you how to secure your network. . . That has started to change. I think we still have some way to go. Some people talk about that shift-left approach, as in design security as early as possible." We note that participants attempted to remove security risks whenever possible. For example, there was concern with the use of third-party cloud providers (Insulin Pump S=4/7;NGS S=5/8), which many participants felt were over-relied on for security protections as they suggested other MDM security experts might assume the cloud is "secure by default." Instead, participants pointed to the potential for a third-party cloud provider to introduce a security vulnerability through misconfiguration or make changes that break their security assumptions, causing non-trivial risk introduction. For these reasons, they chose to control their deployment as much as possible.

5.4 Safety Considerations

MDM security experts prioritize their analysis for safety risks. For example, S10S said, "If you have two other servers that are used for data analytics or some other intended use to do some type of calculation in the background but doesn't impact what the robot is performing on the patient, then I can look at that from a risk perspective that, 'Oh, it's not impacting the intended use. It may be impacting patient data, but from a safety aspect, it's not posing a safety risk to a patient.'"

Safety as something that can go wrong. As expected, safety was a paramount concern when discussing threats (P=11). For instance, a denial of service on the Surgical Robot could have safety impacts if it delayed instructions sent by the surgeon. M3C explained, "integrity of the data that flows across the system, as well as the availability of the data flow and both, could result in harm to the patient." M4S mentioned assessing threats for safety, "when we assess them, we also do potential safety impact. . . the ones with the potential safety impact, we are translating them into our risk management process."

We also observed some participants focusing first on potential safety impacts (S=11) and using that to guide the properties they might want to guarantee. For example, S9C said, "So if the surgeon moves the robot for one millimeter, you need to make sure that somewhere in all those data connection points on the hospital servers in the software, a millimeter doesn't, by accident, get translated into a centimeter. . . So the integrity is almost as important as the availability for this data."

As discussed in our Regulatory Overview [1], the FDA asks MDMs to analyze the potential for multi-patient harm. Interestingly, we saw only two mentions (S=2/7) of this explicitly. S5L, when talking about the Insulin Pump referred to the manufacturer's cloud data storage (specified in our diagram [1]) and that an attacker with access to the manufacturer's cloud data storage could steal data or manipulate all the users.

Weighing mitigations in light of safety. Participants thought about safety as it relates to controls that they can use. For example, ensuring the longevity of an implantable device is critical. S5L explained when discussing the Insulin Pump, referring to their experiences with implantable cardiac devices, "You can't have high technology, high energy solutions for authentication." M3C mentioned validating packets separately from the operation of the Surgical Robot to ensure that it would not slow down time-critical operations rather than waiting for the validation before execution. They went on later to discuss using a heartbeat to ensure the connection is not interrupted, which would cover both a system failure as well as Denial of Service. For the Surgical Robot, S10S discussed how they focus in on "clinical workflows for devices is something where we, in product security, typically don't jump right into, but especially when you come into things like threat modeling...if we're not protecting those critical services, essentially we are missing a big piece of it. We can't just look at where data resides, we can't just say "Hey, harden your servers," and things of that general statements. We have to really look at the function and what the data that's flowing between each component to understand and wrench its impact to affecting that clinical workflow."

Integrating safety and security risk management is essential, but there is no clear process. Our participants discussed cases where security and safety conflicted. M1C, our most experienced participant, talked about how they think about safety is different from security and that they "cannot hold both concepts in my mind at the same time, it is one or the other. I'll look at this and go, "Oh yeah, we can add those things." And then I stop, put on my other hat, look at it, and go. But that creates these problems." For example, replacing an insecure device may improve security but introduce safety harms. S8C described this saying, "extra security is not needed to guarantee safety. For example, a completely insecure device can be perfectly safe if every decision is second-guessed and it never directly touches a patient." This trade-off between security and safety might be in conflict, and some standards discuss establishing risk/benefit processes to balance these considerations [2, 34, 35].

Because it is necessary to consider this balance between security and safety, several participants (P=6) discussed the need to integrate security and safety risk management processes. By integrating these processes, they can determine the safety impact of system failures due to exploitation or system degradation caused by mitigation deployment costs (e.g., slowing down network communication by adding encryption). Our participants indicated there are no suitable methods for integrating their threat model findings into their organizations' formalized safety risk management processes. This makes evaluating and prioritizing mitigations difficult for MDMs. Participants focused on the communication challenge between experts (i.e., security experts explaining security problems to safety experts and vice versa) as the primary issue facing integrating safety and security. S8C explained, "the integration of this is very important, and we have separate processes that have synchronization points, but without necessarily the two

groups understanding each other, it [potential miscommunication] is pretty dangerous." S7S mentioned using a spreadsheet to gauge the potential for safety risk quickly, but they did not fully understand the mechanics of the safety analysis.

6 System Navigation (RQ2)

In the prior section, we described how participants determine threats and mitigations when considering each system component. We now describe how they navigated components to produce a thorough system threat model. As we did not restrict participants to a specific framework or order, our results demonstrate their natural approach to the task. We observed four general approaches: walking through the components sequentially (P=4, S=6), reviewing each component interaction (i.e., communicating pairs of components) sequentially (P=1, S=1), focusing on specific use cases (S=6), or using an ad-hoc approach (S=11) (i.e., using a combination of the prior approaches, varying the way they navigated the system). Participants were not consistent in the process used, with almost half (P=6) following different approaches between scenarios.

Sequentially going through the model was not common. The most obvious approach to ensure a thorough review would be to work across the system diagram left-to-right or top-to-bottom (or in whichever order the MDM security expert would usually read), considering each component or connected pair of components in turn. These per-component or per-interaction approaches are recommended by threat modeling guides [91, 96, 101]. These approaches can have the benefit of simplicity, allowing the user to know they have considered all components (or interactions) once they reach the other side of the diagram. However, we observed few participants following this approach; only four participants (and only on six of the eight systems they reviewed) considered the systems per component in sequential order. Only one participant used a per-interaction approach for one scenario.

Use cases to organize navigation. Instead of working through the system sequentially, some participants (P=5) considered paths associated with potential use cases (an additional three participants leveraged use cases in their ad-hoc approach). Participants considered how components would coordinate end-to-end to enable a clinical workflow and identified threats impacting the workflow and mitigations to ensure the function could continue in the face of an adversary. For example, S10S started by looking at the critical clinical workflows for the *Surgical Robot*. Here, the clinical workflow is performing surgery whereby the surgeon will manipulate controls on their workstation, translating to small movements on the robot and the tools that are physically operating on the patient, as well as the telemetry and video feed sent back to the surgeon. Other workflows were discussed around less critical functionality. For example, S9C focused on how components interacted to allow patching, a process that regulators highlight in their cybersecurity guidances [17, 66, 103]. S9C walked through how they would want to perform an update for the NGS, "contact an update server and pull to see if there is a new version. If a new version is ready, again, I would always implement a manual step there... after there should be some kind of mechanism to calibrate the machine again... to ensure the integrity of this machine, make sure that the data that comes out is actually correct."

For a thorough review, MDM security experts must consider all use cases, as each use case only considers a subset of component functionality (i.e., those relevant to the use case). Enumerating all use cases is complicated and, like the sequential approach described above, does not guarantee a complete threat model. However, our participants pointed out multiple benefits. It helps MDM security experts consider interactions across the system, and see non-obvious issues arising from these interactions that would be missed if viewed in isolation. S10S talked about focusing first on critical workflows to understand priority assets, "When you have specific function and workflows that you designate that this product is going to do. Automatically, those become critical high-severity items because if those workflows aren't protected, essentially the device is not performing at its intended use... Because if you look at the clinical workflows that are critical for this device to do its intended use, it's going to actually help drive what you should be looking at from a security standpoint."

Participants navigated the system in an ad-hoc manner. While we did see some participants follow a single approach, they most often conducted an ad-hoc review (P=8, S=11) where they used a mix of previously described approaches or the approach did not have a discernible structure. For example, participants might begin using a per-component approach, but switch to a use-case approach, when their elicitation of threats on the component led them to think about a critical use case. As another example, multiple participants considered a potential threat or mitigation and paused their sequential or use-case approach to walk through the system and identify other points where the threat or mitigation might be applied.

While one might assume an ad-hoc approach would not be seen as effective (as it is not systematic and difficult to track) our results suggest this was not the case. Instead, many participants chose this more flexible process. It removed structural restrictions and allowed them to consider threats and mitigations as they came to them and review issues associated with interactions across the system. They also leveraged their expertise to focus on the most critical aspects of the system. For example, S8C said "there are rabbit holes I want to go down that I'm not allowed to go down by the threat modeling process" as a reason for taking a more ad-hoc approach. It could be that this unstructured approach is an artifact of our study design due to the mock nature of our scenarios. However, when asked whether their approach to our scenarios matched their real-world approach, all participants indicated they used this ad-hoc approach in practice during threat modeling.

7 Discussion

There has been a significant body of knowledge developed over time to support threat modeling through method and tool development [6, 26–28, 37, 51, 54, 58, 71, 88, 92, 105, 106, 109, 118] and evaluation [7, 16, 56, 57, 62, 64, 108, 116]. However, until now, it has been unclear how security experts threat model in practice. We begin to close this gap by taking a first step toward understanding the real-world threat modeling practices of MDM security experts, building a necessary foundation for future research both for more general threat modeling and "secure-by-design" medical devices. Specifically, we identify a two-part process model enumerating first the ways MDM security experts elicit threats and mitigations while looking at components in the system and then how they choose to navigate through the system. We found that while MDM security experts' considerations were generally in line with existing frameworks, specifically the Four Question approach [118], we identify further sub-questions MDM security experts ask. We saw MDM security experts do not restrict the order in answering questions, allowing ideas to flow and associate as they have them. Similarly, when navigating between components, we found participants did not follow a specific process, instead taking an ad-hoc approach, switching between sequential, use case, and other less defined approaches. These results suggest it is vital for this process to be free-flowing, avoiding strict structures that can limit creativity and brainstorming [77]. However, our findings do not assess actual process efficacy. Therefore, future work should evaluate ad-hoc and use-case-based approaches similarly to prior work which has tested sequential approaches [7, 16, 57, 62, 64, 108]. Additionally, our findings suggest that more work should be done to integrate safety evaluations with threat modeling proposed in prior work [83, 117], and evaluated in human subjects studies. Building on our observations, we describe how researchers and tool designers can use our results.

Recommendations for Researchers. As our research is primarily exploratory, our first contribution is laying the foundation for future work. Our process model can be used to structure experiments and focus on process-critical aspects. For example, considering the component review part of the MDM security experts' process, researchers could design tasks to focus participants on a single question (e.g., eliciting threats). This would limit experiment participation time requirements, allowing broader recruitment to provide sufficient statistical power. While our results suggest significant interrelation between questions and wide variance in how they are approached, we outline these relationships, allowing the researcher to provide some control information to the user (e.g., a list of mitigations) to allow natural exploration while removing the burden of mitigation identification.

Our process model is valuable to future researchers evaluating new approaches to threat modeling. As we described in Section 2, prior work has evaluated new threat modeling approaches using systematic audits conducted by the research, meant to mimic practitioner behaviors [7, 16, 57, 62, 64, 108]. Our results suggest this evaluation approach does not accurately reflect how practitioners threat model. Instead, researchers developing novel approaches designed to be used by practitioners (rather than fully automated threat modeling) should consider our more ad-hoc and fluid process model.

We have made our device scenarios available for future research (see supplemental materials [1]). These scenarios were carefully built through community engagement and viewed as clear and representative of real-world systems by our participants. They allow future work to skip this challenging step of realistic scenario development, lowering the barrier to entry.

Guidelines for Threat Modeling Tools. Automated tools can increase threat modeling adoption, support MDM security experts with less experience, and guide a more systematic review. As discussed in Section 2, threat modeling tools exist, but several participants mentioned they do not use tools other than for creating data-flow diagrams, as it was not central to our research questions; we did not probe this further. Psychology research on creativity has shown that brainstorming processes (e.g., threat modeling) are best fostered by supporting natural ideation alongside systematic structure to guide the process [77]. To achieve this goal, automation must match MDM security experts' natural, flexible approach. We provide guidelines from our results for naturally usable automation:

G1 Support free-flowing process through interaction. Users have varied approaches to threat modeling, often adjusting their process while looking at the same system. Automation should fade into the background [94, pg. 19-22], suggesting threats and mitigations related to the current focus, only broadening focus when it appears the user is stuck.

G2 Support for multiple configurations. Various configurations may exist for a single system. Threat modeling tools should support this annotation and recommend common configurations the user may not have considered.

G3 Provide use-case views. Because there are a variety of approaches when navigating the system, tooling should allow them to cycle through alternative visualizations. This includes allowing them to isolate specific use cases, as this was common among participants.

G4 Prompt for multi-patient harm. Few participants (P=2) considered multi-patient harm. This determination is important from a regulatory and safety perspective. Tools should take care to help users consider these harms as they are unlikely to flag them otherwise.

G5 Integrate with the safety risk process. Safety risk evaluation is an essential process for MDM security experts. MDMs follow well-established processes for safety risk evaluation and rely on existing tooling. It is prudent that tools integrate with existing safety evaluation tooling.

Acknowledgements

We thank all of our participants for their time and expertise. We also want to thank the anonymous reviewers who provided helpful comments on drafts of this paper; Peter Ney, Rock Stevens, Seth Carmody, Naomi Schwartz, and Shannon Lantzy for valuable insights on the design of the study; and Greg Garcia and HSCC [3] for helping us with recruitment. Gifts from MedCrypt and Cisco supported this project.

References

[1] Medical device threat modeling interviews supplemental materials. https://doi.org/10.17605/OSF.IO/P9XKY.

[2] Principles for medical device security - Risk Management. Technical Report TIR57, Association for the Advancement of Medical Instrumentation, 2016.

[3] Health Sector Coordinating Council, 2022.

[4] Yasemin Acar, Michael Backes, Sascha Fahl, Simson Garfinkel, Doowon Kim, Michelle L Mazurek, and Christian Stransky. Comparing the usability of cryptographic apis. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 154–171. IEEE, 2017.

[5] Agilent. In Vitro Diagnostic Regulations | Agilent, 2023.

[6] Christopher Alberts, Audrey Dorofee, James Stevens, and Carol Woody. Introduction to the OCTAVE Approach:. Technical report, Defense Technical Information Center, Fort Belvoir, VA, 2003.

[7] Hussain Almohri, Long Cheng, Danfeng Yao, and Homa Alemzadeh. On threat modeling and mitigation of medical cyber-physical systems. In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 114–119, 2017.

[8] Edward G Amoroso. *Fundamentals of computer security technology*. Prentice-Hall, Inc., 1994.

[9] Archimedes Center for Health Care and Medical Device Cybersecurity. Medical Device Security 101.

[10] S. Ards, C. Chung, and S. L. Jr Myers. The effects of sample selection bias on racial differences in child abuse reporting. *Child abuse & neglect*, 22(2):103–115, 1998.

[11] Hala Assal and Sonia Chiasson. Security in the software development lifecycle. In *Proceedings of the Fourteenth USENIX Conference on Usable Privacy and Security*, pages 281–296, 2018.

[12] Hala Assal and Sonia Chiasson. 'think secure from the beginning': A survey with software developers. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA, 2019. Association for Computing Machinery.

[13] J.D. Aurini, M. Heath, and S. Howells. *The How To of Qualitative Research*. SAGE Publications, 2021.

[14] Angela H. Beckett, Kate F. Cook, and Samuel C. Robson. A pandemic in the age of next-generation sequencing. *The Biochemist*, 43(6):10–15, 2021.

[15] Tamara Bonaci, Jeffrey Herron, Tariq Yusuf, Junjie Yan, Tadayoshi Kohno, and Howard Jay Chizeck. To make a robot secure: An experimental analysis of cyber security threats against teleoperated surgical robots, 2015.

[16] Matteo Cagnazzo, Markus Hertlein, Thorsten Holz, and Norbert Pohlmann. Threat modeling for mobile health systems. In *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 314–319, 2018.

[17] Center for Devices and Radiological Health, Food & Drug Administration. Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions, 2023.

[18] Kathy Charmaz. *Constructing grounded theory: A practical guide through qualitative analysis*. sage, 2006.

[19] Christian Johner. Threat Modeling – An Introduction. *Johner-Institute*, 2023.

[20] CISA. St. Jude Merlin@home Transmitter Vulnerability, 2017.

[21] CISA, NSA, FBI, ACSC, NCSC-UK, CCCS, BSI, NCSC-NL, CERT NZ, and NCSC-NZ. Shifting the Balance of Cybersecurity Risk: Principles and Approaches for Security-by-Design and -Default. Technical report, 2023.

[22] Claudio Cobelli, Eric Renard, and Boris Kovatchev. Artificial pancreas: past, present, future. *Diabetes*, 60(11):2672–2682, 2011.

[23] Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70:213–220, 1968.

[24] Juliet Corbin and Anselm Strauss. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications, 2014.

[25] DefCon. Biohacking Village, 2022.

[26] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16(1):3–32, 2011.

[27] Tamara Denning, Adam Lerner, Adam Shostack, and Tadayoshi Kohno. Control-alt-hack: the design and evaluation of a card game for computer security awareness and education. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 915–928, 2013.

[28] Michael Eddington, Brenda Larcom, and Eleanor Saitta. Trike.

[29] Erik C. Decker and Julie Chua. Health industry cybersecurity practices: Managing threats and protecting patients. Technical report, HSCC, 2019.

[30] Elaine Bochniewicz et al. Playbook for Threat Modeling Medical Devices, 2021.

[31] Center for Devices and Food & Drug Administration Radiological Health. Discussion Paper: Strengthening Cybersecurity Practices Associated with Servicing of Medical Devices: Challenges and Opportunities. *FDA*, 2021.

[32] Center for Devices and Food & Drug Administration Radiological Health. The Artificial Pancreas Device System, 2022.

[33] International Organization for Standardization. Iso/iec/ieee 24765: 2017 systems and software engineering—vocabulary, 2017.

[34] International Organization for Standardization. Health software and health it systems safety, effectiveness and security — part 5-1: Security — activities in the product life cycle IEC 81001-5-1:2021, 2021.

[35] International Organization for Standardization. Health informatics — device interoperability — part 40101: Foundational — cybersecurity — processes for vulnerability assessment ISO/IEEE 11073-40101:2022, 2022.

[36] Forrest Shull. Evaluation of Threat Modeling Methodologies. In *SEI 2016 Research Review*, 2016.

[37] OWASP Foundation. OWASP Threat Dragon, 2020.

[38] Sylvain Frey, Awais Rashid, Pauline Anthonysamy, Maria Pinto-Albuquerque, and Syad Asad Naqvi. The good, the bad and the ugly: A study of security decisions in a cyber-physical systems game. In *2018 IEEE/ACM 40th International Conference on Software Engineering (ICSE)*, pages 496–496, 2018.

[39] Kelsey R. Fulton, Anna Chan, Daniel Votipka, Michael Hicks, and Michelle L. Mazurek. Benefits and Drawbacks of Adopting a Secure Programming Language: Rust as a Case Study. pages 597–616, 2021.

[40] Kelsey R. Fulton, Daniel Votipka, Desiree Abrokwa, Michelle L. Mazurek, Michael Hicks, and James Parker. Understanding the how and the why: Exploring secure development practices through a course competition. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*, CCS '22, page 1141–1155, New York, NY, USA, 2022. Association for Computing Machinery.

[41] William J. Gordon, Naruhiko Ikoma, Heather Lyu, Gretchen Purcell Jackson, and Adam Landman. Protecting procedural care—cybersecurity considerations for robotic surgery. *npj Digital Medicine*, 5(1):148, 2022.

[42] Chauncey F Graetzel, Alexander Sheehy, and David P Noonan. Robotic bronchoscopy drive mode of the auris monarch platform. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 3895–3901. IEEE, 2019.

[43] Greg Guest, Arwen Bunce, and Laura Johnson. How many interviews are enough? an experiment with data saturation and variability. *Field Methods*, 18(1):59–82, 2006.

[44] Daniel Halperin, Thomas S Heydt-Benjamin, Kevin Fu, Tadayoshi Kohno, and William H Maisel. Security and privacy for implantable medical devices. *IEEE pervasive computing*, 7(1):30–39, 2008.

[45] Daniel Halperin, Thomas S. Heydt-Benjamin, Benjamin Ransford, Shane S. Clark, Benessa Defend, Will Morgan, Kevin Fu, Tadayoshi Kohno, and William H. Maisel. Pacemakers and implantable cardiac defibrillators: Software radio attacks and zero-power defenses. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 129–142, 2008.

[46] Julie M Haney, Mary Theofanos, Yasemin Acar, and Sandra Spickard Prettyman. " we make it a big deal in the company": Security mindsets in organizations that develop cryptographic products. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 357–373, 2018.

[47] Health Canada. Guidance Document: Pre-market Requirements for Medical Device Cybersecurity - Summary, 2021.

[48] Illumina. MiSeq Dx IVDR EU Declaration of Conformity, 2022.

[49] International Medical Device Regulators Forum. Principles and Practices for Medical Device Cybersecurity, 2020.

[50] Irius Risk. Threat modeling medical devices, 2023.

[51] IriusRisk. Threat Modeling Platform.

[52] Pedro Isaías, Sara Pífano, and Paula Miranda. Subject recommended samples: snowball sampling. In *Information systems research and exploring social artifacts: Approaches and methodologies*, pages 43–57. IGI Global, 2013.

[53] John Giantsidis. Medical Device QMS Cybersecurity: Threat Modeling. *Med Device Online*, 2023.

[54] Pontus Johnson, Robert Lagerström, and Mathias Ekstedt. A meta language for threat modeling and attack simulations. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, pages 1–8, 2018.

[55] Harjot Kaur, Sabrina Amft, Daniel Votipka, Yasemin Acar, and Sascha Fahl. Where to Recruit for Security Development Studies: Comparing Six Software Developer Samples. pages 4041–4058, 2022.

[56] Shaymaa Mamdouh Khalil, Hayretdin Bahsi, Henry Ochieng' Dola, Tarmo Korõtko, Kieran McLaughlin, and Vahur Kotkas. Threat modeling of cyber-physical systems - a case study of a microgrid system. *Computers & Security*, 124:102950, 2023.

[57] Rafiullah Khan, Kieran McLaughlin, David Laverty, and Sakir Sezer. Stride-based threat modeling for cyber-physical systems. In *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, pages 1–6, 2017.

[58] Loren Kohnfelder and Praerit Garg. The threats to our products. *Microsoft Interface, Microsoft Corporation*, 1999.

[59] Larry Mraz and David Remick. Medical device cyber security & privacy. Technical report, KPMG, 2017.

[60] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research Methods in Human Computer Interaction (Second Edition)*. Morgan Kaufmann, Boston, second edition edition, 2017.

[61] JGraph Ltd. draw.io.

[62] Patrick Luckett, Jeffrey McDonald, and William Glisson. Attack-graph threat modeling assessment of ambulatory medical devices. In *Proceedings of the 50th Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences, 2017.

[63] Jacques Marescaux, Joel Leroy, Francesco Rubino, Michelle Smith, Michel Vix, Michele Simone, and Didier Mutter. Transcontinental robot-assisted remote telesurgery: feasibility and potential applications. *Annals of surgery*, 235(4):487–492, 2002.

[64] Goncalo Martins, Sajal Bhatia, Xenofon Koutsoukos, Keith Stouffer, CheeYee Tang, and Richard Candell. Technical Report: Towards a Systematic Threat Modeling Approach for Cyber-physical Systems. *Proceedings, 2015 Resilience Week (RSW) : Hilton Philadelphia at Penn's Landing, Philadelphia, PA, 18-20 August, 2015. Resilience Week (2015 : Philadelphia, Pa.)*, 2015, 2015.

[65] MedCrypt. Threat Modeling Training for Medical Devices, 2022.

[66] Medical Device Coordination Group. MDCG 2019-16 Rev.1 Guidance on Cybersecurity for medical devices. Technical report, 2020.

[67] Medical Device Cybersecurity Working Group. Principles and Practices for the Cybersecurity of Legacy Medical Devices. Technical report, International Medical Device Regulators Forum, 2022.

[68] Medical Device Innovation Consortium. 2023 Threat Modeling Bootcamps, 2023.

[69] Medtronic. MiniMed™ 770G Hybrid Closed Loop System.

[70] Medtronic Diabetes. URGENT MEDICAL DEVICE RECALL MiniMed™ Remote Controller (MMT-500 or MMT-503), 2021.

[71] Microsoft. Microsoft Threat Modeling Tool overview, 2022.

[72] Rahul Mohanani, Iflaah Salman, Burak Turhan, Pilar Rodríguez, and Paul Ralph. Cognitive biases in software engineering: a systematic mapping study. *IEEE Transactions on Software Engineering*, 46(12):1318–1339, 2018.

[73] Stefania Morganti, Paolo Tarantino, Emanuela Ferraro, Paolo D'Amico, Bruno Achutti Duso, and Giuseppe Curigliano. Next Generation Sequencing (NGS): A Revolutionary Technology in Pharmacogenomics and Personalized Medicine in Cancer. *Advances in experimental medicine and biology*, 1168:9–30, 2019.

[74] Alena Naiakshina, Anastasia Danilova, Eva Gerlitz, and Matthew Smith. On conducting security developer studies with cs students: Examining a password-storage study with cs students, freelancers, and company developers. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2020.

[75] Lily Hay Newman. Hackers Could Increase Medication Doses Through Infusion Pump Flaws. *Wired*, 2021.

[76] Peter Ney, Karl Koscher, Lee Organick, Luis Ceze, and Tadayoshi Kohno. Computer security, privacy, and {DNA} sequencing: compromising computers with synthesized {DNA}, privacy leaks, and more. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 765–779, 2017.

[77] Bernard A. Nijstad, Carsten K. W. De Dreu, Eric F. Rietzschel, and Matthijs Baas. The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence. *European Review of Social Psychology*, 21(1):34–77, 2010.

[78] Daniela Seabra Oliveira, Tian Lin, Muhammad Sajidur Rahman, Rad Akefirad, Donovan Ellis, Eliany Perez, Rahul Bobhate, Lois A. DeLong, Justin Cappos, and Yuriy Brun. API blindspots: Why experienced developers write vulnerable code. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*, pages 315–328, Baltimore, MD, 2018. USENIX Association.

[79] Carly Page. Critical-rated security flaw in Illumina DNA sequencing tech exposes patient data. *TechCrunch*, 2023.

[80] Hernan Palombo, Armin Ziaie Tabari, Daniel Lende, Jay Ligatti, and Xinming Ou. An ethnographic understanding of software (in) security and a co-creation model to improve secure software development. In *Proceedings of the Sixteenth USENIX Conference on Usable Privacy and Security*, SOUPS'20, USA, 2020. USENIX Association.

[81] Nathanael Paul, Tadayoshi Kohno, and David C Klonoff. A review of the security of insulin pump infusion systems. *Journal of diabetes science and technology*, 5(6):1557–1562, 2011.

[82] Kevin Poulsen, Robert McMillan, and Melanie Evans. A Hospital Hit by Hackers, a Baby in Distress: The Case of the First Alleged Ransomware Death. *Wall Street Journal*, 2021.

[83] Sam Procter, Eugene Y Vasserman, and John Hatcliff. Safe and secure: Deeply integrating security in a new hazard analysis. In *Proceedings of the 12th International Conference on Availability, Reliability and Security*, pages 1–10, 2017.

[84] Qiagen. Quality Assurance, 2023.

[85] Harry T Reis, Harry T Reis, Charles M Judd, et al. *Handbook of research methods in social and personality psychology*. Cambridge University Press, 2000.

[86] Rev.com. Transcribe Speech to Text.

[87] Andrew Ruef, Michael Hicks, James Parker, Dave Levin, Michelle L Mazurek, and Piotr Mardziel. Build it, break it, fix it: Contesting secure development. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, pages 690–703, 2016.

[88] Bruce Schneier. Attack trees. *Dr. Dobb's journal*, 24(12):21–29, 1999.

[89] Security Innovation. Medical Device Threat Modeling, 2023.

[90] Senator Mark R. Warner. Cybersecurity is Patient Safety. Technical report, Senate Select Committee on Intelligence, 2022.

[91] Nataliya Shevchenko. Threat modeling: 12 available methods. Carnegie Mellon University, Software Engineering Institute's Insights (blog), 2018. Accessed: 2023-Mar-21.

[92] Nataliya Shevchenko, Brent R. Frye, and Carol Woody. Threat Modeling for Cyber-Physical System-of-Systems: Methods Evaluation. Technical report, Carnegie Mellon University Software Engineering Institute, 2018.

[93] Zhenpeng Shi, Kalman Graffi, David Starobinski, and Nikolay Matyunin. Threat modeling tools: A taxonomy. *IEEE Security & Privacy*, 20(4):29–39, 2022.

[94] Ben Shneiderman and Catherine Plaisant. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson, 4th edition, 2016.

[95] Ben Shneiderman, Catherine Plaisant, Maxine S Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. *Designing the user interface: strategies for effective human-computer interaction*. Pearson, 2016.

[96] Adam Shostack. *Threat modeling: Designing for security*. John Wiley & Sons, 2014.

[97] 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*, 48(05):1515–1528, 2022.

[98] Benjamin Shreeve, Joseph Hallett, Matthew Edwards, Pauline Anthonysamy, Sylvain Frey, and Awais Rashid. "So If Mr Blue Head Here Clicks the Link..." Risk Thinking in Cyber Security Decision Making. *ACM Trans. Priv. Secur.*, 24(1), 2020.

[99] Rock Stevens, Daniel Votipka, Elissa M. Redmiles, Colin Ahern, Patrick Sweeney, and Michelle L. Mazurek. The battle for new york: A case study of applied digital threat modeling at the enterprise level. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 621–637, Baltimore, MD, 2018. USENIX Association.

[100] CyberMed Summit. CyberMed DC, 2022.

[101] Microsoft Software Development Lifecycle Team. The stride per element chart, 2007.

[102] The White House. National Cybersecurity Strategy, 2023.

[103] Therapeutic Goods Administration. Medical device cyber security guidance for industry. Technical Report Version 1.2, Department of Health and Aged Care, Australian Government, 2022.

[104] Thermo Fisher Scientific. IVDR Certificates, 2023.

[105] ThreatModeler. ThreatModeler^R.

[106] threatmodeler. What is VAST?, 2018.

[107] Shawn Tsuda, Dmitry Oleynikov, Jon Gould, Dan Azagury, Bryan Sandler, Matthew Hutter, Sharona Ross, Eric Haas, Fred Brody, and Richard Satava. Sages tavac safety and effectiveness analysis: da vinci® surgical system (intuitive surgical, sunnyvale, ca). *Surgical endoscopy*, 29:2873–2884, 2015.

[108] Katja Tuma, Riccardo Scandariato, Mathias Widman, and Christian Sandberg. Towards security threats that matter. In Sokratis K. Katsikas, Frédéric Cuppens, Nora Cuppens, Costas Lambrinoudakis, Christos Kalloniatis, John Mylopoulos, Annie Antón, and Stefanos Gritzalis, editors, *Computer Security*, pages 47–62, Cham, 2018. Springer International Publishing.

[109] Tony UcedaVelez and Marco M Morana. *Risk Centric Threat Modeling*. John Wiley & Sons, 2015.

[110] Dimitri Van Landuyt and Wouter Joosen. A descriptive study of assumptions in STRIDE security threat modeling. *Software and Systems Modeling*, 21(6):2311–2328, 2022.

[111] Daniel Votipka, Kelsey R Fulton, James Parker, Matthew Hou, Michelle L Mazurek, and Michael Hicks. Understanding security mistakes developers make: Qualitative analysis from build it, break it, fix it. In *29th USENIX Security Symposium (USENIX Security 20)*, pages 109–126, 2020.

[112] Daniel Votipka, Rock Stevens, Elissa Redmiles, Jeremy Hu, and Michelle Mazurek. Hackers vs. testers: A comparison of software vulnerability discovery processes. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 374–391. IEEE, 2018.

[113] Dominik Wermke, Noah Wöhler, Jan H. Klemmer, Marcel Fourné, Yasemin Acar, and Sascha Fahl. Committed to trust: A qualitative study on security trust in open source software projects. In *2022 IEEE Symposium on Security and Privacy (SP)*, pages 1880–1896, 2022.

[114] Emily Skahill and Darrell M. West. Why hospitals and healthcare organizations need to take cybersecurity more seriously. *Brookings*, 2021.

[115] Nicole Wetsman. The pandemic revealed the health risks of hospital ransomware attacks. *The Verge*, 2021.

[116] Wenjun Xiong and Robert Lagerström. Threat modeling – a systematic literature review. *Computers & Security*, 84:53–69, 2019.

[117] William Young and Nancy Leveson. Systems thinking for safety and security. In *Proceedings of the 29th Annual Computer Security Applications Conference*, ACSAC '13, page 1–8, New York, NY, USA, 2013. Association for Computing Machinery.

[118] Zoe Braiterman, Adam Shostack, Jonathan Marcil, Stephen de Vries, Irene Michlin, Kim Wuyts, Robert Hurlbut, Brook S.E. Schoenfield, Fraser Scott, Matthew Coles, Chris Romeo, Alyssa Miller, Izar Tarandach, Avi Douglen, and Marc French. Threat Modeling Manifesto.

A Interview Script

We include the specific questions asked of participants in the interview. Before we asked these we reviewed the consent documents and explained the structure of the interview.

A.1 Background

Great, so let us now talk a little bit about your background and how you became involved in medical device security.

- Can you briefly go over the different jobs that you have had during your career?
- So what got you into medical devices?
- What experience do you have with security?

A.2 Scenarios

We will look at two specific device development scenarios. We will present you with a set of requirements for a device and we want you to imagine you're working to design and build that device. As part of this we'll give you a system diagram that will lay out the basic functionality of the device, as well as any specific features we would like you to include. Our goal in each of these scenarios is to understand your process better, specifically how you formulate ideas on security and safety issues and conceptualize these threats. Because we're interested in the process itself, we would love for you to talk through your thinking at each step to the extent you are comfortable. Also if at any point something comes to mind about a previous scenario when we are on a different one, we can always go back and discuss that more. You are also welcome to ask for us to repeat any information about the requirements or for more details at any point. Note, we do not have a final solution in mind. There is no "correct" solution. We just want to see how you think about the problem.

Questions to be used for each scenario.

Let's discuss how you would approach ensuring this device is secure & safe. What are some of the initial thoughts or concerns you would have?

- What are some of the security concerns you have? How did you come up with them?
- What are some of the safety concerns you have? How did you come up with them?
- What are some of the mitigation strategies you would use?
- How do you think about these threats? Did you come up with this on your own, from a previous experience, or something you read/saw?
- What sort of fail safe systems would you put in place if the device or the software breaks? Who would be in charge of implementing and maintaining these systems?
- How would you go about shipping software updates? Who is responsible for dealing with software updates?
- What kind of security training are end users provided with?
- Would the device connect via Bluetooth or another network protocol? If another network protocol, would it be a dedicated network or regular WiFi?

Scenario 1: Implantable Medical Device. You are in the process of developing a networked insulin pump that can monitor blood glucose and allow patients to control insulin levels via their smartphone. The goal of the device is accurate and automatic insulin delivery, as well as easy monitoring for patients, caregivers, and medical providers.

This insulin pump will connect to the smartphone using bluetooth. The app on the phone will allow the patient, as well as their medical provider and any trusted individuals/caregivers, access and control over the device. The device's default will be an "automatic mode" that automatically pumps insulin based on the patient's current levels every few minutes, and the device also offers an "exercise mode" where the patient can opt to adjust the glucose target in order to avoid high and low spikes in insulin levels.

Scenario 2: Surgical Device. You are developing a robotic surgical system that is going to be used in Operating Rooms. This device should be able to perform general surgery procedures such as an appendectomy or bariatric surgery, as well as cardiac and orthopedics procedures. The system will consist of a surgeon console and a patient cart that holds the instruments as well as a way to show other providers the surgeon's viewpoint.

The system must have low latency in order to make sure there's minimal delay between the surgeon's movements and the instruments corresponding movements. Additionally, the view that the surgeon sees must be able to be broadcasted on a network so that others in the OR as well as possibly remote providers are able to watch the surgery. This must also have low latency. The hardware requirements are up to you, and we'll only discuss these as it relates to security rather than efficacy.

Scenario 3: Next-Gen Sequencer. You are working on a new Next-Generation Sequencer. Your company is developing both the physical device as well as software to analyze the data. The goal of this product is to be a one stop shop that allows pre-processing of samples as well as the breaking of samples into reads before sequencing as well as provide initial analytics to researchers.

The sequencer should be capable of high throughput and would be used for targeted sequencing, and whole gene sequencing, a machine that would be able to handle almost all sequencing needs. The device would also need to be able to assist in the preparation of samples.