

Knowledge Modelling for Automated Risk Assessment of Cybersecurity and Indirect Patient Harms in Medical Contexts

Samuel M. Senior¹^a, Laura Carmichael¹^b, Steve Taylor¹^c, Mike Surridge¹^d, and Xavier Vilalta²

¹IT Innovation Centre, University of Southampton, Southampton, U.K.

²Debiotech SA, Lausanne, Switzerland

{s.senior, l.e.carmichael, s.j.taylor, ms8}@soton.ac.uk, x.vilalta@debiotech.com

Keywords: Automated Risk Assessment, Connected Medical Devices and In Vitro Diagnostic Devices, Cybersecurity, Indirect Patient Harms, Knowledge Modelling

Abstract: The use of connected medical and in vitro diagnostic devices (CMD&IVD) as part of individual care and self-care practices is growing. Significant attention is needed to ensure that CMD&IVD remain safe and secure throughout their lifecycles — as if a cybersecurity incident were to occur involving these devices, it is possible that in some cases harm may be brought to the person using them. For the effective safety management of these devices, risk assessment is needed that covers both the cybersecurity and patient safety domains. To this end, we present knowledge modelling of indirect patient harms (e.g., misdiagnosis, delayed treatment, etc.) resulting from cybersecurity compromises, along with a methodology for encoding these into a previously developed automated cybersecurity risk assessment tool, to begin to bridge the gap between automated risk assessment related to cybersecurity and patient safety.

1 INTRODUCTION

It is increasingly common for connected medical devices and in vitro diagnostic devices (CMD&IVD) to be used as part of individual care and self-care practices — e.g., for remote monitoring by clinicians, for individuals to manage their conditions through health apps (e.g., National Health Service [NHS] England, 2023). Special attention is needed to ensure that CMD&IVD remain safe and secure throughout their lifecycles — especially given that CMD&IVD cyberattacks “*may put at severe risk the health and safety of patients*” (Biasin & Kamenjasevic, 2022).

To understand potential harms to patients from such devices, **risk management** is necessitated, particularly at patient safety, cybersecurity, and privacy and data protection levels, so that the risks of cybersecurity incidents can be understood in terms of the potential patient harms that may result. In other words, there is a need for cybersecurity risk assessment for CMD&IVD to “*explicitly consider the health care outcomes, systems and processes for*

which that information is used” (Piggin, 2017). Additionally, risk-benefit analysis is also of importance as tensions between the level of cybersecurity controls on a device and its treatment or diagnostic effectiveness may need to be considered. As highlighted in the Medical Device Co-ordination Group (MDCG, 2019) guidance on cybersecurity where issues may be caused by “weak security” — referring to security measures that are inadequate in the given circumstances — and “restrictive security” — relating to those security measures that offer “a high level of protection may have a safety impact”.

This paper describes knowledge extensions to an existing knowledge-based expert system and automated risk simulator of cyber-physical systems called Spyderisk (Phillips et al., 2024) that follows **ISO/IEC 27005: 2022** “Information security, cybersecurity and privacy — Guidance on managing information security risks” (ISO, 2022) and **ISO/IEC 27000: 2018** “Information technology — Security

^a <https://orcid.org/0000-0002-3428-9215>

^b <https://orcid.org/0000-0001-9391-1310>

^c <https://orcid.org/0000-0002-9937-1762>

^d <https://orcid.org/0000-0003-1485-7024>

techniques — Information security management systems — Overview and vocabulary” (ISO, 2018).

The Knowledge Base of this expert system contains pre-existing information about threats and risks related to cyber-physical systems. As part of our recent work for the Horizon Europe NEMECYS project, we have begun to explore how this Knowledge Base can be extended for use in the specific domain of cybersecurity for CMD&IVD systems so that automated risk assessment can be performed for it. For example, **ISO/TR 24971: 2020** (ISO, 2020) provides guidance on risk assessment for medical devices and so starts to bridge this gap between the relationship between domain-specific risk management for medical devices with risk management for cybersecurity.

The core extension to Spyderisk described in this paper is domain model extensions corresponding to **indirect patient harms resulting from cybersecurity compromises**. For the purposes of this paper, indirect patient harms are described as harms that arise *“as a consequence of the medical decision or action taken/not taken on the basis of information or result(s) provided by a device”* (MDCG, 2023). Indirect patient harms resulting from cybersecurity compromises are translated into the terminology and structure of this Knowledge Base, and mapped to cybersecurity risks and threats already present within it. Then, this new information and mapping is encoded into the Knowledge Base, thus starting to bridge and link between the domains of cybersecurity and indirect patient harms for the automated risk assessment and management of CMD&IVD. An illustrative workflow guiding the extensions to our tool for treatment-based indirect patient harms is given. An equivalent workflow for diagnosis-based indirect patient harms can also be formed through following the same methodology, and both have been successfully implemented in our tool (Spyderisk, 2024). Our work is driven by four use cases, focusing on different types of connected medical and IVD devices (NEMECYS, 2023).

2 BACKGROUND AND RISK MANAGEMENT CONTEXT

ISO 27000 (ISO, 2022) and ISO 27005 (ISO, 2018) have guided the development of the Spyderisk risk modelling approach and continue to do so for the extensions outlined here. In this section, we provide an overview of how cybersecurity risk concepts from these two standards have been interpreted for the

Spyderisk. We then outline our approach to extending its Knowledge Base with domain-specific information for the cybersecurity of CMD&IVD.

2.1 Risk Assessment Schema

Figure 1 shows a risk assessment schema derived from ISO 27000 (ISO, 2018) and adapted for the trustworthiness-based approach of Spyderisk. It shows relationships between the different elements involved in ISO 27000-based risk assessment.

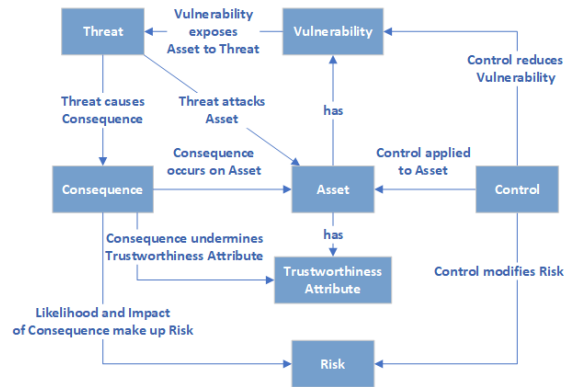


Figure 1: Risk assessment schema. Adapted from (Taylor, 2024).

Here, **assets** are entities of value within the modelled system and can have **vulnerabilities**, which enable **threats**. A successful threat acts on an asset to cause a **consequence**, which is typically adverse. **Risk** is the impact of a consequence combined with the likelihood of the causing threat. **Controls** modify risk by modifying the likelihood of the threat through mitigative or preventative means. **Trustworthiness Attributes (TWAs)** “model the expected behaviour of an Asset, are (generally) desirable properties and are closely related to the Consequences: each Trustworthiness Attribute is undermined by a Consequence.” (Phillips et al., 2024).

Spyderisk contains a **Knowledge Base of Assets, Consequences, TWAs, Controls and Threats** for cybersecurity and cyber-physical systems. The Asset types include data, IT hosts, software processes, networks, stakeholders, and physical spaces, amongst others. TWAs include cybersecurity concepts like confidentiality, integrity, and availability for data assets, reliability for software processes and privacy for humans. Consequences include (typically) adverse behaviours affecting assets and also the undermining of TWAs at assets, such as “loss of confidentiality” on a data asset. Consequences in turn can cause other threats.

The operator of the Spyderisk creates a “System Model” containing a configuration of assets and relationships describing the system to evaluate, and the Knowledge Base automatically determines the threats and risks present, along with their likelihood.

2.2 Domain Modelling

The knowledge extensions to Spyderisk follow a process known as **domain modelling**, which involves the capture and encoding of knowledge relevant to risk assessment for a given domain to integrate it with the existing knowledge and thus to extend it. Here, an essential part of domain modelling is to acquire knowledge relevant to the cybersecurity of CMD&IVD — such as, by working together with cross-domain experts as part of project use cases and examining existing requirements and best practice related to risk assessment for CMD&IVD with principal focus on the EU regulatory framework. For instance, Annexes 1 of the **Medical Device Regulation (MDR)** (Regulation 2017/745) and the **In Vitro Diagnostic Devices Regulation (IVDR)** (Regulation 2017/746) contain cybersecurity requirements for CMD&IVD, and the **MDCG** provides guidance on the cybersecurity for medical devices (2019). Further, international standards on risk management for medical devices are also used to guide the developments — i.e., **ISO 14971:2019** “Medical devices — Application of risk management to medical devices” (ISO, 2019) “*specifies terminology, principles and a process for risk management of medical devices, including software as a medical device and in vitro diagnostic medical devices*” (ISO, 2019); and **ISO 24971** (ISO, 2020) as previously mentioned. A brief overview of the domain modelling process is now described.

In the Knowledge Base, threats are modelled using: (i) **Matching Patterns** and (ii) **Threat Patterns**. “*Broadly, a Matching Pattern describes a set of connected Assets to be looked for in the System Model: particular Asset types connected by specific Relation types*” (Phillips et al., 2024). “*Threat Patterns are matching parts of the System Model where there could be an unwanted incident (of any kind) [...] [and] describes how the Nodes in the pattern relate to its causes*” (Phillips et al., 2024).

An example threat pattern is given in Figure 2. It is based on a matching pattern with an additional cause TWA that enables the threat (blue rounded rectangle), two controls that block the threat (green ovals), and a consequence that results from the threat being successful (red oval).

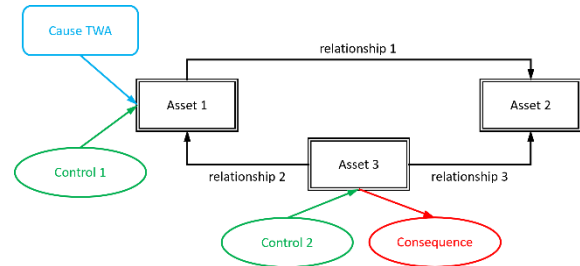


Figure 2: Example threat pattern.

3 MODELLING INDIRECT PATIENT HARMS AS RISKS

Indirect patient harms resulting from cybersecurity compromises of CMD&IVD can occur due to one or more cybersecurity-related incidents causing errors in diagnosis and / or treatment. This work utilises types of indirect harm determined by the MDCG, namely “*absence of diagnosis*”, “*delayed diagnosis*”, “*misdiagnosis*”, “*absence of treatment*”, “*delayed treatment*”, and “*inappropriate treatment*” (MDCG, 2023). These are modelled as **Consequences of Threats** that are adverse behaviours affecting patients (considered as “Assets” in ISO 27000 nomenclature).

Typical treatment and diagnosis processes utilising CMD&IVD have been used to derive patient harm-related Consequences and how different cybersecurity incidents lead to them. For this, four key types of purpose for CMD&IVD devices are outlined and explored. Additionally, a generic workflow based on these is presented, which provides a high-level illustrative view of how sensor data generated via these devices, and the resulting examination results they provide, are used to inform treatment decisions actions taken / not taken. This is then used to map elements of the workflow phases to elements of the Knowledge Base so that cybersecurity threats at the different stages can be considered in how they result in indirect patient harms.

The domain modelling activities here have been driven by the discussed standards as well as use cases involving consultation with domain experts and practitioners, with one such use case used as an illustrative example in Section 6.

3.1 Purpose of Use of Medical Devices

A wide variety of decisions are made by individuals, patients, and clinicians (decision-makers) as part of individual care and self-care — some of which may be informed by data generated and collected via

assorted types of CMD&IVD. A CMD or IVD device will have an “intended purpose” — i.e., *“the use for which a device is intended according to the data supplied by the manufacturer on the label, in the instructions for use or in promotional or sales materials or statements and as specified by the manufacturer in the clinical evaluation”*, as defined in Article 2(12) of the MDR (Regulation 2017/745). Types of “specific medical purposes” for medical devices are outlined by Article 2(1) of the MDR as follows:

- “[...] diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease,
- diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability,
- investigation, replacement or modification of the anatomy or of a physiological or pathological process or state,
- providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations [...]”

In this work, the focus is on the specific purposes for medical devices that align with the key types of intended purposes for CMD&IVD as identified through the use cases. In particular, how medical devices may be used for the following purposes:

- **Diagnosis** — some MDCG-based indirect harms (MDCG, 2023) explicitly concern diagnosis (e.g., ‘delayed diagnosis’, ‘misdiagnosis’).
- **Treatment** — some MDCG-based indirect harms (MDCG, 2023) regard treatment (e.g., ‘absence of treatment’, ‘delayed treatment’, ‘inappropriate treatment’).
- **Monitoring** — CMD&IVD can be used for the purpose of monitoring.
- **Providing information** by means of in vitro examination of specimens derived from the human body — as there are instances of CMD&IVD being used for the purpose of providing such information.

How these four key types of purpose are modelled is given in Section 4.

3.2 Generic Workflow: Monitoring and Treatment for Non-Emergency Individual Care and Self-Care

Diagnosis, monitoring, treatment and providing IVD information are complex activities. For instance, diagnosis has been described as *“a complex, patient-centered, collaborative activity that involves information gathering and clinical reasoning with the goal of determining a patient's health problem. This process occurs over time, within the context of a larger health care work system that influences the diagnostic process”* (Balogh et. al., 2015). The aim here is to identify key aspects and abstractions that are required for understanding the risks in these processes to enable their risk modelling and simulation. This is important as medical devices may be used as part of distinct clinical workflows and at different stages of a care pathway. For instance, what data is being collected, gathered or generated by the CMD&IVD, and for what use, needs to be considered. Each stage of the workflow represents data, a process, or a consequence — concepts used within our Knowledge Base.

The generic workflow is presented in Figure 3 and focuses on CMD&IVD used for the purpose of monitoring and treatment as part of individual care (e.g., *“intended for use by clinicians at point-of-care”* (ISO, 2020) or self-care (e.g., intended to be used by individuals). It should be noted the IVDR makes a distinction between IVD medical devices used for *“self-testing”*, *“near-patient testing”*, and testing inside a *“laboratory environment”*, see Article 2(5) and (6) of the IVDR (Regulation 2017/746).

The workflow is organised into phases containing processes, which are operations performing an activity; data, which are generated by the processes and link one process with another; and consequences, which result from incidents occurring at the data and processes. It illustrates how different types of sensor data are generated, interpreted, and acted on for the *monitoring and treatment* as part of non-emergency individual care and self-care practices when those decision-making processes rely on high quality examination results derived from data generated and collected via CMD&IVD.

This workflow is based on the specific medical purposes for medical devices as well as our interpretation of the informative guidance for in vitro diagnostic medical devices given by Annex H of ISO/TR 24971 (ISO, 2020) and the diagnostic process outlined by the Committee on Diagnostic Error in Health Care (Balogh, 2015).

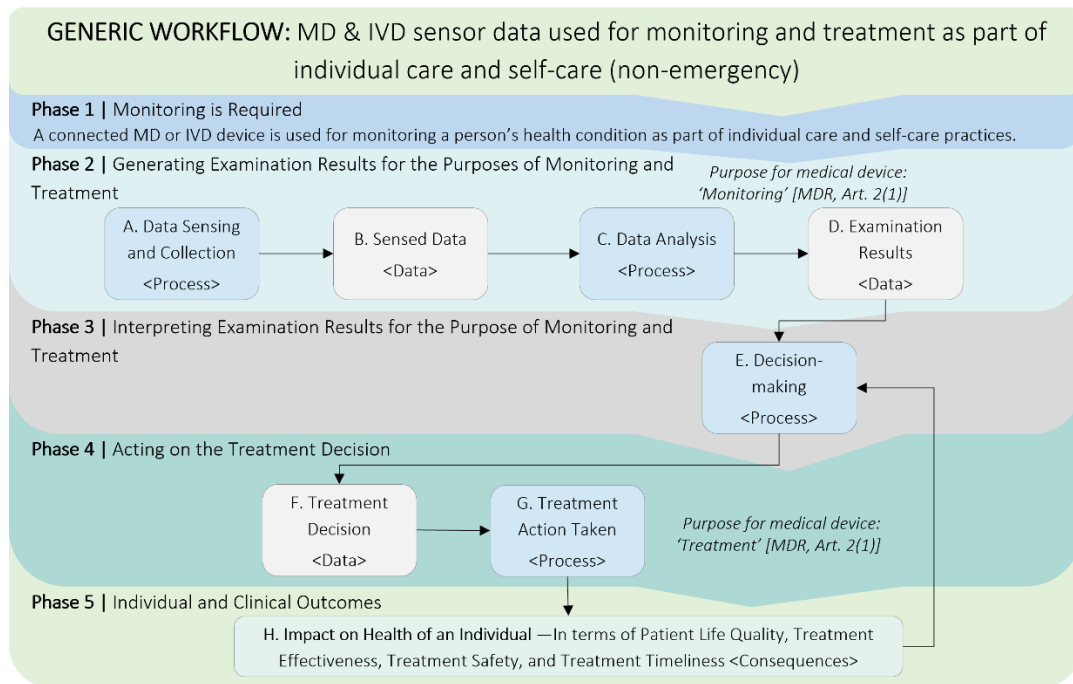


Figure 3: Generic workflow: MD & IVD used for monitoring and treatment in non-emergency individual care and self-care.

The different workflow phases are discussed below. As assumption-making is a key aspect of threat modelling since such underlying assumptions are used to “*postulate system properties of relevance, the implications of which are relied upon during threat documentation, prioritization and mitigation*” (Van Landuyt and Joosen, 2022), key domain modelling assumptions are also described below.

3.3.1 Phase 1: Monitoring is Required

This phase establishes an individual is undergoing care or self-care where CMD&IVD are used to monitor their condition. The workflow is initiated where a CMD&IVD is used for monitoring a person's health condition, that requires ongoing treatment, as part of individual care and/or self-care practices.

It is assumed this workflow concerns situations where an individual has already received a diagnosis. Further, to consider the indirect patient harms rather than direct patient harms, non-emergency situations are specifically concentrated on, where on-going monitoring and treatment of a health condition may be happening e.g., remotely, within a clinical context.

3.3.2 Phase 2: Generating Examination Results

This phase of the workflow uses the sensors of the CMD&IVD to generate the monitoring data that are

then processed to generate the Examination Results Data. This follows three of the key purposes for medical devices given in Article 2(1) of the MDR (Regulation 2017 745), which are for “*monitoring*” and “*treatment*” of “*disease*”, “*an injury or disability*” as well as “*providing information by means of in vitro examination*”.

It is assumed Examination Results Data are derived from Sensed Data and used for monitoring and treatment by clinicians and patients, and is crucial to monitoring a health condition and making necessary treatment decisions. The processes and data for this phase of the workflow are:

- **A. Data Sensing and Collection <Process>**. Data is sensed by an MD or IVD medical device. (Other metadata may also be collected here, such as time-stamp data.)
- **B. Sensed Data <Data>**. The sensed raw data are an input to the data analysis process.
- **C. Data Analysis <Process>**. Raw data is transformed into “*meaningful, actionable knowledge*” (ISO, n.d.) data (i.e., Examination Results).
- **D. Examination Results <Data>**. Output Examination Results provide meaningful, actionable knowledge, interpreted and used as an input to the decision-making process.

Cybersecurity incidents arising in any of the data or processes at this phase impact the decision-making process of the next phase, by either causing incorrect input into it or causing a lack of input to it.

3.3.3 Phase 3: Interpreting Examination Results

In the third phase, Examination Results are processed and interpreted so appropriate treatment decisions can be made by the individual or clinician.

It is assumed Examination Results are not communicated to another medical device that immediately administer treatment (ISO, 2020) and instead, in Phase 4, one or more persons review the Examination Results data before deciding whether to act on it. Additionally, the Examination Results data is viewed as a critical input to the decision-making process. The process for this phase is:

- **E. Decision-Making <Process>.** Examination Results are processed and interpreted to inform patient / clinician treatment decisions.

Cybersecurity incidents arising at this decision-making process cause the output of it to either be incorrect or absent. Additionally, incidents in the prior phase can also result in this.

3.3.4 Phase 4: Acting on Treatment Decision

In Phase 4, treatment decisions are made and actions taken. Treatment actions are taken as part of wider individual care and / or self-care practices and will contribute to realising “*the best possible outcomes for the individual*” (Mukoro, 2011). The process and data for this are:

- **F. Treatment Decision <Data>.** As an output of the decision-making process, a treatment decision is made.
- **G. Treatment Action Taken <Process>.** Treatment decision is acted on by the patient and/or clinician(s) responsible for their care.

For both the data and process of this phase, if there are cybersecurity incidents then the actions taken will either be incorrect or absent, impacting the final phase by causing indirect patient harms.

3.3.5 Phase 5: Individual and Clinical Outcomes

In this final phase, the effectiveness of the treatment actions is evaluated in terms of the impact on the health of the individual. This phase of the workflow

impacts patient harm consequences relating to the outcomes of their treatment actions, described by:

- **H. Impact on the Health of an Individual <Consequences>.** Patient harm consequences related to a lack treatment or lack of necessary treatment. These result from cybersecurity-related incidents in the prior phases.

In summary, patient harm consequences that have been identified. The next section describes how these are caused by cybersecurity threats.

4 MAPPING CYBERSECURITY CAUSES TO INDIRECT PATIENT HARMS

To link cybersecurity threats to the non-emergency indirect patient harms, the workflow is used to make connections with pre-existing information about threats and risks related to cyber-physical systems in the Knowledge Base. The focus is about how the security risk factors related to the generic workflow can be mapped to indirect patient harms. The workflow is concerned with processes, data, and consequences that are indirect patient harms resulting from threats. This process-data-consequence approach can be represented as a threat-consequence state mapping diagram, shown and described below.

4.1 Threat-Consequence State Mapping Diagram

Threat-consequence state mapping diagrams are concerned with the consequences of prior threats leading to further threats and further consequences. One such diagram is given in Figure 3. They contain controls and threats that are connected together to show how chains of these form, leading from one to the next. A **Black Box** is a type of **Threat**, a **Red Oval** a type of **Consequence**, a **Green Oval** a type of **Control**, a **Red Arrow** indicates a type of **Threat enabled by a Control**, and a **Green Arrow** indicates a type of **Threat blocked by a Control**. For the threat-consequence diagram here, the Consequences, Controls and Threats are grouped together between black dashed lines. These groups are based on the Assets at which the Consequences, Controls and Threats occur.

Figure 4 maps security risk factors (including pre-existing cybersecurity knowledge in the Knowledge Base) associated with the generic workflow. The phases of the workflow are highlighted in the diagram

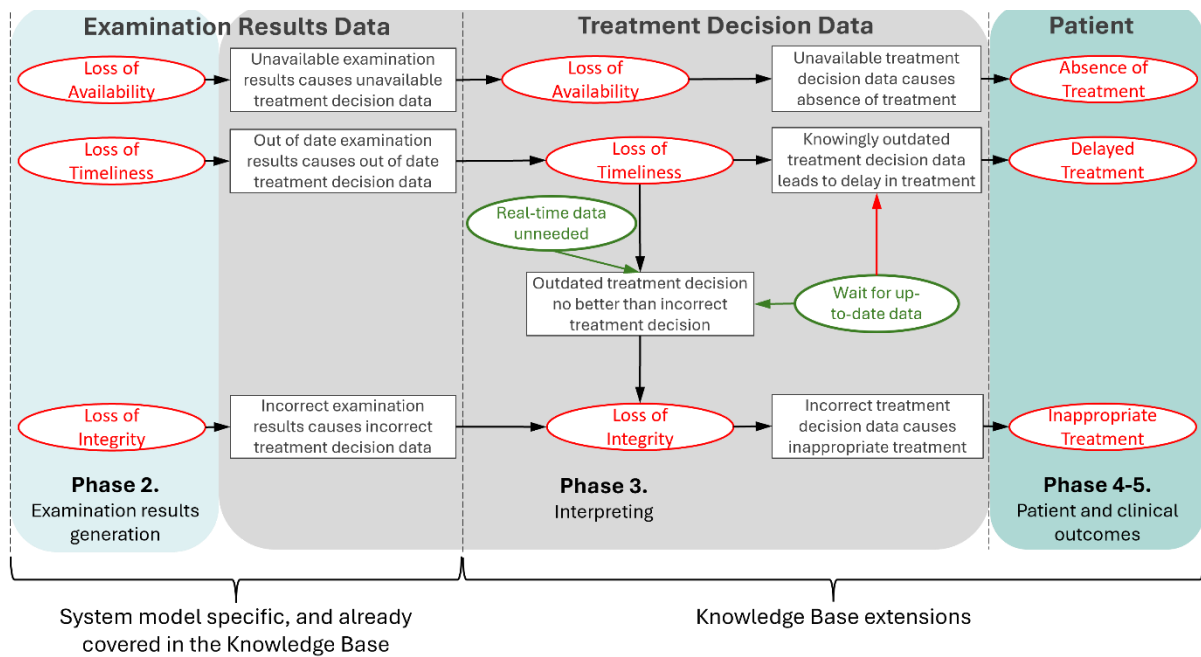


Figure 4: Threat-consequence mapping – Sensor data leading to undermined treatment.

to show which phases of the workflow the different consequences and threats occur in. The processes of the workflow diagram are not explicitly shown, but their actions are implicitly contained within the threats. This then covers the indirect patient harms that occur due to harmful medical decisions for ongoing treatment.

4.2 Risk Modelling for Generic Workflow

In Figure 4, the consequences in Phase 2 act on the Examination Results Data, as an undermining of its TWAs due to prior threats leading into the workflow. The threats in Phase 3 represent flaws occurring in the treatment decision-making process, resulting from cybersecurity-related consequences and lead to patient harm consequences in Phases 4-5 that affect the patient. A key assumption is that the worst-case scenarios are considered (Piggin, 2017).

Following through the diagram (Figure 4):

- **If the Examination Results Data becomes corrupted** (loss of integrity) then through the treatment decision-making processes using this corrupted data the treatment decision data will be corrupted and the

patient will receive an inappropriate treatment.

- **If the Examination Results Data becomes unavailable** (loss of availability) then there is no input to the treatment decision-making process so treatment cannot occur and Treatment Decision Data will be unavailable, leading to an absence of treatment.
- **If the Treatment Decision Data is out of date and near real-time treatment decision data is needed** then the outdated data is no better than being incorrect, leading to a loss of integrity of the Treatment Decision Data. If the Treatment Decision Data does not need to be near real-time then this threat path can be blocked with a control specifying that. Additionally, if the data does need to be up to date then it can be blocked with a control specifying the decision-making process will wait for up to date data. Waiting for the data to be up-to-date again, however, enables a threat resulting in delayed treatment.

With the state mapping diagram defined, the Threats, Controls, Consequences, and paths between them can be encoded into the Knowledge Base. This is discussed in the next section.

5 ENCODING DOMAIN MODELLING EXTENSIONS

Three key types of domain modelling extensions have been undertaken: **(i) a review of existing Asset types** to determine what already exists in the domain model that can be used towards modelling indirect patient harms; **(ii) new relationships** between Assets have been encoded; as have **(iii) new matching patterns**.

5.1 Use of Existing Asset Types

Some of the key assets identified for socio-technical CMD&IVD sensor-based systems already exist in the Knowledge Base and so can be modelled using these existing Asset types. These are the Clinician being modelled as the Adult asset type, the Individual / Patient as a Human, Adult or Child, the CMD&IVD Sensor as an IoT Sensor, the Data Sensing and Collection process, Data Testing process as a Process or Interactive Process, the Decision-making process as an Interactive Process, and the Diagnostic Decision, Examination Results, Treatment Decision and Sensed Data as Health Data asset type.

5.2 New Relationships

To model the identified four key purposes of CMD&IVD, new *Relationship* types between Assets in the Knowledge Base are encoded for diagnosis and treatment. Indicating a type of action from one asset type to another, the new relationship types identified and encoded are:

- **administersTreatment:** Relation between a Human and Data indicating the Human is administering treatment specified by Data.
- **diagnoses:** Relation between two Humans to indicate one Human is a Clinician diagnosing the other, who must be a Patient.
- **diagnosisFor:** Relation to indicate that Health Data relating to a Patient is the diagnosis, as determined by the Clinician.
- **senses:** Relation between a data asset and a sensor, indicating the data is sensed output of the sensor. This was pre-existing but not user-assertable, and this has been changed.
- **treats:** Relation between two Humans indicating one is a Clinician treating the other, who must be a Patient.
- **definesTreatmentFor:** Relation to indicate Health Data relating to a Patient is the

treatment instructions for them, which can be carried out by the Clinician or Patient.

Monitoring is already covered in the Knowledge Base so no new relationship need to be encoded for it.

5.3 New Matching and Threat Patterns

New types of Matching Pattern were encoded so that the following sets of connected Assets and Relationships can be found in modelled systems:

1. A clinician diagnosing a patient, as determined through a *Clinician-diagnoses-Patient* relationship in a wider matching pattern to indicate the clinician creates and interacts with the diagnosis data that is the patient diagnosis.
2. A clinician treating a patient, as determined through a *Clinician-treats-Patient* relationship in a wider matching pattern to indicate the clinician creates and uses the treatment instructions data, which forms the patient treatment actions.
3. A patient treating themselves via self-care practices, as determined through a *Data-definesTreatmentFor-Patient* relationship in a wider matching pattern to indicate the patient is interacting with an interactive application to view and act on the treatment instructions data.
4. A sensor sensing user-asserted data, determined through a *Sensor-senses-Data* relationship.

New Threat Patterns have been specified, which are based on the first three Matching Patterns. These encode the threats and their consequences identified in the threat-consequence state mapping diagram. One such threat pattern is given in Figure 5, where timeliness of data being undermined results in a “loss of integrity”, unless one of the two controls is active.

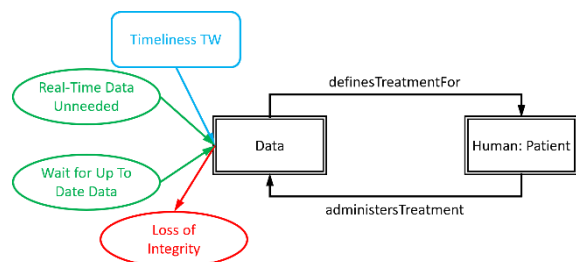


Figure 5: Threat pattern linking loss of timeliness to loss of integrity.

Here the Data (treatment decision data) becomes out of date (timeliness trustworthiness), causing a loss of integrity of the Data, due to the starting assumption that near real-time data is needed for the individual to

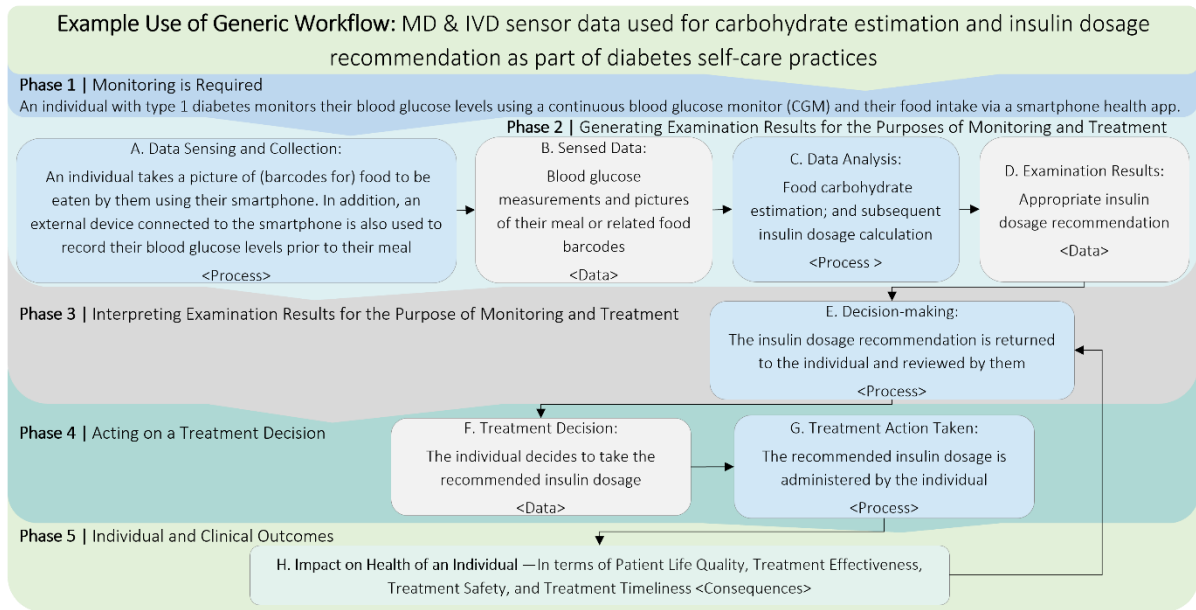


Figure 6: Use of Workflow: Use case example — IVD sensor data used for food calorie estimation and insulin dosage recommendation in diabetes self-care practices.

correctly treat themselves. There are two controls that block this threat from occurring. Firstly, it can be asserted that near real-time data is not needed, and secondly, the patient can wait for up-to-date data before taking the treatment actions defined by it.

With the patient harms encoded, an illustrative example is given next.

6 ILLUSTRATIVE EXAMPLE

Worked use cases with clinical and industrial partners, as part of the NEMECYS project, have guided the development of the indirect patient harms-related extensions. Here one such use case is used as an illustrative example for our approach. This scenario was chosen as it builds on the examples in Annex H of ISO/TR 24971 (pp. 62-85) (ISO 2020).

For this, the generic workflow of Figure 3 has been specialised to illustrate the workflow and patient harm domain modelling applied to the sensor data being used in the self-care management of diabetes. This **specialised workflow** is shown in Figure 6 and describes a scenario where an individual with type 1 diabetes regularly monitors their blood glucose levels and food intake to help manage their condition. The individual monitors both their blood glucose levels prior to a meal through using a continuous blood glucose monitor (CGM) and their food intake by taking pictures of their meal or the barcodes of the

food, that are then analysed in the cloud. This data is collected via a “Software as a Medical Device” (SaMD) app and uploaded to the cloud where food carbohydrate levels are estimated and, with this and the initial blood glucose levels, an appropriate insulin dosage is calculated for the individual and their given meal. This is communicated back to the individual so they can administer the recommended insulin dosage.

The **system model** shown in Figure 7 models this use case and focuses on the cybersecurity threats involved in the data sensing, collection, processing and communication, which then link to the newly included indirect patient harms. There are two types of sensed data within this use case: the pictures taken by the individual of their food / barcodes; and their initial blood glucose measurements. Within the system model, these two data elements are given as a single Data asset that achieves the same results.

The use case workflow above has steps identified by letters A-H. Steps A-F (up to the treatment decision) are covered within the system model:

- Health Sensor (A) senses User Phone Data (B), which encapsulates both the blood glucose data and food images / barcodes.
- User Phone Data (B) is stored locally on the individual’s smartphone and also uploaded to the cloud (C)
- In the cloud, User Phone Data is used to update User Cloud Data (D).

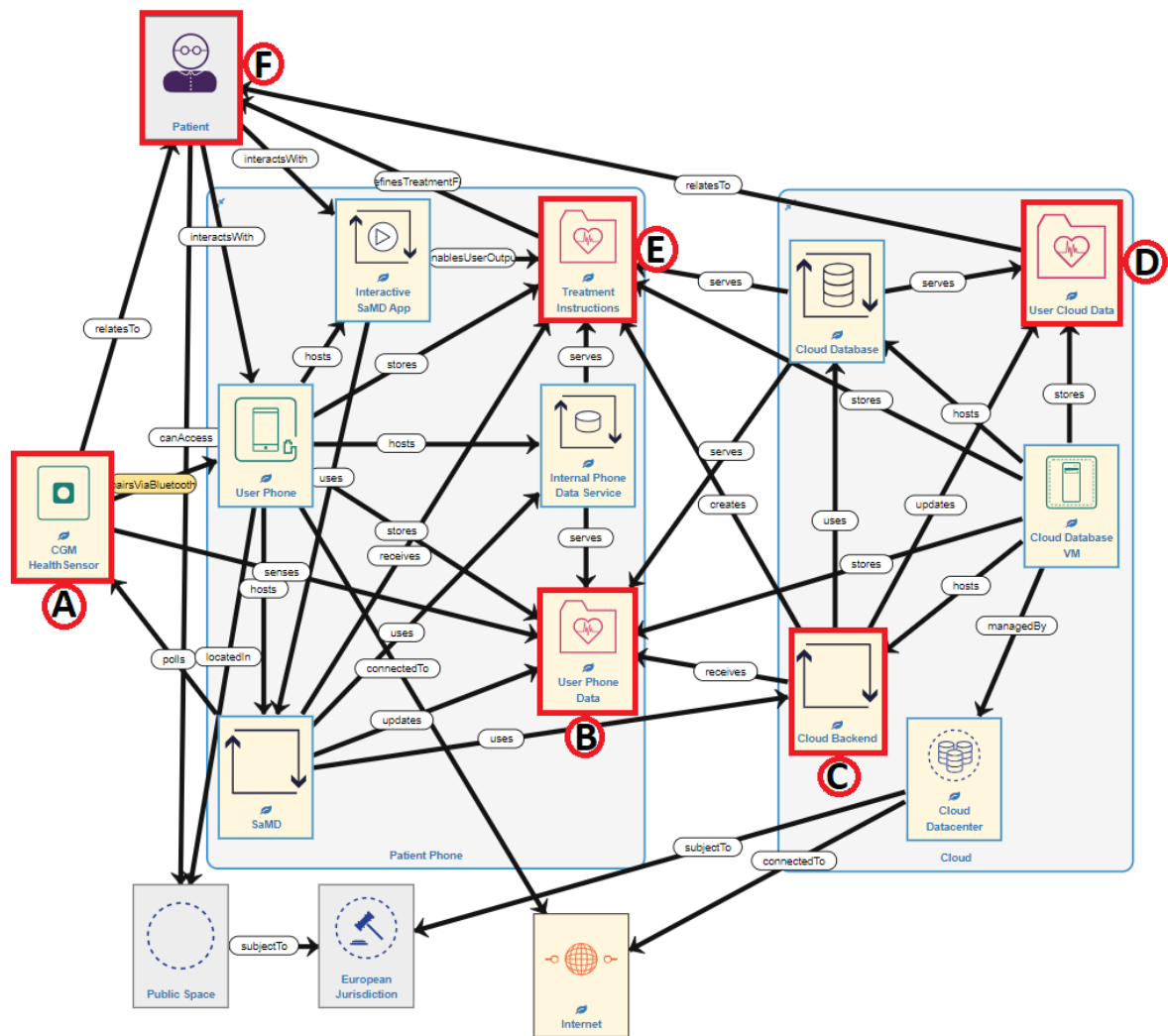


Figure 7: System model for self-managed type 1 diabetes use case.

- These are used to calculate the Treatment Instructions Data (E), which is transmitted back to the individual's smartphone.
- The individual views Treatment Instructions via the smartphone Interactive SaMD App and actions the Treatment Instructions (F).

Figure 8 shows the risks present. Each row corresponds to a risk, which is a Consequence occurring at an Asset, with associated Impact and Likelihood. To manage typical cybersecurity risks like Loss of Confidentiality of the health data assets, typical cybersecurity controls have been applied, like the different health data assets being encrypted and transmitted securely between the smartphone and cloud. This reduces most risks to a medium risk level or lower. However, **Inappropriate Treatment of the**

Patient and Loss of Authenticity of the User Phone Data are still High risks.

Spyderisk is then used to trace between Consequences and Threats to work backwards to the **Root Cause Threat**. A Root Cause Threat is the very first threat in a Threat Path, enabling the path to occur. The root cause to **Inappropriate Treatment of the Patient and Loss of Authenticity of the User Phone Data** is that if the Health Sensor is spoofed when paired with the patient smartphone, particularly if pairing occurs in the Public Space, then the "imposter" Health Sensor will have a **Loss of Authenticity** and report incorrect blood glucose levels to the SaMD. This leads to a **Loss of Reliability** in that process, propagating to a **Loss of Reliability** in the Cloud Backend, a **Loss of Authenticity** in the User Cloud Data, and a **Loss of Authenticity and Integrity** in the Treatment

Instructions, finally resulting in Inappropriate Treatment of the individual.

Consequence	Asset	Direct Impact	Likelihood	Direct Risk
InappropriateTreatment	Patient	High	Medium	High
LossOfAuthenticity	User Phone Data	High	Medium	High
AbsenceOfTreatment	Patient	High	Low	Medium
DelayedTreatment	Patient	High	Low	Medium
LossOfAuthenticity	Treatment Instructions	High	Low	Medium
LossOfAuthenticity	User Cloud Data	High	Low	Medium
LossOfIntegrity	Treatment Instructions	Medium	Medium	Medium
LossOfIntegrity	User Cloud Data	Medium	Medium	Medium
LossOfIntegrity	User Phone Data	Medium	Medium	Medium
LossOfAvailability	Treatment Instructions	Medium	Low	Low

Figure 8: Initial system risks.

A Simple Secure Pairing control on the smartphone so the correct Health Sensor pairs with it blocks this threat path, and applied it reduces the Inappropriate Treatment risk to Medium, shown in Figure 9. The Medium risk level remains due to the consequence impact being high as it involves correct treatment of a patient, whereas the likelihood of it occurring becomes Low, reduced from Medium.

The control to wait for up-to-date data on the Treatment Instructions has been applied, indicating the patient will wait for up-to-date treatment instructions before following them and taking their insulin dosage. This blocks the threat paths for Loss of Timeliness in the Treatment Instructions and Inappropriate Treatment of the individual as they will now not be following treatment instructions that are based on outdated information. However, it enables a threat path leading to Delayed Treatment, which

could affect the individual as they may require taking the correct insulin dosage soon after having their meal. This illustrates that there are different potential control strategies and trade-offs that may need to be considered. Spyderisk does not make the trade-off decision, though it does provide decision support information in terms of the possible consequences of an intervention, which can be taken into consideration when determining an appropriate course of action.

This demonstrates that non-emergency, indirect patient harm risks related to cybersecurity incidents are modelled in Spyderisk automated risk assessment approach, and initial risk-benefit analyses can be conducted to consider how some patient harm-related controls can block certain threats and risks, but enable others, and so need to be considered carefully.

7 CONCLUSIONS

Knowledge of non-emergency indirect patient harms has been collected and translated into the terminology of the Knowledge Base that is part of an existing cybersecurity risk simulator called Spyderisk. This knowledge has then been used to determine threat paths linking cybersecurity incidents to indirect patient harms and this has been encoded inside that Knowledge Base. An illustrative example following a guiding use case has also been presented.

We envision this work will continue by increasing the link and modelling between these two domains, and see this work as important since understanding, preventing and mitigating cybersecurity risks that result in patient harms is important as they can have profound effects on the health and wellbeing of individuals using CMD&IVD.

Control Strategies (1/1)

★ **SimpleSecurePairing.Local** (Safe effectiveness)

Simple Secure Pairing (SSP) is used between "CGM HealthSensor" and "User Phone", following the Just Works association model with user confirmation at "User Phone". This is effective in preventing spoofing in insecure locations, but depends on there being one secure location where the numerical comparison can be made safely, and then the result stored for subsequent use.

● SimpleSecurePairing at "User Phone"

☐ Accept threat

Consequence	Asset	Direct Impact	Likelihood	Direct Risk
InappropriateTreatment	Patient	High	Low	Medium
LossOfAuthenticity	User Phone Data	High	Low	Medium
AbsenceOfTreatment	Patient	High	Low	Medium
DelayedTreatment	Patient	High	Low	Medium
LossOfAuthenticity	Treatment Instructions	High	Low	Medium
LossOfAuthenticity	User Cloud Data	High	Low	Medium
LossOfIntegrity	Treatment Instructions	Medium	Low	Low
LossOfIntegrity	User Cloud Data	Medium	Low	Low
LossOfIntegrity	User Phone Data	Medium	Low	Low
LossOfAvailability	Treatment Instructions	Medium	Low	Low

Figure 9: Effect of secure pairing controls on treatment.

ACKNOWLEDGEMENTS

Project report. This conference paper has been adapted from part of a NEMECYS project deliverable report: D2.1 Risk Benefit Schemes (initial) (found at: <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e50edeaf43&appId=PPGMS>).

Funding statement. This work has been conducted as part of the NEMECYS project, which is co-funded by the European Union (101094323), by UK Research and Innovation (10065802, 10050933 and 10061304), and by the Swiss State Secretariat for Education, Research and Innovation.

We would like to give thanks to our project partners for their involvement in risk modelling related use case discussions. Please note that all views expressed in this paper are those of the authors, and do not necessarily represent those above.

REFERENCES

- Balogh, E. P., Miller, B. T., Ball, J. R., Committee on Diagnostic Error in Health Care, Board on Health Care Services, Institute of Medicine, & The National Academies of Sciences, Engineering, and Medicine (Eds.). (2015). *Improving Diagnosis in Health Care*. National Academies Press (US).
- Biasin E, Kamenjasevic E. Cybersecurity of Medical Devices: Regulatory Challenges in the European Union. In: Cohen IG, Minssen T, Price II WN, Robertson C, Shachar C, eds. *The Future of Medical Device Regulation: Innovation and Protection*. Cambridge University Press; 2022:51-62. <https://doi.org/10.1017/9781108975452.005>.
- International Organisation for Standardization (ISO). (n.d.). An easy guide to understanding healthcare data analytics. <https://www.iso.org/healthcare/data-analytics>
- International Organisation for Standardization (ISO). (2018). *Information technology – Security techniques – Information security risk management systems – Overview and vocabulary* (ISO Standard No. 27000:2018).
- International Organisation for Standardization (ISO). (2019). *Medical devices — Application of risk management to medical devices* (ISO Standard No. 14971:2019).
- International Organisation for Standardization (ISO). (2020). *Medical devices — Guidance on the application of ISO 14971* (ISO Standard No. TR 24971:2020).
- International Organisation for Standardization (ISO). (2022). *Information security, cybersecurity and privacy protection – Guidance on managing information security risks* (ISO Standard No. 27005:2022).
- Medical Device Coordination Group (MDCG) (2019). *MDCG 2019-16 Guidance on Cybersecurity for medical devices*. <https://ec.europa.eu/docsroom/documents/41863>
- Medical Device Coordination Group (MDCG) (2023). *MDCG 2023-3. Questions and Answers on vigilance terms and concepts as outlined in the Regulation (EU) 2017/745 on medical devices*. https://health.ec.europa.eu/system/files/2023-02/mdcg_2023-3_en_0.pdf
- Mukoro, F. (2011). *Care Planning — Mini Topic Review*. NHS Kidney Care. <https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/Care-Planning-Mini-Topic-Review.pdf>
- NEMECYS. (2023). NEMECYS Use Cases. <https://nemecys.eu/about-us/use-cases>
- Phillips, S. C., Taylor, S., Boniface, M., Modafferi, S., Surridge, M. (2024). Automated Knowledge-Based Cybersecurity Risk Assessment of Cyber-Physical Systems. *IEEE Access*, 12, 82482-82505. doi: 10.1109/ACCESS.2024.3404264
- National Health Service [NHS] England. (2023). Medical devices and digital tools. Version 1.2, 25 May 2023. <https://www.england.nhs.uk/long-read/medical-devices-and-digital-tools/>.
- Piggin, R. (2017). Cybersecurity of connected medical devices. BSI White Paper. Available at: <https://www.bsigroup.com/meddev/LocalFiles/en-US/Whitepapers/bsi-md-whitepaper-cybersecurity.pdf> (Accessed 18 October 2024).
- Regulation 2017/745. *Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance.)*. <http://data.europa.eu/eli/reg/2017/745/oj>
- Regulation 2017/746. *Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU (Text with EEA relevance.)*. <http://data.europa.eu/eli/reg/2017/746/oj>
- Spyderisk. (2024). Spyderisk Open Project. <https://github.com/Spyderisk>
- Taylor, S.; Jaatun, M.; Mc Gibney, A.; Seidl, R.; Hrynchenko, P.; Prosvirin, D. and Mancilla, R. (2024). A Framework Addressing Challenges in Cybersecurity Testing of IoT Ecosystems and Components. In *Proceedings of the 9th International Conference on Internet of Things, Big Data and Security*, ISBN 978-989-758-699-6, ISSN 2184-4976, pages 226-234.
- Van Landuyt, D., Joosen, W. (2022). A descriptive study of assumptions in STRIDE security threat modeling. *Softw Syst Model* 21, 2311–2328