HORIZON 2020

The EU Framework Programme for Research and Innovation

CAPABLE

Cancer Patients Better Life Experience

Grant Agreement No. 875052

Start Date: 01/01/2020 (48 Months)

Deliverable No. 7.2

AI ethics and incidental findings policy

Due Date: [31/12/2020]

Submitted On: [31/12/2020]

Resbmitted On: [15/08/2023]

Coordinator	University of Pavia (UNIPV)
Deliverable Lead Partner	Prof. Lucia Sacchi [UNIPV]
Contributing Partners	[UNIPV, UH, IBM, BIT, PUT, ICS MAUGERI, NKI, Deontics, AIMAC, UPM]
Contact	Prof. Silvana Quaglini
Email	silvana.quaglini@unipv.it
Website	www.capable-project.eu

Deliverable Type		
R	Document, report	
DEM	Demonstrator, pilot, prototype	[X]
DEC	Websites, patent fillings, videos etc.	
OTHER		
Dissemination Level		
PU	Public	[X]
CO	Confidential (Consortium members including the Commission Services)	
CI	Classified Information (Commission Decision 2015/444/EC)	

This deliverable is a part of a project receiving funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 875052.

Table of contents

1 Executive Summary 2
2 Preliminary and brief description of the purposes of the project 2
3 Incidental findings 3
3.1 Introduction 3
3.2 Events 6
4 Communication to participants 12
5 Data Protection Impact Assessment 12
5.1 Risk Assessment 13
6 References 15

1 Executive Summary

The aim of this deliverable is to describe the approach that the CAPABLE consortium will have with respect to the ethics issues that might arise during the development and use of the CAPABLE AI-based clinical decision support system. In addition, the document also describes the policy related to incidental findings detection and management.
This is an initial document that provides an overview of the foreseen issues, with specific focus on ethical and legal evaluations about data management, incidental findings, and residual risks. Some specific cases related to the project development are detected and described. The document is structured as follows: after a general introduction on the project and its objectives (Section 2), incidental findings and risks are defined, together with the policy that CAPABLE will adopt to detect and manage them (Section 3). In Section 4 we present how CAPABLE will deal with the communication of such findings and risks to the users of the system. Section 5 details the data protection impact assessment (DPIA).

2 Preliminary and brief description of the purposes of the project

After the primary intervention, most cancer patients are managed at home, facing long-term treatments or sequelae, making the disease comparable to a chronic condition. Despite their benefit, strong therapeutic regimens often cause toxicity, severely impairing quality of life. This may decrease adherence to treatment, thus compromising therapeutic efficacy. Also due to age-related multimorbidity, patients and their caregivers develop emotional, educational and social needs.
CAPABLE is developing a cancer patient coaching system with the objective of facing these needs/issues. It will fully exploit Artificial Intelligence (AI) and Big Data potentialities for cancer care and bring them to patients' homes.
CAPABLE will rely on predictive models based on both retrospective and prospective data (clinical data, data from unobtrusive environmental and wearable sensors, data from social media and questionnaires). Models will be integrated with existing clinical practice guidelines and made available to oncologists.
Thanks to the mobile coaching system for patients, CAPABLE will allow providing patient-specific decision support. This feature, together with the chance of discovering unknown adverse effects of most recent treatments, makes CAPABLE more than a personalised tool for improving life quality, an advance for the whole research community.

The Consortium acknowledges and believes that is it extremely important to manage possible risks in this type of systems. This is the starting point of the following document, which reflects our initial evaluation of the scenarios that CAPABLE may face during its development and use.
It is preliminary to note here that:
1) The risks identified in this Deliverable are subject to an ongoing monitoring process specifically designed for CAPABLE that includes internal and external monitoring processes and bodies.

2) This monitoring is also part of the general risk management strategy.

It is also important to provide some definitions. *Incidental risks* are the risks of misuse of the system caused by internal and external factors. These risks can be mitigated (reduced) using suitable policies. The risks that remain after these policies have been implemented and taken full effect are termed *residual risks*.
To identify possible sources of incidental risks we have reviewed privacy and data protection inquiries and the academic literature on the subject. This investigation has focused particularly on the recent literature and experience following the introduction of new European Data protection regulations, the General Data Protection Regulation (GDPR) in 2016 (enacted on May 25th 2018).

This Deliverable is focused on ethical and legal evaluations about data management, incidental findings, and residual risks, to set a policy for the CAPABLE project.

3 Incidental findings

3.1 Introduction

The notion of incidental findings originated in medical and genetic research. Incidental findings are traditionally defined as results that are outside the original purpose for which a test or procedure was conducted.
According to the literature [1], incidental findings are distinct from *primary findings*, which are the results that are actively sought as the primary target of a test or procedure.
They can be either “anticipatable” or “unanticipatable.” An *anticipatable incidental finding* is one that is known to be associated with a test or procedure. Anticipatable incidental findings need not be common or even likely to occur—their defining characteristic is that the possibility of finding them is known.
Unanticipatable incidental findings include findings that could not have been anticipated given the current state of scientific knowledge. Researchers cannot plan for these types of findings specifically. However, they can consider in advance what they might do if a particular kind of unexpected finding arises, for example, one that could be actionable or lifesaving.
A *secondary finding*, by contrast, is not the primary target of the test or procedure; rather, it is an additional result actively sought by the practitioner. Secondary findings might be sought deliberately when doing so is recommended by an expert body or by a consensus of practitioners. Table 3.1.1 provides examples of each type of finding.

TYPE OF RESULT DISCOVERED	DESCRIPTION	EXAMPLE
Primary Finding	Practitioner aims to discover A, and result is relevant to A	In a child with unknown vaccine history, a test done to determine a child's immunity status before the chickenpox vaccine is administered
Incidental Finding: *Anticipatable*	Practitioner aims to discover A, but learns B, a result known to be associated with the test or procedure at the time it takes place	Discovering misattributed paternity when assessing a living kidney donor and potential recipient who believe they are biologically related
Incidental Finding: *Unanticipatable*	Practitioner aims to discover A, but learns C, a result not known to be associated with the test or procedure at the time it takes place	When a DTC genetic testing company identifies a health risk based on a newly discovered genetic association not knowable at the time a previous sample was submitted
Secondary Finding	Practitioner aims to discover A, and also actively seeks D per expert recommendation	ACMG recommends that laboratories conducting large-scale genetic sequencing for any purpose should actively look for variants underlying 24 phenotypic traits

Table 3.1.1 - Description and examples of research results. Source: [2].

The most pressing ethical questions in the debate on the management of incidental findings as it pertains to medical research are often summarised as follows [3]:
• should the physician be obligated to report all such findings back to the patients, or just some findings—in that case, which ones, or none?
• should the patients have a right to demand such results to be delivered to them under all circumstances, or should they be allowed to refuse to receive any such information?
• should a patient with a genetic variant implicated in the development of serious, but preventable/treatable clinical condition be allowed to refuse to know such information and consequently withhold it from family members that can also be carriers of that same genetic variant?
• should some genetic variants that can cause preventable/treatable clinical conditions that come up as incidental results in genome-scale screening testing be actively sought in such testing, becoming thus a secondary instead of incidental finding, or, in fact, a regular finding of the clinical screening?
Detection and feedback of incidental findings can be a "double edged sword", as they may allow for timely treatment (thus leading to medical benefit) but may also harm research participants because of the burdens of costs of follow-up testing and (possible) over- treatment [4].

The ethical path we have elaborated includes:

(i) Thinking about anticipatable incidental findings
(ii) Preparing a set of information provisions for the informed consent of the research participants (research participants will be given the opportunity either to opt out of receiving information about incidental findings or to withdraw from the study);
(iii) communication of the incidental finding policies to the research participant should align with national regulations and customs;
(iv) definition of the person/institution who will take responsibility for the clinical follow-up of the research participant.

The project, also thanks to the incidental findings policy, states its:

(i) compliance with laws, regulations and court assessments,
(ii) technological conformance with existing standards,
(iii) congruence with ethical principles, could be better achieved if based on empirical knowledge and reasonable estimation.

According to the literature, it should be considered carefully whether to allocate time and resources to seeking secondary findings, or to interpreting, assessing, and disclosing incidental findings, especially when these decisions might benefit individuals in the research study but stall broader societal benefits of the research activity. Researchers do not have an ethical duty to seek secondary findings. However, researchers must determine how their incidental findings management policy will affect participants as individuals, and how it will affect their ability to contribute to generalizable knowledge.
The Consortium takes the responsibility to make choices according to the principles listed in Table 3.1.2.

Principle	*Definition*	*Application*
Respect for Persons	This principle recognizes the fundamental human capacity for rational self- determination.	Researchers must communicate the fundamental aspects of their research – including the possibility of discovering incidental or secondary findings and the plan for their disclosure or management – so that participants can make informed decisions about whether to enrol.
Beneficence	This principle calls on professionals to take action to ensure the wellbeing of others. Its corollary, non-maleficence, requires not imposing harm on others.	This principle supports returning findings when disclosure might help forestall or prevent harm. By contrast, disclosing an incidental finding for which no preventive or positive action can be taken has the potential to cause anxiety and distress with no corresponding medical benefit.
Justice and Fairness	This principle requires fair and equitable distribution of the potential benefits and burdens across society.	The principle of justice and fairness calls upon researchers to take into account how policies for returning incidental and secondary findings could benefit or burden some participants or, alternatively, could burden the research enterprise and the ability to contribute to generalizable knowledge.

Intellectual Freedom and Responsibility	This principle protects sustained and dedicated creative intellectual exploration that furthers scientific progress, while requiring that researchers take responsibility for their actions.	This principle supports affording wide latitude to researchers in pursuing their scientific goals and engaging in intellectual exploration for the good of society, while also expecting that researchers uphold and respect the trust placed in them by participants. Ethical conduct of research with human participants includes acknowledgment and planning for incidental and secondary findings.

Table 3.1.2 - Principles that constitute the foundation of the CAPABLE incidental findings policy. Source: [2].

These principles are explicitly recognized and promoted also by important European documents, such as:

- the Declaration of Helsinki, Ethical Principles for medical research involving human subjects;
- the Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine (Oviedo, 4 April 1997) (Oviedo Bioethics Convention)- Commission Directive 2005/28/EC of 8 April 2005 laying down principles and detailed guidelines for good clinical practice as regards investigational medicinal products for human use as well as the requirements for authorization of the manufacturing or importation of such products (OJ L 91, 9.4.2005, p. 13).
- The Regulation No 536/2014 of the European Parliament and of the Council on clinical trials on medicinal products for human use, repealing Directive 2001/20/EC (OJ L158, 27/5/2014).

3.2 Events

Tables 3.2.1 and 3.2.2 represent possible cases of Incidental Findings and Clinical Incidental Risks in the CAPABLE project. They will be periodically updated during the development of the project, and especially during the clinical trial phase.

Incidental findings	Management	Subject involved
From collected data, physicians discover a new disease	The patient will be informed about the possibility of anticipatable or unanticipatable cases in the informed consent. The physician who is taking care of the patient will inform him/her about the disease, offering all possible information	Physicians

	about treatment and contacts of specialized physicians	
From collected data, the decision support system reveals a risk of diseases linked to the cancer (e.g. cardiovascular risk)	The patient will be informed about the possibility of the discovery of new risks in the informed consent. The physician who is taking care of the patient will inform him/her about the increased risk, proposing the patient the useful/necessary actions to take.	Physicians

Table 3.2.1 - examples of incidental findings within the CAPABLE project. For each finding, we provide a possible management strategy, and we identify the subjects that are involved in the management.

Incidental Risks	**Management**	**Subject/component involved**
Malfunction of the system that delivers wrong/missing recommendation to the patient (e.g. nutrition advice)	Human periodical check and immediate information to the patient about how to change behaviour	Physicians who take care of the patient; Software engineers
Malfunction of the system that delivers wrong medical indication to the patient (e.g. which pills to take)	Human frequent periodical check and immediate information to the patient about how to change behaviour	Physicians who take care of the patient; Software engineers
Patients' misbehaviours such as taking supplements or substances (herbals, etc) that have not been reported.	Informing initially and periodically the patient about the necessity to report all new supplements or substances taken during the trial period	Physicians and app notifications
Patients' misbehaviours such as not being compliant with the therapeutic contract (drugs combined with physical and mental activity)	A clear statement in the Informed Consent about the terms of the therapeutic contracts and delivering periodical recommendations about the compliance to it	Physicians and app notifications.
Patient does not report his data	Both patients and healthcare professionals will be reminded of the importance of providing complete data. In case the system detects a persistent abstention from the use by the	Physicians and app notifications.

	patient, an alert will be sent to the patient and the healthcare provider.	

Table 3.2.2 - examples of incidental risks within the CAPABLE project. For each case, we provide a possible management strategy, and we identify the subjects or system components involved.

Monitoring of incidental findings and risks

During the clinical study, when real users will test the system, we will implement a monitoring strategy to detect possible events that generate a potential risk or an incidental finding.
Both the users and the technical partners will take part in the monitoring process.
Each time a potentially risky behaviour of the system or incidental finding is detected, it should be reported. In particular, the following information are needed:

- who detected the event
- date of the event
- type of event (incidental finding -anticipatable or not- or incidental risk)
- description of the event
- patient(s) involved
- user and component involved
- was the event detected at the same time it happened?
- was the event managed? how? when?
- was there any consequence of the event?

Some of the information will be reported by the user who experienced the event, some others will be completed by the CAPABLE research team. In general, the system users will notify the first level support at the clinical center, who will be responsible to report the event to the CAPABLE research team, who will meet periodically to analyze the cases.

Managing the risks of wrong or unethical decisions

Explainability and traceability

In order to support explainability and interpretability of decision models, we promote formal and symbolic representation of the domain knowledge. In particular we employ the PROforma language [5] to represent broadly understood clinical rules and workflows, such as clinical practice guidelines, clinical pathways and other algorithms employed by the CAPABLE system (for example, algorithms for recommending the so-called “well-being capsules” to a patient, which are non-pharmacological interventions to improve the mental well-being and managing stress). PROforma requires explicit specification of considered data items and conditions imposed on these items that are associated with specific decisions, which contributes to *intrinsic* interpretability of applied models. In order words, a model can be relatively easily verified and tested before it is deployed to practical use. A sample

guideline modeled in PROforma is presented in Figure 3.2.1.

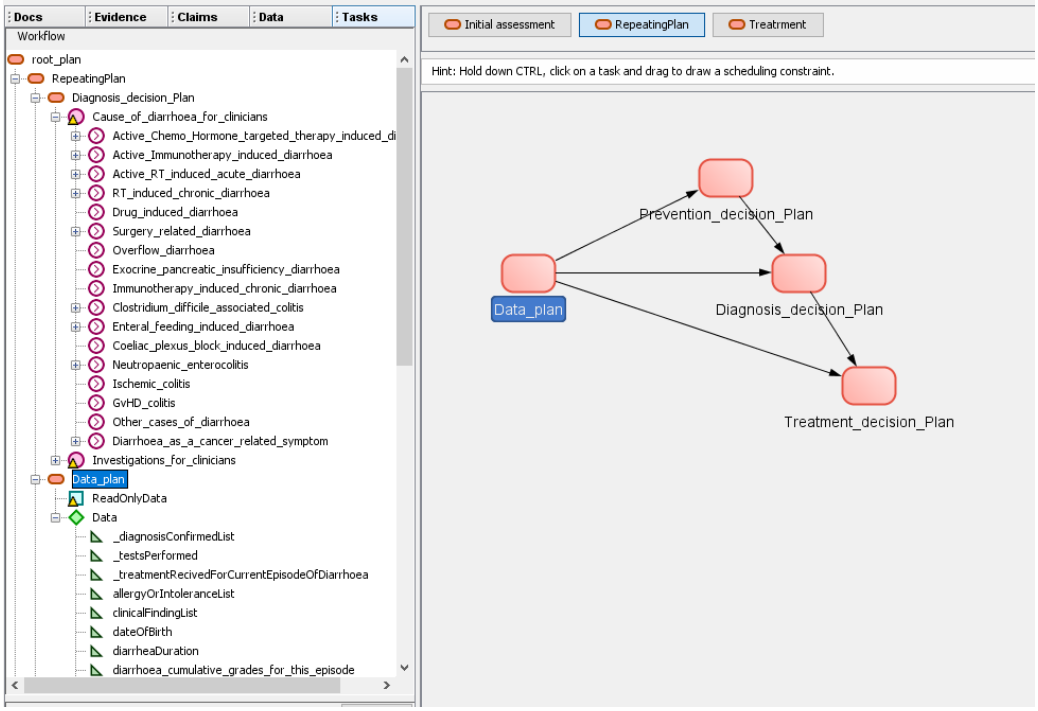

Figure 3.2.1 - Example of a guideline modeled using the PROforma language.

Moreover, PROforma modeling and execution tools allow for associating relevant parts of evidence-based documents, such as guidelines or systematic reviews. These parts can be presented to a decision maker (physician) when considering possible options for a given decision. This further enhances the *post-hoc* interpretability of the models -- the decision maker is not only able to trace the "path" in the model followed for a specific patient, but is also offered sound justification behind specific choices. This is illustrated in Figure 3.2.2, where the upper part presents possible treatment options (candidates), and the bottom part brings relevant parts of the guideline.

We should also note that decision models in PROforma are meant to be used as assistants to human decision makers (physicians and patients) in order to ensure the required level of human autonomy. Thus, physicians and patients are provided by decision recommendations established by CAPABLE, and they can either accept or discard them and freely choose options that are not among the ones suggested by the system.

During later stages of the system development we plan to employ "blackbox" models, such as convolutional neural networks, for personalization of capsules. These models will rely on sensor data, such as blood volume pulse or other biomarkers that may be related to stress. Since such models are not directly interpretable, we will combine them with explainable AI methods, e.g., LIME, to provide better insight into operations and reasoning of these models. In any case, no critical decision will be taken by the system without the intervention of the healthcare personnel.

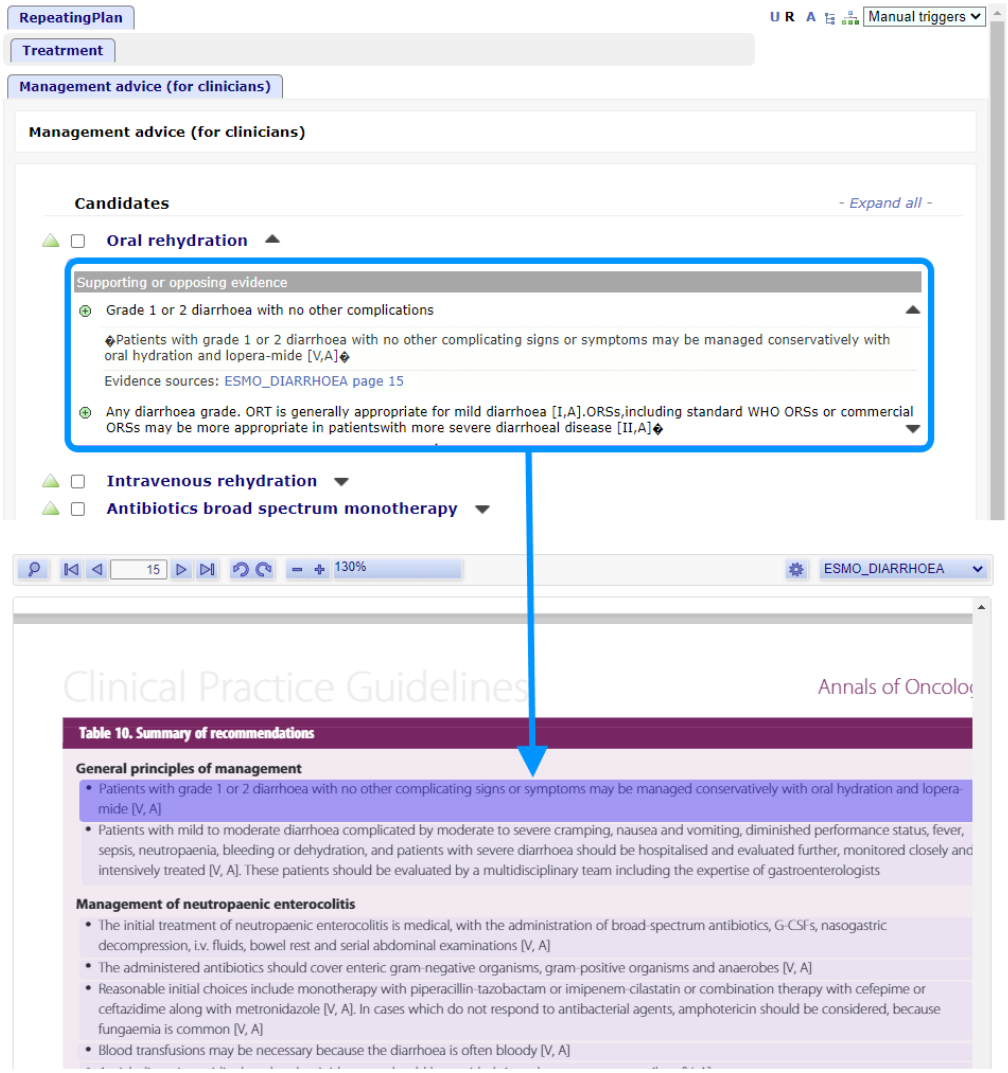

Figure 3.2.2 - Explanation of evidence in PROforma.

Avoidance of unfair bias

CAPABLE will employ two types of data-driven predictive models -- personal and population-based. Personal models have been already mentioned above in the context of capsule personalization. We plan to build a separate model for each patient managed by the system, thus bias should have negligible impact on models' operations. However, the bias, related for example, to age, gender, or specific type of therapy, may influence population-based models. We plan to identify possible biases during the data preprocessing stage by applying exploratory data analysis methods and address them by applying appropriate resampling techniques to make data sets more "representative" for considered problems. This step will be conducted through cooperation of technical and clinical teams.

4 Communication to participants

Possible incidental findings and risks will be communicated to participants during the informed consent process.
This allows individuals to choose not to participate in research if they are uncomfortable with the project's management plan. The consent materials will include information about the following elements:

- Secondary findings that will be actively sought and returned to participants should be conveyed in the informed consent process, and there should be a specific plan for their return.

- A plan for anticipatable incidental findings (e.g., that researchers will or will not return some or all potential findings)

The plan for managing incidental findings detailed in the informed consent will include also a description of the research team's responsibilities following disclosure of such a finding, also providing:

- basic educational information about the nature of the finding;
- advice regarding how to seek care from a clinician or specialist;
- guidance about obtaining health insurance to secure treatment; and/or
- a referral to a clinical specialist, if one is required.

The contact details of the support in case of malfunctioning or unexpected behaviour will also be included in the user manuals of the system. In this way, if during the pilot study one of the users experienced some problems, he/she is informed on how to proceed to report the issue.

5 Data Protection Impact Assessment

The first step is to assess the necessity and proportionality of the intended data processing, which the Article 29 Working Party (WP29) has advised to be done, considering: a) the specified, explicit and legitimate purpose of the processing; b) the lawfulness of processing, and c) the principle of minimization, which requires the data to be adequate, relevant and limited to what is necessary for the objective.

As already detailed in the Deliverable D1.2 (Data Management Plan), the overall objective of CAPABLE is to combine the most advanced technologies for data and knowledge management with a sound socio-psychological approach in order to develop a coaching system for improving the quality of life of cancer patients.
The system aims at early detecting and managing cancer-related issues and at satisfying the needs of patients and their home caregivers.

Ultimately, CAPABLE will exploit several different datasets using AI techniques to effectively monitor individual patients, with the final goal to improve quality of life after cancer treatment. More specifically, the data collection and analysis activities in the CAPABLE project will help achieve the following objectives:

- Identifying, classifying and ranking new cancer patients' and their home caregivers' needs, mostly leveraging on data provided by the AIMAC patients' association, interviews and questionnaires to be administered in the requirements elicitation work package (WP2).
- Improving patients' compliance to treatment by acquiring Patient Reported Outcomes (PROs) and Patient Reported Experiences (PREs).
- Collecting data for early identification of deterioration in quality of life or emotional issues.
- Improving healthcare professional workflows by promptly identifying priority patients and shortening the duration of control visits due to a better understanding of the patient conditions, thanks to data collected in-between visits at the patient home.
- Identifying adverse events of (relatively) new therapies or unknown long-term effects of cancer treatment.
- Developing new, data-driven AI models for the course of cancer, which could drive more personalized interventions.

The system will rely on both data already available to partners at the beginning of the project and on data that will be collected during the clinical study, which will last the entire fourth year of the project. The clinical study, that will take place at the two clinical partner organizations ICSM and NKI, will enroll kidney cancer (ICSM) and melanoma patients (NKI). Thus, the data collected by the project pilot will be focused on these two cancer patient populations, but many of the findings intend to be generalizable to other cancer domains. Table 1 provides a summary of the data that CAPABLE will collect. All the data collected by the project during the pilot studies will be stored in a centralized data repository based on the OMOP CDM [1], in order to improve standardization and promote reusability. Every action will be taken according to the principle of minimization, including only data that have been evaluated useful for the project.

The lawfulness of the processing is granted by the Informed Consent that will be signed by users and participants.

5.1 Risk Assessment

The initial risk-related scenarios that we envisage are summarized in Table 5.1.1.

Data protection risks	Management
Sharing patient's data with the caregiver (both at home, e.g. for not self sufficient patients, and at nursing institutions)	The caregiver is the data processor and thus legitimate to manage data, according to the recommendations that will be delivered and signed.
Data breach: theft of the	The access to the patient's data is protected by a password

smartphone or the tracker	The Capable app can be deactivated remotely (as long as data connection is on) by the technical team/hospital IT. This will prevent any further attempt to login and use the Capable patient app. Also, clinical and personal data stored locally in the patient phone by the Capable app will be encrypted. If the patient reports the phone was lost or stolen, the deactivation will also cause a full data wipe-out from the smartphone local storage. Data will be kept in the centralized Capable Data Platform.
Data breach; a third party access the patient's data on the smartphone or Hacking activity	The access to the app is protected by a password In case of hacking, there will be an immediate notification to the national Data Protection Authority as provided by the GDPR. The access to the patient's data is protected by a password The Capable app can be deactivated remotely (as long as data connection is on) by the technical team/hospital IT. This will prevent any further attempt to login and use the Capable patient app. Also, clinical and personal data stored locally in the patient phone by the Capable app will be encrypted. If the patient reports the phone was lost or stolen, the deactivation will also cause a full data wipe-out from the smartphone local storage. Data will be kept in the centralized Capable Data Platform.
From collected data, patient's habits can be inferred	All people professionally involved in the project shall be compliant with the data protection regulations and will be entitled to use inferred information only for project-related uses.

Table 5.1.1- Description of data protection risks and their management.

While any DPIA must be carried out before the processing of personal data begins, it should be considered a 'live' document. This means this document will be subject to regular review or re-assessment during the development of the project.

6 Document revision July 2023

This section is an addendum to D7.2 required after the Ethics Check performed in April 2023. Besides the revisions requested by the Ethical Review Committee, we also provide an update on Section 5 related to the Data Protection Impact Assessment and the Risk Assessment.

To perform the CAPABLE pilot study in compliance with the current regulations in terms of Data Protection and Medical devices, we have prepared a Data Protection Impact Assessment document (DPIA) and a set of documents related to the risk assessment process. The DPIA was approved by

the DPOs of the centres involved in the clinical studies and signed by their legal representatives, whereas the risk assessment procedure was submitted and approved by the National Authorities involved in the MD evaluation of the clinical study. The documents are described and attached to deliverable D7.6.

6.1 Assessment based on the Ethics Guidelines for Trustworthy AI and Ethics By Design and Ethics of Use Approaches for AI

Since CAPABLE relies on a number of AI-based components, in this section we provide an assessment of the system according to the relevant guidelines in the field [6,7]. The scope of this assessment is to show how ethical and robust AI principles have been operationalised in the CAPABLE system. To prepare this assessment we rely on the Trustworthy AI assessment list proposed in [6], adapted and contextualised to the scope of the CAPABLE project. This is presented in the Table below.

Human agency and oversight		
Fundamental rights	The AI system interacts with decisions by human (end) users by providing guideline-based recommendations both to patients and clinicians	Human autonomy in the decision making process is always preserved – any action proposed by the system requires medical intervention for approval. Any unintended behaviour is considered and mitigated in the risk assessment process
	The AI system should communicate to (end) users that a decision, content, advice or outcome is the result of an algorithmic decision.	Users are informed that advices and recommendations are the result of an algorithmic decision. This process takes place during training and information is also provided in the instructions for use.
Human agency	The AI system is implemented in clinical practice (during a clinical pilot study). It enhances human capabilities by providing automatic guideline-based recommendations.	Users are trained to prevent overconfidence in or overreliance on the AI system.

Human oversight	The CAPABLE project considers the appropriate level of human control for the particular AI system.	The system is developed in order to guarantee human control over the AI. Suggestions to physicians from the AI do not imply any automatic actions but they need to be preliminarily accepted. A risk based approach was adopted for suggestions to patients. Recommendations that have a higher risk need a previous approval by the physician whereas educational tips and management of mild symptoms are automatically sent by the virtual coach on the basis of a validated flow chart. AI rules are supervised and the implemented guidelines underwent both clinical and technical validation. We also carried out a risk assessment regarding the whole CAPABLE system highlighting possible issues related to the autonomy of the AI and proposed a mitigation strategy to deal with them. CAPABLE is not a self-learning or autonomous AI system.
Technical robustness and safety		
Resilience to attack and security	The system was assessed to identify potential forms of attacks to which it could be vulnerable.	To assess potential vulnerabilities related to the technical features of the CAPABLE system, we performed a risk assessment analysis in compliance with ISO 14971 and a Data Protection Impact Assessment according to GDPR. The results of both assessments identified all risks as acceptable. Moreover, penetration tests were performed on the system deployed for the clinical study.
Fallback plan and general safety	CABAPLE has a fallback plan in case of adversarial attacks or other unexpected situations.	CAPABLE has considered adversarial attacks and unexpected situations from the very beginning of the project. As the system involves health data and fundamental rights of patients,

		security measures were taken from the moment of the design of the project to its deployment. For example, the system is installed into servers of the clinical sites. DPIAs were carried out as well as a risk assessment analysis. As the system is considered a Medical Device, it underwent the approval of the Ethical Committees and the approval of National Authorities according to EU Regulation 745/2017.
	CAPABLE considers the level of risk raised by the AI system.	A risk assessment analysis, including specific issues related to the AI components of the system, has been carried out. In this analysis we identify potential safety risks of foreseeable uses of the technology, including accidental or malicious misuse, and propose a plan to mitigate or manage those risks. Users (both patients and physicians) are trained and aware of the system features. Moreover patients are informed on their rights to withdraw the clinical investigations and a contact point for doubts is always available upon request. We have insurance policies that cover damages deriving from the use of the CAPABLE system.
	The project assesses whether there is a probable chance that the AI system may cause damage or harm to users or third parties.	In the risk assessment analysis we considered the chance of affecting patients and the measures taken in order to avoid or, at least, mitigate potential harms. A matrix of the risks has been put in place. The matrix assesses and quantified the likelihood, potential damage, impacted audience and severity of each of the identified risks. Risk analysis includes whether security or network problems scould pose safety risks or damage due to unintentional behaviour of the AI system. Such risks have been

	The project estimates the likely impact of a failure of the AI system when it provides wrong results, becomes unavailable, or provides societally unacceptable results.	evaluated both in the DPIAs and in the general risk assessment analysis of the system. In the risk assessment analysis we have estimated the impact of a failure of the AI system by quantifying its likelihood and providing mitigation plans.
Accuracy	The system assesses what level and definition of accuracy is required in the context of the AI system. The project considers action to be put in place if the AI system makes inaccurate predictions.	The CAPABLE AI system is an expert system that relies on (complex) rules that model computer interpretable clinical practice guidelines. The accuracy of the system is measured as the number of correct recommendations in response to the data that are entered on the patients. The accuracy of the expert system has been assessed both by domain experts and by running a technical evaluation based on simulations.
Reliability and reproducibility	We put in place a strategy to monitor and test if the AI system is meeting the goals, purposes and intended applications.	The AI system is going to be tested, as well as the CAPABLE system, in different clinical investigations in Italy and in The Netherlands. Before the clinical investigations, system underwent an evaluation aimed at testing if the AI system is meeting the goals, purposes and intended applications. After the investigations, the clinical efficacy of the intervention will be evaluated by comparing data extracted from the CAPABLE system to those coming from a control cohort treated with standard care.
Privacy and data governance		
Respect for privacy and data Protection	Privacy and personal data protection has been considered as a central and crucial issue in developing the project and deploying it.	Respect of privacy is crucial for CAPABLE and the AI embedded in the system. All Consortium partners were fully involved in the privacy process as well as their DPOs. Fundamental privacy principles such as data minimization and maximum retention period have been considered and applied.

		Legal basis is the patient's consent. In the patient's informed consent, their right to withdraw, revoke or portability are fully reported and explained. As the CAPABLE system deals with special category of data (health data) and fundamental rights and freedom of patients are at risk, DPIAs (one for each clinical site) were adopted. The CAPABLE system is deployed in the servers of the clinical sites, that are safer than other servers outside the hospitals. Personal data coming from previous studies used for developing the models were fully anonymized (NL) or pseudonymized (IT) according to the GDPR. According to the hospitals' privacy policies, specific procedures were defined to allow users and technical staff to flag issues related to privacy or data protection of the data processing.
Quality and integrity of data	We align the system with relevant standards and widely adopted protocols for daily data management and governance. Oversight mechanisms for data collection, storage, processing and use have been established. Processes to ensure the quality and integrity of your data have been put in place. We assess the extent to which you are in control of the quality of the external data sources used.	CAPABLE system is aligned to the UNI CEI EN ISO 14971 In the CAPABLE system data are collected according to the OMOP common data model and exchanged in adherence to the HL7 FHIR standard. Data are stored on servers provided by the clinical centres and managed by their IT staff according to policies and rules established internally. Such rules ensure regular backups, single fault tolerance, access control and firewall rules to prevent unauthorised access and preserve data integrity. Data collected by the smartwatches and stored in the proprietary vendor cloud, which is external to the CAPABLE infrastructure, are used only for post study analysis and not by the AI system.

Access to data	Protocols, processes and procedures were followed to manage and ensure proper data governance.	As the Consortium is composed of several partners located among European Union, Israel and United Kingdom and each partner does not access all data according to minimization principle, we have a) defined the role of each partner (clinical partners are data controller; other partners are data processors); b) each partner provided the controller with a check list of technical and safety measure implemented in their company / institution in order to allow the controller to evaluate their adequacy in processing data; c) settled a Data Processing Agreement and a Data transfer agreement among data controllers and each data processor. The Agreements describe which data processing are allowed, the safety and technical measures implemented and guidelines in case of data breach / loss; In order to guarantee a higher level of protection, the CAPABLE system is designed as a federated system where each software component is deployed on a separated virtual machine. For the same reason, each partner can access the system deployed in the clinical site servers via a secure VPN connection. This connection is limited to the virtual machine(s) necessary to process the data according to the DPA and DTA. Connection to other virtual machines are not possible as the partner does not have the valid permission to access them.
Transparency		

Traceability	We establish measures that can ensure traceability.	The CAPABLE expert system has been implemented using a state of the art modelling framework based on the ProForma language, which is a well-established formalism for the formalisation and implementation of computer interpretable guidelines. All the methodological and implementation steps have been documented in the deliverables of WP5 (D5.2, D5.3, D5.5). The predictive models that were developed during the project, even though not included in the system being piloted on real users, are also documented in the deliverables of WP5 (deliverable D5.1, D5.2, and D5.4). These documents include the details on the methodology used for training the algorithms (including a description of the input data that was selected), and the information about the validation process.
Explainability	We ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand. We designed the AI system with interpretability in mind from the start.	CAPABLE as an expert system is explainable by nature. The module used to implement clinical guidelines include an explainability component that, for each rule, is able to provide the data and the conditions that were responsible for it to be triggered. As for the predictive models, we tried to use the simplest and most interpretable model possible for the application in question.
Communication	We ensure communication to end-users about their interaction with an AI system. We establish a mechanism to inform (end-)users on the reasons and criteria behind the AI system's outcomes.	End-users are informed that the CAPABLE system provides interactions based on AI. This interaction occurs in specific sections of the CAPABLE application that end-users are trained to identify. End-users are also informed on how the AI is working to produce the suggestions and advice, and on its potential limitations. This information is included in the instruction for use and in the consent form signed by end-users.

	We establish processes that consider users' feedback and use this to adapt the system. We clarify the purpose of the AI system and who or what may benefit from the product/service. We specify usage scenarios for the product and clearly communicate these to ensure that it is understandable and appropriate for the intended audience.	The CAPABLE system is developed following a user centred design strategy. We followed an iterative approach that, at each step, considered the user feedback to adapt the system to their needs. Intended users, usage scenarios and intended audience were defined in a system usability process (composed of a plan and a report).
Diversity, non-discrimination and fairness		
Unfair bias avoidance	Measures adopted for unfair bias avoidance.	The CAPABLE system is a first prototype of a monitoring system for cancer patients treated at home. As such it has considered AI strategies that rely on the one hand on the implementation of clinical practice guidelines and on the other hand on predictive models. Both these AI methodologies are prone to bias. The system has been designed by taking this into account. As regards the predictive models, we decided not to include them in the system being piloted on real users, and one of the reasons was related to the lack of a robust evaluation of the sources of bias, the possibility of generalising to different populations, and the possibility of continuous updating of the models. Such evaluation is being performed on new data sources. As regards the expert system, we are going to evaluate possible biases during the analysis of the pilot study data and identify possible strategies to deal with such bias, for example by coupling clinical guidelines to real world evidence.
Accessibility and universal design	Measures to ensure: ● accessibility by people with special needs or disabilities. ● Involvement of the community in the development of the AI system. ● Representativeness of the target users audience during	The CAPABLE system is meant to be used by oncological patients, which is a category of users with special needs and vulnerabilities. For this reason, we took such special needs into account from the start of the project. A patient association was deeply involved in the design of the

	system design. ● Users feedback	system with the goal of being as inclusive as possible for adoption, also considering groups who might tangentially be impacted. Several iterations of usability and user experience studies were performed throughout the development process, to ensure that the whole system is accessible and easy to use for the target population. The team that performed such evaluations was selected to represent as widely as possible the target users population.
Stakeholder participation	We consider a mechanism to include the participation of different stakeholders in the AI system's development and use.	The project has considered the participation of different stakeholders from the very beginning and since the proposal of the project. Indeed, in order to guarantee stakeholder participation, clinical partners, technical ones and cancer patient associations have been involved in the Consortium. Patient participation was also guaranteed during the development of the CAPABLE system as each development step of the end-user interface underwent a usability evaluation from patients.
Societal and environmental well-being		
Sustainable and environmentally friendly AI	Mechanisms to measure the environmental impact of the AI system's development.	As the system is deployed on servers of the clinical partners, we are not in charge of evaluating the environmental impact of the AI system's development. As the CAPABLE system is a pilot project, our aims, at this stage, were focused on other primary issues such as its usefulness, data protection and usability.
Social impact	Evaluation of the social interactions of AI with the users and of the risks of de-skilling of the workforce.	The CAPABLE system does not include a social interaction with the users, so there is no risk to develop attachment or empathy towards the system. Being a clinical decision support system, human intervention is always requested in the loop. The use of CAPABLE is meant to support

		the physicians in their daily activities and improve the patients conditions by constant monitoring. No risk of job loss or de-skilling is foreseen.
Society and democracy	Assessment of the broader societal impact of the AI system's use beyond the individual (end-)user.	During the lifetime of the project the main goal has been to develop the system to satisfy the needs of the intended end users. In case the system is exploited into a product or service, an assessment of the broader societal impact of CAPABLE will need to be planned.
Accountability		
Auditability	Mechanisms to facilitate the system's auditability. Independent auditing.	The CAPABLE system components, including the AI ones all include logging functionalities and traceability. Since CAPABLE has a federated system architecture, it is possible to decouple the individual components and perform separate tests able to replicate the behaviour of the production system. This would allow the independent auditing of the AI system.
Minimising and reporting negative Impact	Rsk or impact assessment of the AI system, taking into account different stakeholders that are (in)directly affected. Processes to report potential vulnerabilities in the AI system. Assessment of the ethics practices of the project	A risk and impact assessment was carried out for the AI system, according to ISO 14971. The CAPABLE team has defined processes for reporting potential vulnerabilities in the AI components. For end users, these can be found in the instructions for use of the system. As a research project funded by EU CAPABLE underwent an ethics review by an independent panel of experts.
Documenting trade-offs	Did you establish a mechanism to identify relevant interests and values implicated by the AI system and potential trade-offs between them?	In the DPIAs and in the risk assessment analysis we identified relevant interests and values implicated by the AI system and

	How do you decide on such trade-offs? Did you ensure that the trade-off decision was documented?	potential trade-offs between them.
Ability to redress	We establish a set of mechanisms that allows for redress in case of the occurrence of any harm or adverse impact.	Specific insurance policies for the project were set up. Moreover, patients have a contact point at their disposal where explanation on how to sue an action can be provided.

6 References

1. The Report of the Presidential Commission for the Study of bioethical issues, 2016 “For Researchers: Incidental and Secondary Findings”
2. Presidential Commission for the Study of Bioethical Issues (PCSBI). (2013, December). Anticipate and Communicate: Ethical Management of Incidental and Secondary Findings in the Clinical, Research, and Direct-to-Consumer Contexts. Washington, DC: PCSBI, p. 27. Note: DTC = direct-to-consumer, ACMG = American College of Medical Genetics and Genomics.
3. Damjanovicova M. (2016) Incidental Findings. In: Boniolo G., Sanchini V. (eds) Ethical Counselling and Medical Decision-Making in the Era of Personalised Medicine. SpringerBriefs in Applied Sciences and Technology. Springer, Cham. http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-319-27690-8_9
4. Bunnik, Eline M., et al. "Ethical framework for the detection, management and communication of incidental findings in imaging studies, building on an interview study of researchers' practices and perspectives." BMC Medical Ethics, vol. 18, no. 1, 2017
5. Sutton DR, Fox J. The syntax and semantics of the PROforma guideline modeling language. J Am Med Inform Assoc. 2003;10(5):433-443. doi:10.1197/jamia.M1264
6. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
7. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf